

RESEARCH ARTICLE OPEN ACCESS

Operational Machine Learning Post-Processing of Short-Range Temperature, Humidity, Wind Speed and Gust Forecasts

Leila Hieta  | Mikko Partio

Finnish Meteorological Institute, Helsinki, Finland

Correspondence: Leila Hieta (leila.hieta@fmi.fi)**Received:** 30 December 2024 | **Revised:** 25 April 2025 | **Accepted:** 26 June 2025**Funding:** The authors received no specific funding for this work.**Keywords:** bias correction | nowcasting | operational forecasting | XGBoost

ABSTRACT

Statistical methods can be used to create bias correction models that learn from past forecast errors and reduce systematic errors in real-time forecasts. This study presents a machine learning (ML) approach using extreme gradient-boosted (XGBoost) trees to address biases in a numerical weather prediction (NWP) nowcast model for key meteorological parameters: 2-m temperature, 2-m relative humidity, 10-m wind speed, and 10-m wind gust. These ML models have been integrated into the Finnish Meteorological Institute's (FMI) operational nowcasting framework, Smartmet nowcast. Results show that, even with a relatively modest set of meteorological predictors, the ML bias correction method significantly improves forecast accuracy, reducing the root mean square error (RMSE) by 24%–29% compared to the direct NWP model output. The implementation of this new bias correction method not only improves the quality of FMI's short-range forecasts, but also extends the availability of bias-corrected data for longer forecast lead times, offering substantial improvements over the previously implemented bias correction method. The codebase for this machine learning bias correction is available at (https://github.com/fmidev/snwc_bc).

1 | Introduction

The production of frequently updated, high-quality nowcast information, defined by the World Meteorological Organization (WMO) to cover a timescale of 0–6 h (Wang et al. 2017), is essential for supporting social functions, enhancing weather preparedness, and assisting the public in daily decision making. Traditional numerical weather prediction (NWP) models with hourly update frequency and sub-hourly temporal resolution are typically used to produce short-range forecasts. However, these models often fail to fully exploit the potential of available observational data during data assimilation and are prone to systematic errors (Huang et al. 2012; Dee 2005). The accuracy of forecasts in the nowcast range can be significantly improved by

integrating observation-derived information into the traditional NWP approach.

Methods for generating observation-based nowcast information vary from bias correction to observation extrapolation techniques, depending on the weather parameter being forecast and the available observational data. Purely observation-based methods, traditionally grounded in motion vector techniques, are widely used in many nowcasting systems, especially for precipitation (Bowler et al. 2006; Germann and Zawadzki 2002, 2004; Pulkkinen et al. 2019). More recently, neural network-based methodologies have gained prominence, particularly for parameters where high spatial resolution observational data is available (Ayzel et al. 2020; Gao et al. 2021; Leinonen et al. 2023;

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Meteorological Applications* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

Ravuri et al. 2021). Although observation-based methods excel in tracking the movement of weather systems, they often face challenges in representing the evolution of weather conditions (Pulkkinen et al. 2019). To address this limitation, an observation-based nowcast can be seamlessly integrated into the NWP forecast to provide a more comprehensive representation of weather evolution. A comprehensive guide to nowcasting and related techniques can be found in Wang et al. (2017).

Bias correction methods have long been employed to improve the accuracy of operational weather forecasts by reducing the systematic errors of NWP models. In bias correction, a statistical model correlating historical forecasts and observations is used to correct the biases in future forecasts. Bias correction methods can be divided into methods that are based solely on recent forecast error information, such as Kalman filtering (Galanis and Anadranistakis 2002) and decaying average (Cui et al. 2012), and methods that utilize a wider set of predictors and a longer history of error correlations, such as model output statistics (MOS) (Glahn and Lowry 1972). In recent years, more sophisticated machine learning methods have emerged in which bias correction models can learn to account for more dynamical error corrections in operational forecasts, for example, Ben Bouallègue et al. (2023).

Gradient boosted decision trees (GBDT) (Friedman 2001; Mason et al. 1999) are a class of ensemble learning methods that have shown strong predictive performance in a variety of real-world tasks. Extreme gradient boosting (XGBoost) (Chen and Guestrin 2016), the implementation used in this study, is a widely used framework for GBDT. In meteorological applications, GBDT-based models have been applied to various forecasting problems, including wind speed prediction (Fang et al. 2023; Xu et al. 2020), solar radiation forecasting (Bakker et al. 2019), and fog occurrence estimation (Castillo-Botón et al. 2022), among others. These examples illustrate the potential of the method to handle complex, nonlinear relationships common in atmospheric data, supporting its application to bias correction in weather forecasting.

Several national meteorological services (NMS) have developed integrated nowcasting frameworks that combine multiple nowcasting techniques with NWP data to provide automated, frequently updated nowcasts for the public. Notable examples of such systems are the German weather service's (DWD) NowCastMIX (James et al. 2018), developed especially for convective events, and SINFONY (Seamless INtegrated FOrecastiNg sYstem) (Blahak et al. 2021), also developed by DWD, which produces ensemble output. Other examples are Geosphere Austria's INCA (integrated nowcasting through comprehensive analysis) (Haiden et al. 2011) particularly suited for complex terrain, and IMPROVER by Met Office in the United Kingdom (Roberts et al. 2023) which focuses on probabilistic post-processing. While machine learning-based methods have been explored in operational nowcasting systems, to the best of our knowledge, a bias correction scheme employing an advanced gradient-boosting technique, as presented in this study, has not yet been implemented operationally.

The Finnish Meteorological Institute (FMI) has developed a nowcasting system known as Smartmet nowcast (Hieta

et al. 2021), which integrates observation-based data and forecasts from NWP models to produce seamless, rapidly updating, and high-resolution deterministic nowcasts for key meteorological parameters. This system was incorporated into FMI's operational weather forecast production in November 2021, leading to improvements in forecast accuracy while reducing both operational delays and the need for manual forecast editing by forecasters. However, areas for improvement were identified, particularly in the bias correction method applied to parameters such as 2-m temperature, 10-m wind speed, and 2-m relative humidity. The initial bias correction method, described in Hieta et al. (2021), relied on the latest forecast—observation error information and was used to correct the present state and the subsequent four forecast hours using static error weighting based on 1 year of historical data. Although this approach improved the accuracy of the NWP output, it did not account for the diurnal variation of errors and could not be extended for longer lead times. In addition to the parameters addressed in this article, the Smartmet nowcast system also provides nowcasts for 1-h precipitation accumulation generated using the pySTEPS framework (Pulkkinen et al. 2019), total cloud cover based on a neural network model (Partio et al. 2024), and probability of thunder derived from a pySTEPS-based technique (Korpinen et al. 2024). The latter two parameters were added after the system was operationalized in 2021.

To address the identified limitations in FMI's operational nowcasting system, this study aims to develop and evaluate a machine learning bias correction framework that improves forecast accuracy for key meteorological parameters. An extreme gradient-boosted model (XGBoost) is developed to replace the previous bias correction method in Smartmet nowcast. The new framework dynamically adapts to varying error patterns, including diurnal variations, and extends bias correction lead times up to 11 h. Additionally, wind gust is incorporated as a new parameter within the bias correction process. By leveraging machine learning techniques, this study seeks to enhance the precision and adaptability of FMI's nowcasting system, representing a significant step forward in operational weather forecasting.

2 | Machine Learning Bias Correction Method

2.1 | Training Data

The MetCoOp nowcast (MNWC) NWP model data forms the main input for the Smartmet nowcast system and was the primary source of training data for this study. The MNWC is a product of the Meteorological Cooperation on Operational NWP (MetCoOp) (Müller et al. 2017), a collaborative effort between several NMS, providing high-resolution limited area forecasts for Northern Europe. The MNWC is a convection-permitting model based on HARMONIE-AROME (Bengtsson et al. 2017; Seity et al. 2011), using a 3D-Var data assimilation method that incorporates surface observations such as 2-m temperature, 2-m relative humidity, and surface pressure (MetCoOp 2024). The MNWC model is run hourly, generating 12-h forecasts at a horizontal resolution of 2.5 km with 15-min output intervals. The high update frequency allows for more timely incorporation of recent observations, making it particularly suitable for nowcasting applications. For bias

TABLE 1 | The features and their short names available for XGBoost model training for the target parameters: temperature, relative humidity, wind speed and wind gust. The features marked with X are included in the training. The rows marked with gray define the features that are direct MNWC model output.

Feature	Short name	Temperature	Wind speed	Wind gust	Relative humidity
2 m temperature	T2M	X	X	X	X
2 m relative humidity	RH2M	X	X	X	X
10 m wind speed	S10M	X	X	X	X
10 m wind gust	GMAX		X	X	
2 m specific humidity	Q2M	X			X
Low level cloud cover	CCLOW	X			X
Surface pressure	PMSL	X	X	X	X
10 m wind direction	D10M		X	X	
Observation elevation	obs_elevation	X	X	X	X
Latitude/longitude	obs_lat/lon	X	X	X	X
Latest model error	1bias	X	X	X	X
MNWC—obs elevation	ElevD	X	X	X	X
Time of day	hour_sin/cos	X	X	X	X
Month	month_sin/cos	X	X	X	X
Lead time	leadtime	X	X	X	X

correction purposes, only the forecasts with a 1-h output interval were used, as this matches the temporal resolution of the 10-day operational forecast data into which the bias-corrected data is integrated. In contrast, for other parameters within the Smartmet nowcast system (precipitation, total cloud cover, and probability of thunder) higher temporal resolution datasets with 15-min update and output intervals are produced to support time-sensitive nowcasting applications with shorter lead times.

For the training of the XGBoost model, only data bilinearly interpolated from MNWC fields to synoptic weather station locations within the MNWC domain were available. The available MNWC dataset shown in Table 1 was relatively limited, focusing mainly on the surface-level parameters regularly used for verification. As no gridded MNWC data were accessible for training, the point data were used instead.

Approximately three and a half years of data, spanning January 2020 to July 2023, was used to train the bias correction model. One year of data, from August 2023 to July 2024, was reserved as an entirely independent test dataset. To address computational demands and improve training efficiency, a data reduction scheme was applied. Instead of including data from all hourly MNWC model runs, only data from every other model run was selected for the training dataset. To further reduce the data volume, only every other forecast lead time was used from these selected model runs. The data reduction scheme is illustrated in Table 2, where the lead times highlighted in green indicate those chosen for the final training dataset. This approach effectively reduced the volume of the training data to approximately one-third of the total available

TABLE 2 | An example illustrating the data reduction scheme. The rows represent seven consecutive, hourly MNWC model runs starting from 00 UTC analysis time, with forecast lead times of 0–9 h. The forecast lead times selected for the dataset used to develop XGBoost machine learning models are marked in green.

Analysis time	Forecast lead time (h)									
00 UTC	0	1	2	3	4	5	6	7	8	9
01 UTC	0	1	2	3	4	5	6	7	8	9
02 UTC	0	1	2	3	4	5	6	7	8	9
03 UTC	0	1	2	3	4	5	6	7	8	9
04 UTC	0	1	2	3	4	5	6	7	8	9
05 UTC	0	1	2	3	4	5	6	7	8	9
06 UTC	0	1	2	3	4	5	6	7	8	9

data, streamlining the training process and mitigating redundancy in the hourly updating model runs. In January 2023, the MNWC forecast length was extended from 9 to 12 h, and therefore the training data include only the shorter lead times available before this extension. The final training dataset comprised approximately 40 million rows.

Hourly SYNOP observations of 2-m temperature, 2-m relative humidity, 10-m wind speed, and 10-m hourly maximum wind gust were used as ground truth data within the MNWC domain. For wind speed observations for Finland, observations corrected to represent the WMO standard for wind speed measurement were used that are produced operationally

at FMI (Aaltonen and Frisk 2019; World Meteorological Organization 2023). The wind speed correction adjusts the raw observations for both the measurement height and the surface roughness that the wind has passed over. For stations where corrected wind speed observations are not available, the SYNOP observations were used instead. Forecast–observation errors served as the target variable in the bias correction model. The training dataset contains data from 795 to 888 stations depending on the parameter, the wind gust having the lowest amount of SYNOP observation stations.

Quality control was performed on the observational data to remove erroneous observations. The implemented method is based on z -score, also known as the standard score, which indicates how many standard deviations a value deviates from the mean of the distribution. Values with deviations larger than a defined z -score threshold are identified as outliers or, in this context, erroneous observations. The quality control method is applied separately to each parameter by evaluating the distribution of all SYNOP observations within the MNWC domain at each hour. The use of hourly distributions was chosen to allow the same method to be efficiently applied in the real-time bias correction system, which is run operationally on an hourly basis. Separate z -score thresholds were applied for the upper and lower tails, varying according to the season, to take into account the skewed distributions. This approach allows, for example, colder temperatures during the winter months to the distribution without being classified as outliers.

Although quality control may occasionally exclude true observations, particularly important extreme values, the total number of detected outliers across all parameters is relatively low, accounting for only 0.01%–0.07% of the entire observational dataset, depending on the parameter.

In addition to the MNWC forecast data, several supplementary features summarized in Table 1 were used for model training across all parameters: elevation of the observation station, latitude and longitude of the observation station, the elevation difference between the MNWC model and observation station, time of day and month expressed in cyclical sine and cosine representations, forecast lead time, and the forecast–observation error from the most recent observation.

2.2 | Machine Learning Model

The objective of the machine learning model developed in this study is to learn and predict bias in the MNWC NWP model using historical data, making it a supervised regression problem. In addition to XGBoost, multiple linear regression and random forest models (Breiman 2001) were considered when selecting a suitable machine learning method. Multiple linear regression is widely used in many practical applications due to its simplicity and interpretability (Kutner et al. 2004). However, linear regression is only capable of representing linear relationships between features and the target parameter. To provide a baseline for comparison, a multiple linear regression model was also trained and evaluated alongside XGBoost during the validation phase. The random forest model originates from decision trees, which are flowchart-like structures splitting data into branches based on

feature values. In random forest, an ensemble of decision trees is created from random subsets of training data and features, and the final model is averaged across the trees, making the model generalize better on unseen data than a single tree. Gradient-boosted trees (e.g., XGBoost) take a similar approach to random forest models, but instead of building the ensemble of trees in parallel, trees are built sequentially, optimizing the model to learn the errors of the previously built trees in a computationally efficient fashion. XGBoost is known to be robust, relatively easy to develop, and a popular choice in various real-world applications, all of which are desirable characteristics for the ML model used for operational production (Chen and Guestrin 2016). For these reasons, XGBoost was selected as the machine learning model for this study.

Four separate XGBoost models were constructed to predict the bias for different target parameters. The developed models are universal in nature: all station-points were used to train a single model per parameter, which can be applied across all stations and lead times. This approach is particularly practical for operational use, as it allows one model to serve all points simultaneously. Moreover, if new observation stations become available, the method can be extended to these locations without retraining, provided that the required input features are available.

A SHAP analysis (SHapley Additive exPlanations) (Lundberg and Lee 2017) was performed to evaluate the importance of the features used in the machine learning models. Figure 1 shows the average impact of each feature on the prediction of the model, highlighting that the most recent forecast error information and the MNWC forecast of the target parameter are often the most influential features. The importance of the latest error information reflects the persistence of forecast errors, which often remain consistent over short lead times in NWP models (Wang et al. 2017). The analysis shows that time-dependent features, such as the time of day and month, have a stronger influence on the models developed for relative humidity (RH) and temperature (T2m), compared to those for wind speed (WS) and wind gust (WG), where the impact of time-related features is less significant. Based on the SHAP values and expert knowledge, some feature selection was performed. Table 1 represents the set of features used for different target parameters. However, all supplementary features that are not direct NWP output were retained, even those with minimal impact on the models, because of the already limited number of available features. This decision ensures that the model has access to all potentially useful information, despite the small set of features.

Hyperparameter tuning was performed using blocked cross-validation with four annual time-based splits. As the training dataset spans approximately three and a half years, the final block represents a half-year period. The hyperparameters of the models were optimized using the BayesSearchCV method from the scikit-optimize library (Head et al. 2021), which employs a cross-validated search to identify the optimal combination of hyperparameters. This optimization process led to a substantial improvement in model performance, with the estimated RMSE of the forecasts reduced by 15%–24% compared to models trained using the default XGBoost hyperparameter settings.

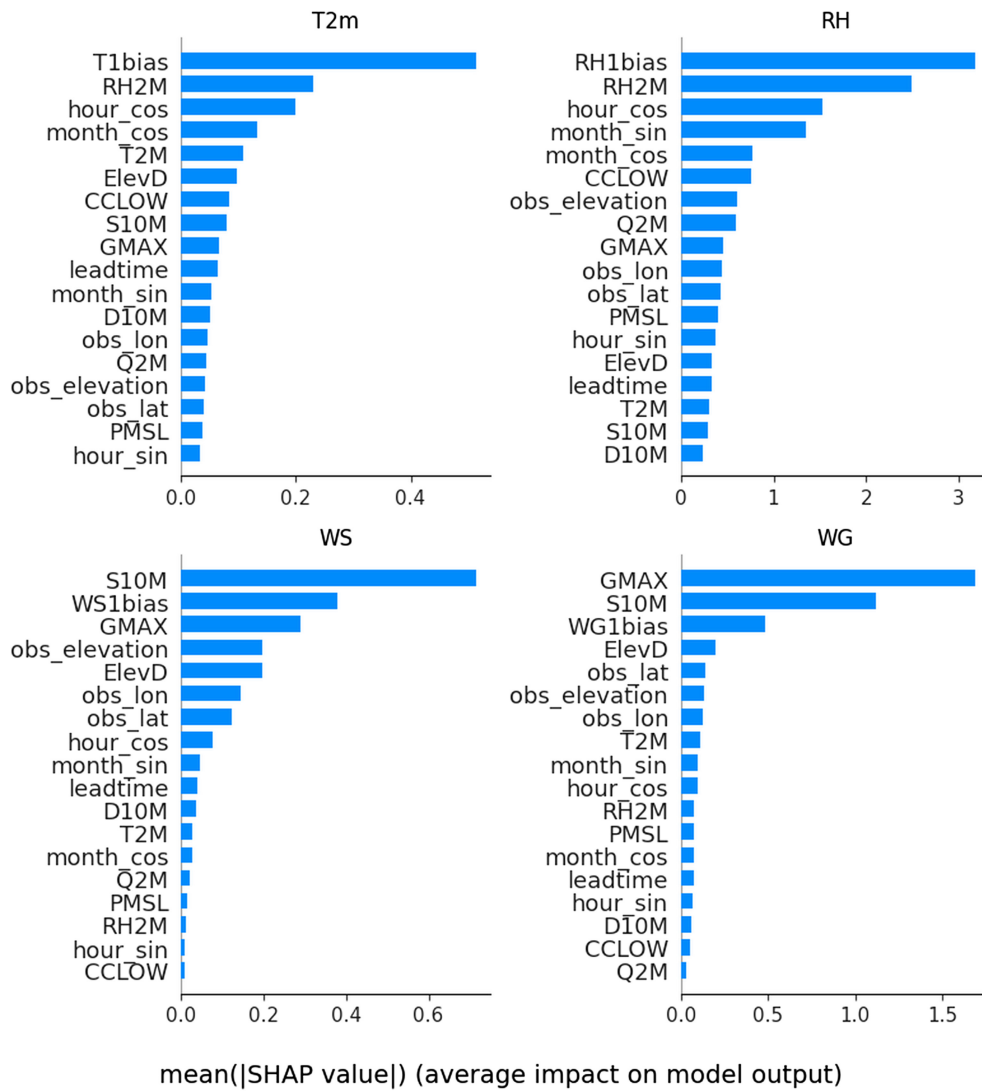


FIGURE 1 | Mean SHAP values showing the relative importance of features for predicting each target parameter: 2-m temperature (T2m), relative humidity (RH), wind speed (WS), and wind gust (WG). Higher SHAP values indicate features with a greater influence on the model's predictions. The short names used for the features are explained more comprehensively in Table 1.

2.3 | Operational Implementation

The XGBoost bias-correction models are used in FMI's operational Smartmet nowcast system to produce bias-corrected MNWC forecasts. The production is run hourly and uses both the latest observation and real-time MNWC forecast information. The MNWC forecast data is available approximately 1 h after the model initialization. Due to this delay, the 1-h forecast (instead of the 0-h analysis time) from MNWC is used as the present state, and the rest of the available MNWC lead times (2–12 h) are bias-corrected, resulting in an 11-h forecast in total. The production is run 20 min past the exact hour to allow sufficient time to collect observations. These SYNOP observations are used to calculate the latest forecast—observation error for the present hour. This information is used to create an accurate representation of the current state of the weather by correcting the MNWC 1-h forecasts. This latest forecast error information is also used as a predictor in XGBoost models. Although the training data only contain the MNWC forecasts with lead times up to 9 h, the XGBoost models are also run for the currently available longer lead times unseen to the model training.

The operational production relies on gridded data, which is why the pointwise biases produced for each lead time are interpolated to a regular grid using the gridpp software package (<https://github.com/metno/gridpp>), developed by the Norwegian Meteorological Institute. Gridpp provides the Optimal Interpolation (OI) (Gandin 1963) method, a statistical interpolation technique commonly used in data assimilation to combine the gridded NWP background field with meteorological point observations. Several user-defined parameters control the interpolation process. For instance, the weighting of point data and the background field can be adjusted based on their respective uncertainties. Additionally, horizontal and vertical decorrelation length scales and land area fraction information can be specified to control the spatial extent of influence for individual observations.

The Optimal Interpolation is applied to generate gridded bias-correction fields from the forecasted pointwise bias information. These resulting gridded bias fields are then subtracted from the original MNWC forecasts to produce bias-corrected forecasts. The forecasts are then integrated into the Smartmet nowcast

system and blended with longer forecast lead times, as described in Hieta et al. (2021).

3 | Results

To validate the XGBoost models, an independent dataset was used covering the period of 1 year from August 2023 to July 2024 to account for the different seasons. The validation was done for bias corrected point forecasts before interpolation to grids. The test dataset was subjected to the same data reduction process as applied to the training data explained in Section 2.1. The MNWC model forecasts used as reference were taken from runs initialized 1 h after those used for the XGBoost-based bias correction. This setup reflects the actual operational scenario and compares bias correction with MNWC data that would have been available to end-users at the corresponding time. To benchmark the performance of the XGBoost model, a multiple linear regression model was also trained and evaluated using the same training and test datasets and the same features as XGBoost bias correction. SYNOP observations were used as ground truth in the verification. The RMSE and mean error (ME) results for MNWC,

XGBoost bias correction, the linear regression bias correction, and the previously used weight-based bias correction, hereafter referenced as “old bc” (Hieta et al. 2021), are shown in Figure 2 across all seasons and SYNOP stations. The 0-h lead time represents the present state corrected by observations in old bc, XGBoost, and linear regression-based bias corrections, which explains the zero ME and RMSE values. The ME results in Figure 2 show that for MNWC the bias for temperature and humidity is close to zero, implying that the model is well calibrated, but for wind speed and gust a slight overforecasting is observed. The XGBoost and linear regression bias correction are able to reduce the bias close to zero for all the parameters as expected.

The RMSE results show that the XGBoost bias correction is able to improve the results for all lead times compared to MNWC, the old bc method, and the linear regression used as a benchmark. The improvement in RMSE compared to MNWC varies from 17% to 43% depending on the lead time and the parameter, whereas the improvement achieved by the linear regression model compared to MNWC varies from 5% to 35%. For wind speed and gust, the RMSE results appear relatively static, showing minimal variation across lead times, compared

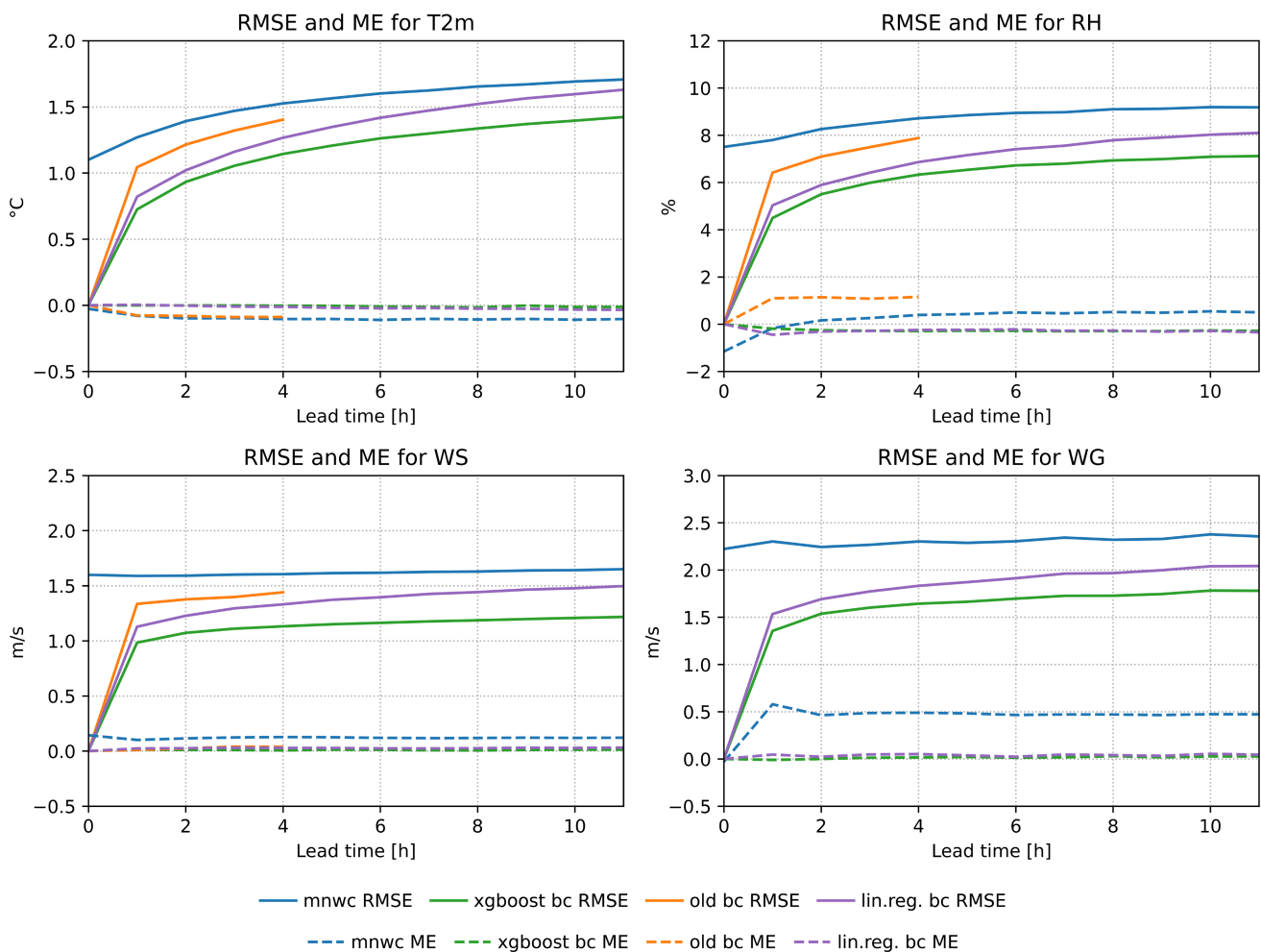


FIGURE 2 | RMSE (solid line) and ME (dashed line) results by forecast lead time for MNWC (mnwc), XGBoost bias correction (xgboost bc), linear regression bias correction (lin.reg.bc), and for the old weight-based bias correction (old bc) for temperature (T2m), relative humidity (RH), wind speed (WS) and wind gust (WG). Note that old bc is not shown for WG, as the method was not originally developed for this parameter. The results are calculated for the period of August 2023 to July 2024 across all station points.

to temperature and relative humidity where the values increase slightly for longer lead times. For MNWC, the lack of data assimilation for WS and WG explains the static RMSE results across different lead times. Furthermore, XGBoost bias correction improves the forecasts also for lead times 9–11, despite these lead times not being included in the training data. Figure 3 shows the monthly RMSE and ME of the MNWC, XGBoost bias-corrected, and linear regression bias-corrected forecasts. The winter months (DJF) exhibit slightly higher values for the monthly RMSE, as winter is the hardest season to forecast for T2m, WS, and WG. For temperature, this is explained by the large variability during winter when solar radiation is minimal and the temperature does not have predictable diurnal variation, but is strongly dependent on the low-level cloud cover (Hyrkkänen et al. 2016). For RH, the spring (MAM) has the highest RMSE values, with MNWC ME indicating slight overforecasting of humidity. Overall, the bias-corrected forecasts are performing as expected, following the overall monthly RMSE of the MNWC forecast, but consistently having slightly lower values, whereas the ME values over all the months are close to zero. Among the bias correction methods,

the XGBoost-based correction yields the lowest RMSE, outperforming the linear regression approach.

The mean absolute error skill score (MAESS), calculated as shown in Equation (1), measures the relative performance of a forecast compared to a reference forecast in terms of Mean Absolute Error (MAE). Positive MAESS values indicate that the forecast has better accuracy than the reference, and negative values indicate the opposite.

$$\text{MAESS} = 1 - \frac{\text{MAE}_{\text{forecast}}}{\text{MAE}_{\text{reference}}} \quad (1)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

The MAESS values calculated at SYNOP station points are shown in Figure 4, where MAE values for the XGBoost bias correction are used as the forecast and MAE values for MNWC as reference. The MAE values are calculated over the entire test dataset and for all the lead times from 1 to 11 h. The

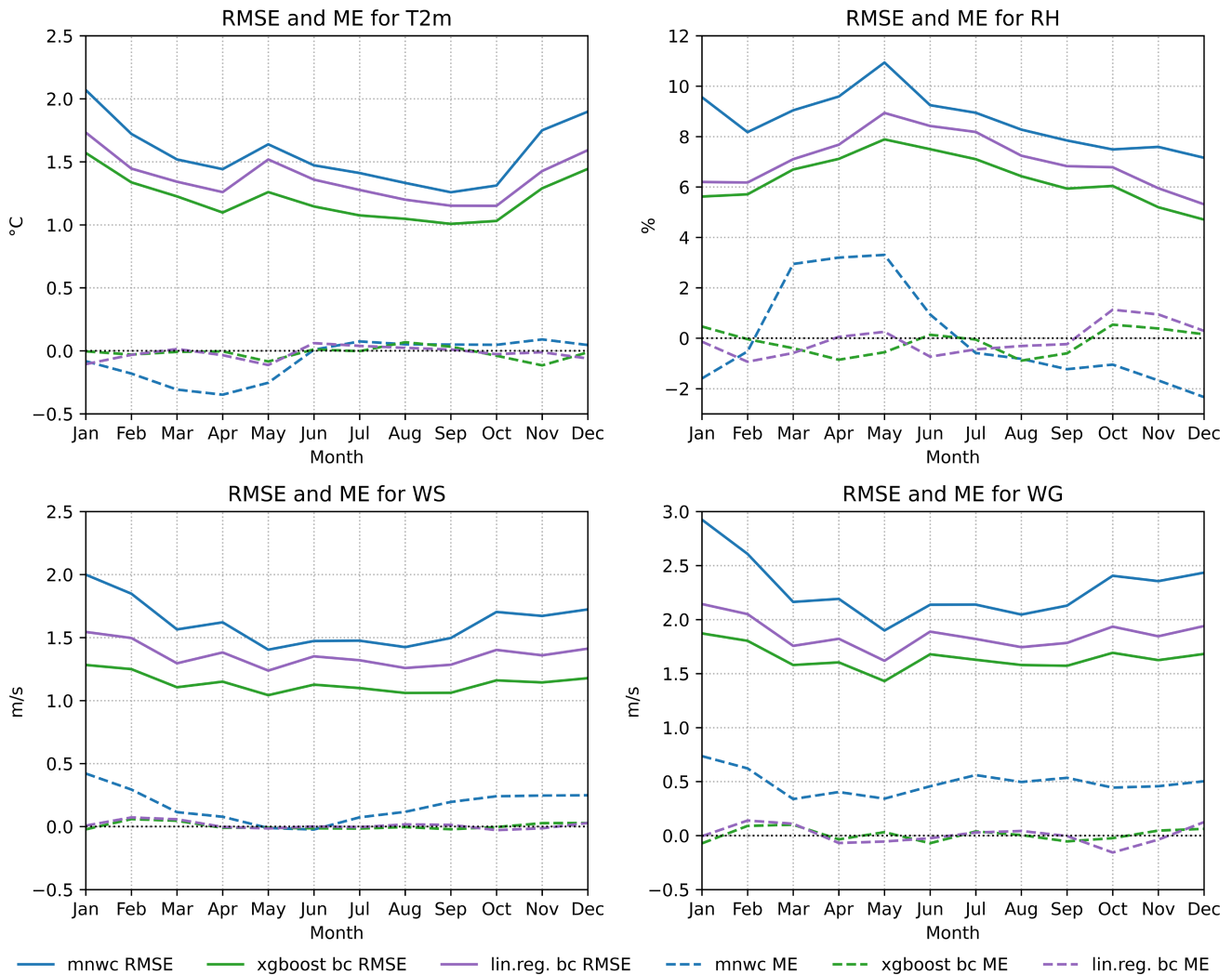


FIGURE 3 | RMSE (solid line) and ME (dashed line) results by month for MNWC (mnwc), XGBoost bias correction (xgboost bc), and linear regression bias correction (lin.reg.bc) for Temperature (T2m), Relative humidity (RH), Wind speed (WS) and Wind gust (WG). The results are calculated for the period of August 2023 to July 2024 across all station points.

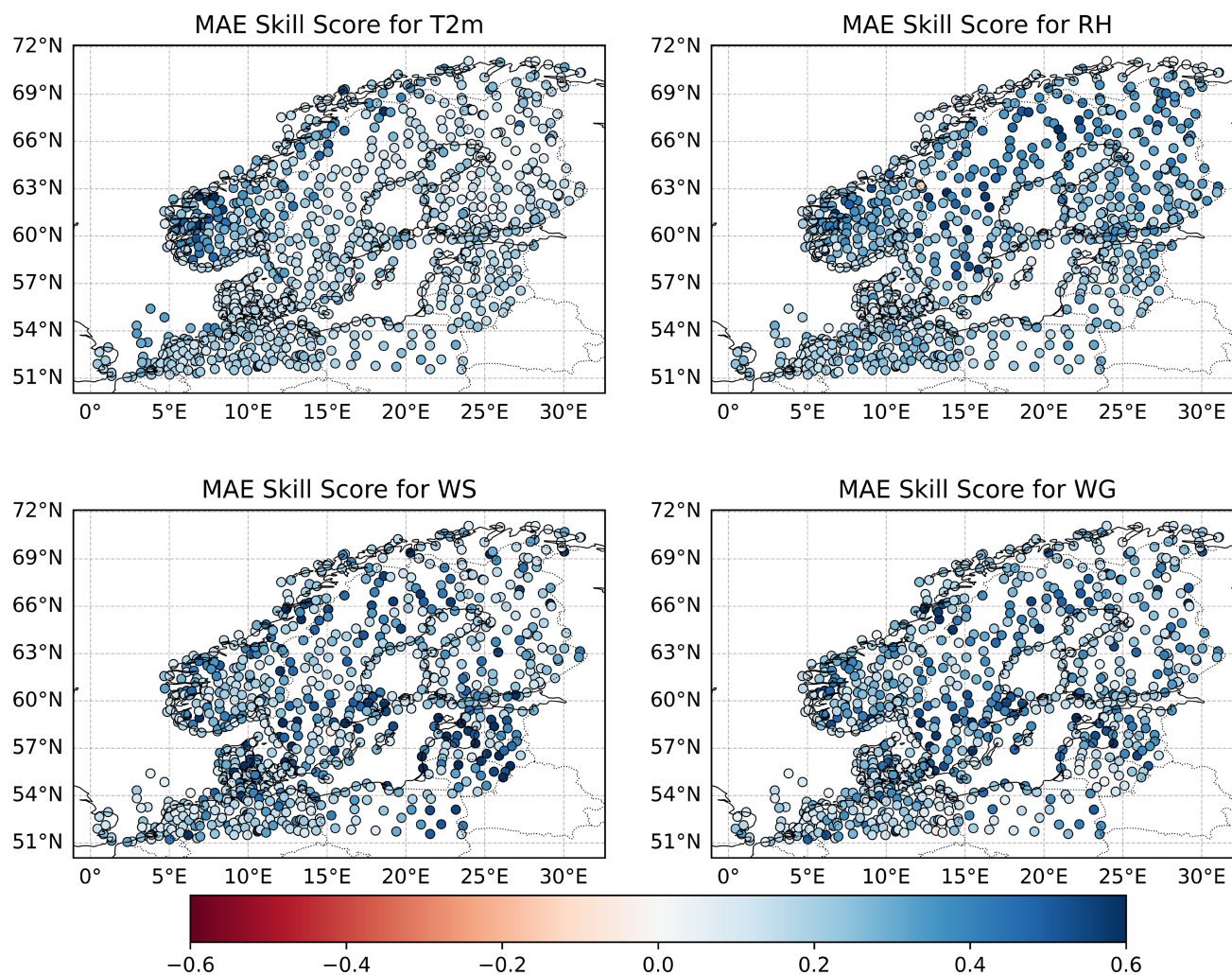


FIGURE 4 | Stationwise mean absolute error skill score (MAESS) results for temperature (T2m), relative humidity (RH), wind speed (WS) and wind gust (WG) for August 2023 to July 2024. In MAESS the MAE results of XGBoost bias correction are compared to the MAE results of MNWC which serves as reference. The positive (blue) values indicating that the XGBoost bias-corrected forecast has better accuracy than the MNWC reference and the negative (red) values the opposite.

results demonstrate that the XGBoost bias correction improves the MNWC forecasts at nearly all station points. However, a slight decrease in performance can be seen at one station point for relative humidity over the Scandinavian mountain range. No specific reason could be identified for this behavior. The results also show spatial variability in the MAESS values for wind speed and wind gusts, suggesting regional differences in XGBoost bias correction performance. For temperature and relative humidity, the MAESS values exhibit less variability across station locations, and for temperature, the greatest improvements are observed in the Scandinavian mountain range.

4 | Conclusions

XGBoost machine learning models were developed to bias correct NWP nowcast model forecasts for key meteorological parameters: 2-m temperature, 2-m relative humidity, 10-m wind speed, and 10-m wind gust. This bias correction approach significantly improved the accuracy of the nowcast forecasts,

resulting in a 24%–29% reduction in RMSE compared to the raw output of the NWP model.

The integration of these models into the FMI's operational Smartmet nowcast system extended the availability of accurate, bias-corrected forecasts by 7 h beyond the previously implemented method based on static error weights (Hieta et al. 2021). This advancement not only enhanced the forecast quality, but also provided more reliable information for lead times up to 11 h. This new machine learning-based scheme offers a more dynamic and accurate approach to bias correction in the nowcasting range.

Future research could explore the use of gridded training data instead of point data to avoid interpolation in the process, provided that suitable datasets become available. Additionally, applying more advanced machine learning models offers an opportunity to further enhance forecast accuracy and reliability. Methods that incorporate ensemble forecasts to account for uncertainty, as well as approaches with higher update and output frequency, may also be considered as potential extensions. Recent progress

in regional AI weather forecasting models has shown promising results (Oskarsson et al. 2023; Nipen et al. 2024). These models, which are based on artificial neural networks, can simultaneously predict multiple weather parameters at high resolution. Exploring such approaches also in nowcasting represents an exciting opportunity for future developments in the field.

Author Contributions

Leila Hieta: writing – original draft, methodology, validation, visualization, software, writing – review and editing. **Mikko Partio:** writing – review and editing, software, data curation.

Acknowledgments

We thank Victoria Sinclair and Marko Laine for providing helpful comments on a final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Aaltonen, A., and M. Frisk. 2019. "Potential Wind Speed – Homogenization of Wind Measurements Over Heterogeneous Terrain." In *EMS Annual Meeting 2019*. Copernicus Publications. <https://meetings.copernicus.org/EMS2019/EMS2019-854.pdf>.
- Ayzel, G., T. Scheffer, and M. Heistermann. 2020. "RainNet v1.0: A Convolutional Neural Network for Radar-Based Precipitation Nowcasting." *Geoscientific Model Development* 13: 2631–2644. <https://doi.org/10.5194/gmd-13-2631-2020>.
- Bakker, K., K. Whan, W. Knap, and M. Schmeits. 2019. "Comparison of Statistical Post-Processing Methods for Probabilistic NWP Forecasts of Solar Radiation." *Solar Energy* 191: 138–150. <https://doi.org/10.1016/j.solener.2019.08.044>.
- Ben Bouallègue, Z., F. Cooper, M. Chantry, P. Düben, P. Bechtold, and I. Sandu. 2023. "Statistical Modeling of 2-m Temperature and 10-m Wind Speed Forecast Errors." *Monthly Weather Review* 151: 897–911. <https://doi.org/10.1175/MWR-D-22-0107.1>.
- Bengtsson, L., U. Andrae, T. Aspelien, et al. 2017. "The Harmonie-Arome Model Configuration in the Aladin-Hirlam NWP System." *Monthly Weather Review* 145: 1919–1935. <https://doi.org/10.1175/MWR-D-16-0417.1>.
- Blahak, U., J. Keller, and SINFONY-Team. 2021. "SINFONY - The Combination of Nowcasting and Numerical Weather Prediction at the Convective Scale at DWD." In *EMS Annual Meeting 2021*. Copernicus Publications. <https://doi.org/10.5194/ems2021-284.EMS2021-284>.
- Bowler, N. E., C. E. Pierce, and A. W. Seed. 2006. "STEPS: A Probabilistic Precipitation Forecasting Scheme Which Merges an Extrapolation Nowcast With Downscaled NWP." *Quarterly Journal of the Royal Meteorological Society* 132: 2127–2155. <https://doi.org/10.1256/qj.04.100>.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Castillo-Botón, C., D. Casillas-Pérez, C. Casanova-Mateo, et al. 2022. "Machine Learning Regression and Classification Methods for Fog Events Prediction." *Atmospheric Research* 272: 106157. <https://doi.org/10.1016/j.atmosres.2022.106157>.
- Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, 785–794. Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>.
- Cui, B., Z. Toth, Y. Zhu, and D. Hou. 2012. "Bias Correction for Global Ensemble Forecast." *Weather and Forecasting* 27: 396–410. <https://doi.org/10.1175/WAF-D-11-00011.1>.
- Dee, D. 2005. "Bias and Data Assimilation." *Quarterly Journal of the Royal Meteorological Society* 131: 3323–3343. <https://doi.org/10.1256/qj.05.137>.
- Fang, Y., Y. Wu, F. Wu, et al. 2023. "Short-Term Wind Speed Forecasting Bias Correction in the Hangzhou Area of China Based on a Machine Learning Model." *Atmospheric and Oceanic Science Letters* 16: 100339. <https://doi.org/10.1016/j.aosl.2023.100339>.
- Friedman, J. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29: 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
- Galanis, G., and M. Anadranistakis. 2002. "A One-Dimensional Kalman Filter for the Correction of Near Surface Temperature Forecasts." *Meteorological Applications* 9: 437–441. <https://doi.org/10.1017/S1350482702004061>.
- Gandin, L. S. 1963. *Objective Analysis of Meteorological Fields*, 242. Israel Program for Scientific Translations.
- Gao, Z., X. Shi, H. Wang, D.-Y. Yeung, W.-c. Woo, and W.-K. Wong. 2021. *Deep Learning and the Weather Forecasting Problem: Precipitation Nowcasting*, chap. 15, 218–239. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119646181.ch15>.
- Germann, U., and I. Zawadzki. 2002. "Scale-Dependence of the Predictability of Precipitation From Continental Radar Images. Part I: Description of the Methodology." *Monthly Weather Review* 130: 2859–2873. [https://doi.org/10.1175/1520-0493\(2002\)130%3C2859:SDOTPO%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130%3C2859:SDOTPO%3E2.0.CO;2).
- Germann, U., and I. Zawadzki. 2004. "Scale Dependence of the Predictability of Precipitation From Continental Radar Images. Part II: Probability Forecasts." *Journal of Applied Meteorology and Climatology* 43: 74–89. [https://doi.org/10.1175/1520-0450\(2004\)043%3C0074:SDOTPO%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043%3C0074:SDOTPO%3E2.0.CO;2).
- Glahn, H. R., and D. A. Lowry. 1972. "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting." *Journal of Applied Meteorology and Climatology* 11: 1203–1211. [https://doi.org/10.1175/1520-0450\(1972\)011%3C1203:TUOMOS%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011%3C1203:TUOMOS%3E2.0.CO;2).
- Haiden, T., A. Kann, C. Wittmann, G. Pistotnik, B. Bica, and C. Gruber. 2011. "The Integrated Nowcasting Through Comprehensive Analysis (INCA) System and Its Validation Over the Eastern Alpine Region." *Weather and Forecasting* 26: 166–183. <https://doi.org/10.1175/2010WAF2222451.1>.
- Head, T., M. Kumar, H. Nahrstaedt, G. Louppe, and I. Shcherbatyi. 2021. "scikit-optimize/scikit-optimize (v0.9.0)." <https://doi.org/10.5281/zenodo.1157319>.
- Hieta, L., M. Partio, M. Laine, et al. 2021. "Smartmet Nowcast: Rapidly Updating Nowcasting System at Finnish Meteorological Institute." *Meteorologische Zeitschrift* 30: 369–377. <https://doi.org/10.1127/metz/2021/1070>.
- Huang, L. X., G. A. Isaac, and G. Sheng. 2012. "Integrating NWP Forecasts and Observation Data to Improve Nowcasting Accuracy." *Weather and Forecasting* 27: 938–953. <https://doi.org/10.1175/WAF-D-11-00125.1>.
- Hyrkkänen, J., J. Kilpinen, P. Nurmi, J. Kaurola, and M. Brockmann. 2016. "Error Characteristics of Temperature Forecast in Finland for the Period

- 1979–2011 in Relation to Various Weather Patterns.” *Meteorological Applications* 23: 244–253. <https://doi.org/10.1002/met.1550>.
- James, P. M., B. K. Reichert, and D. Heizenreder. 2018. “NowCastMIX: Automatic Integrated Warnings for Severe Convection on Nowcasting Time Scales at the German Weather Service.” *Weather and Forecasting* 33: 1413–1433. <https://doi.org/10.1175/WAF-D-18-0038.1>.
- Korpinen, A., L. Hieta, and M. Partio. 2024. “Thunder Probability Nowcast.” In *EMS Annual Meeting 2024*. Copernicus Publications. <https://doi.org/10.5194/ems2024-471>.
- Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2004. *Applied Linear Statistical Models*. McGraw-Hill Irwin. https://users.stat.ufl.edu/~winner/sta4211/ALSM_5Ed_Kutner.pdf.
- Leinonen, J., U. Hamann, D. Nerini, U. Germann, and G. Franch. 2023. “Latent Diffusion Models for Generative Precipitation Nowcasting With Accurate Uncertainty Quantification.” <https://arxiv.org/abs/2304.12891>.
- Lundberg, S. M., and S.-I. Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777. Curran Associates. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- Mason, L., J. Baxter, P. Bartlett, and M. Frean. 1999. “Boosting Algorithms as Gradient Descent.” In *Advances in Neural Information Processing Systems 12 (NeurIPS 1999)*, 512–518. MIT Press. https://proceedings.neurips.cc/paper_files/paper/1999/file/96a93ba89a5b5c6c226e49b88973f46e-Paper.pdf.
- MetCoOp. 2024. “MetCoOp Nowcasting System (MNWC).” <https://metcoop.smhi.se/dokuwiki/nwp/metcoop/mnwc/start>.
- Müller, M., M. Homleid, K.-I. Ivarsson, et al. 2017. “AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model.” *Weather and Forecasting* 32: 609–627. <https://doi.org/10.1175/WAF-D-16-0099.1>.
- Nipen, T. N., H. H. Haugen, M. S. Ingstad, et al. 2024. “Regional Data-Driven Weather Modeling With a Global Stretched-Grid.” <https://arxiv.org/abs/2409.02891>.
- Oskarsson, J., T. Landelius, and F. Lindsten. 2023. “Graph-Based Neural Weather Prediction for Limited Area Modeling.” <https://arxiv.org/abs/2309.17370>.
- Partio, M., L. Hieta, and A. Kokkonen. 2024. “Cloudcasto – Total Cloud Cover Nowcasting With Machine Learning.” <https://arxiv.org/abs/2410.21329>.
- Pulkkinen, S., D. Nerini, A. A. Pérez Hortal, et al. 2019. “Pysteps: An Open-Source Python Library for Probabilistic Precipitation Nowcasting (v1.0).” *Geoscientific Model Development* 12: 4185–4219. <https://doi.org/10.5194/gmd-12-4185-2019>.
- Ravuri, S., K. Lenc, M. Willson, et al. 2021. “Skilful Precipitation Nowcasting Using Deep Generative Models of Radar.” *Nature* 597: 672–677. <https://doi.org/10.1038/s41586-021-03854-z>.
- Roberts, N., B. Ayliffe, G. Evans, et al. 2023. “IMPROVER: The New Probabilistic Postprocessing System at the met Office.” *Bulletin of the American Meteorological Society* 104: E680–E697. <https://doi.org/10.1175/BAMS-D-21-0273.1>.
- Seity, Y., P. Brousseau, S. Malardel, et al. 2011. “The AROME-France Convective-Scale Operational Model.” *Monthly Weather Review* 139: 976–991. <https://doi.org/10.1175/2010MWR3425.1>.
- Wang, Y., E. Coning, A. Harou, et al. 2017. *Guidelines for bNowcasting Techniques*. World Meteorological Organization.
- World Meteorological Organization. 2023. *Guide to Instruments and Methods of Observation (WMO-No. 8)*. 2023rd ed. World Meteorological Organization. <https://library.wmo.int/records/item/68695-guide-to-instruments-and-methods-of-observation>.
- Xu, W., L. Ning, and Y. Luo. 2020. “Wind Speed Forecast Based on Post-Processing of Numerical Weather Predictions Using a Gradient Boosting Decision Tree Algorithm.” *Atmosphere* 11: 738. <https://doi.org/10.3390/atmos11070738>.