THESIS

BIAS CORRECTION OF TEMPERATURE AND WIND FORECASTS FROM THE NOAA GLOBAL FORECAST SYSTEM (GFS) AND GLOBAL ENSEMBLE FORECAST SYSTEM (GEFS) USING MACHINE LEARNING

Submitted by

Qianya Zhu

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2025

Master's Committee:

Advisor: Haonan Chen

Anura Jayasumana
Yanlin Guo

Copyright by Qianya Zhu 2025

All Rights Reserved

ABSTRACT

BIAS CORRECTION OF TEMPERATURE AND WIND FORECASTS FROM THE NOAA GLOBAL FORECAST SYSTEM (GFS) AND GLOBAL ENSEMBLE FORECAST SYSTEM (GEFS) USING MACHINE LEARNING

Numerical Weather Prediction (NWP) models, such as the National Oceanic and Atmospheric Administration's (NOAA) Global Forecast System (GFS) and the Global Ensemble Forecast System (GEFS), are essential tools for modern weather forecasting. NWP models are the backbones of various applications in weather, climate, and water enterprises. However, due to model limitations, initialization errors, and discretizations of grids, large systematic biases still exist aside from advances in computing capabilities, spatial resolution, and physical parameterization of the models.

This study presents a machine learning-based bias correction framework that is driven two models: Extreme Gradient Boosting (XGBoost) and U-Net. The target variables include 2-meter temperature (2m-T), 10-meter and 100-meter wind speed (10m-WS and 100m-WS). ERA5 reanalysis data are used as the reference for evaluating and correcting forecast biases. The models are trained on both seasonal (summer and winter) and all-season datasets to account for seasonal variability in forecast errors. The GFS-based experiments focus on the CONUS region (24.5°N–49.5°N, 125.0°W–66.75°W), while the GEFS-based experiments cover Germany, a climatically diverse region.

Results show that U-Net significantly outperforms XGBoost in long-term forecasting, particularly beyond 120 hours, due to its capacity to learn complex spatial and temporal dependencies. In contrast, XGBoost exhibits superior performance in short-term forecasts (0–48 hours), especially when data are limited, offering efficient and interpretable bias correction. Seasonal training im-

proves temperature correction across both regions and models—especially during summer—while all-season models enhance generalization for wind speed forecasts.

Quantitative evaluation using root mean square error (RMSE) confirms that both models effectively reduce systematic forecast biases in GFS and GEFS outputs. This work has also indicated that without using sophisticated deep learning structures, a rather simple machine learning model may achieve decent performance when correcting weather forecast products.

ACKNOWLEDGEMENTS

Throughout my graduate studies, I have been aided and encouraged by many individuals, and without them, I would not have been able to complete this important academic journey. I want to acknowledge them all personally.

First, I would like to express my sincere gratitude to my advisor, Dr. Haonan Chen. Throughout the entire process of the thesis research, he has always given me patient and meticulous guidance, firm support, and valuable feedback and suggestions. His profound professional attainments, encouragement, and trust have greatly helped me and profoundly influenced the development of this work in many ways.

I would also like to thank my thesis committee members, Dr. Anura Jayasumana and Dr. Yanlin Guo, for generously dedicating their time and offering valuable suggestions and comments to improve my research.

My sincere appreciation also goes to the faculty and staff of the Department of Electrical and Computer Engineering at Colorado State University (CSU) for providing the academic resources, research platform, and supportive environment that enabled me to carry out this work.

I want to express my sincere gratitude to my family. My family has always provided me with endless support throughout my studies, so that I could walk fearlessly and complete it till the end.

Finally, I would like to express my special thanks to my colleagues in the Artificial Intelligence and Remote Sensing (AIRS) Laboratory at CSU. Thank you for your valuable insights, inspiring exchanges, and consistent support and help during my graduate studies. It is my great honor to grow and move forward with you.

DEDICATION

I would like to dedicate this thesis to all those who accompanied me.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background and Motivation

Reliable weather forecasting has a wide range of applications, such as forecasting meteorological disasters, utilization of renewable energy, agricultural production, and management of water resources [1–5]. Global weather forecasting is used to a large extent, but users and operation centers must perform detailed analyses at the regional level to make forecasting more helpful and precise.

Taking wind energy as an example case study, the U.S. Department of Energy has analyzed a scenario by 2030 in which 20% of energy demand is supplied by wind energy [6]. Not only will it require a significant increase in wind power capacity in the United States to do this, but it can also reduce carbon dioxide emissions by the energy sector by 925 million tons per year [6]. However, the uncertainty and intermittency of wind power create significant problems of dispatching power supply and running wind farms. For these reasons, it is critical to increase the forecasting accuracy of wind energy. A number of studies have investigated the short-term prediction of wind speed and wind direction, especially for changes in the next few hours or days, to support grid dispatch and wind farm operation [7].

In addition, temperature prediction plays a key role in solar power generation management. Research by the US National Renewable Energy Laboratory (NREL) shows that in hotter climates, the loss of performance of photovoltaic systems is twice that in cold climates [8]. Accurate temperature predictions can help operators better predict fluctuations in power generation efficiency caused by temperature changes, so optimizing the management of solar power generation. Therefore, accurate forecasting of temperature and wind speed is very important in increasing efficiency in using renewable energy and enabling large-scale application of clean energy.

Although numerical weather prediction (NWP) models have wide applications in forecasting weather conditions around the world, their precision is still limited by systematic errors [9]. Systematic errors affect the reliability of weather forecasting in real-world applications. The bias of the NWP model is largely caused by the numerical discretizations of motion equations, parameterization uncertainty of sub-grid-scale phenomena, and boundary conditions [10, 11]. Although such errors may be small at initial prediction times, as model integration increases, they will nonlinearly interact with other random and systematic errors and ultimately degrade the model's predictive capability. Therefore, effective bias correction and post-processing techniques must be used to improve NWP models' prediction accuracy.

1.2 Objectives of This Thesis

The primary objective of this study is to evaluate and compare the effectiveness of traditional machine learning and deep learning approaches in correcting forecast biases in numerical weather prediction (NWP) models, particularly for near-surface temperature and wind speed variables.

Specifically, the goals of this research are

1. To implement and apply a tree-based machine learning model (XGBoost) for bias correction of 2-meter temperature, 10-meter wind speed, and 100-meter wind speed forecasts derived from the Global Forecast System (GFS), and 10-meter wind speed and 100-meter wind speed forecasts from the Global Ensemble Forecast System (GEFS).

2. To develop and train a deep learning model (U-Net) designed to learn spatial patterns in meteorological fields and evaluate its performance in correcting the same forecast variables.

3. To systematically compare the performance of XGBoost and U-Net across different meteorological variables and forecast lead times, using ERA5 reanalysis data as ground truth.

4. To assess the models' capability to generalize across different seasons and geographical regions within the study domain.

5. To analyze the trade-offs between accuracy, computational cost, and interpretability of the two approaches and provide guidance on their applicability in real-world post-processing systems.

1.3 Thesis Outline

The structure of the remaining chapters of this thesis is as follows.

- **Chapter 2** presents a background review on numerical weather prediction (NWP) models and bias correction techniques, covering both traditional methods such as Model Output Statistics (MOS) and Kalman filters (KF), and advanced approaches including tree-based models (e.g., XGBoost) and deep learning models (e.g., U-Net).

- **Chapter 3** describes the study domains and datasets used in this research. It outlines the geographical characteristics of the Continental United States (CONUS) and Germany, introduces the forecast datasets (GFS and GEFS), and details the reference dataset (ERA5), forecast variables, and data preprocessing steps.

- **Chapter 4** details the methodology employed in this study, including the design and implementation of the XGBoost and U-Net models, the training strategies, and feature extraction processes.

- **Chapter 5** discusses the results and performance of the model across different seasons, forecast lead times, and meteorological variables, which also includes spatial and quantitative evaluations and compares the effectiveness of XGBoost and U-Net in correcting forecast biases.

- **Chapter 6** presents future directions for expanding the current analysis, including feature extension and advanced modeling architectures.

Chapter 2

Brief Review of Numerical Weather Prediction (NWP) and Postprocessing Techniques

2.1 Overview of Numerical Weather Prediction (NWP)

Numerical Weather Prediction (NWP) is a fundamental technique in modern meteorology that utilizes numerical methods to solve the governing equations of atmospheric physics in order to forecast future states of the atmosphere. These models are based on the fundamental conservation laws of mass, momentum, and energy, typically represented as a system of nonlinear partial differential equations. Due to the complexity of these equations and the vast spatial domain involved, NWP models are discretized on a three-dimensional grid and integrated forward in time using high-performance computing resources.

The accuracy of NWP models is inherently dependent on the quality of initial conditions, which are derived through sophisticated data assimilation techniques that combine diverse observational sources, including satellite radiances, ground-based measurements, radiosondes, and radar. Model outputs include forecasts of key meteorological variables such as temperature, wind, pressure, and precipitation at various lead times.

NWP models can be broadly categorized into global models, which simulate atmospheric processes on a planetary scale (e.g., GFS, ECMWF), and regional models, which operate at higher resolutions over limited spatial domains (e.g., WRF, NAM). In addition, ensemble prediction systems (EPS), such as GEFS or ECMWF-EPS, generate multiple forecasts using slightly perturbed initial states to characterize forecast uncertainty and improve probabilistic decision-making.

Although numerical weather prediction (NWP) models have seen considerable improvements in recent decades, particularly in spatial resolution, physical parameterization, and data assimilation, systematic errors in forecasts are still common. These errors usually arise from several

reasons, such as limited observational coverage, simplifications in representing physical processes like cloud formation or turbulence, and sensitivity to initial and boundary conditions. The resulting biases are often consistent over time and structured in space, which makes raw model output less reliable for applications that demand high accuracy, such as energy forecasting or hydrological modeling. In order to improve the accuracy and reliability of forecasts, researchers generally use post-processing techniques, including both traditional statistical techniques and advanced machine learning methods, to adjust and reduce model biases.

2.2 Traditional Bias Correction Methods

2.2.1 Model Output Statistics (MOS)

Model Output Statistics (MOS) is a classical statistical postprocessing technique that aims to correct systematic biases in numerical weather prediction (NWP) model outputs. Introduced by Glahn and Lowry in 1972, MOS established a statistical relationship between observed weather variables and corresponding outputs from the NWP models, using historical forecast-observation pairs to develop regression-based equations [12]. These equations are then applied to future forecasts to improve their accuracy.

The key idea of MOS is to leverage the strengths of NWP models in capturing synoptic-scale atmospheric patterns while compensating for their deficiencies, such as resolution limitations and simplifications in physical parameterizations. Typically, MOS involves linear or logistic regression models where predictors are direct model outputs (e.g., temperature, wind, humidity) and predictands are observed surface weather variables at specific stations.

One of the strengths of MOS is that it is able to incorporate long-term statistical relationships, thereby minimizing smooth transient model errors and enhancing local forecast skill. Additionally, MOS models are not too computationally expensive and can be easily updated using new data.

However, the MOS method still has several practical limitations in its application. First, it is highly dependent on large-scale and consistent historical observational data, which significantly reduces its effectiveness in areas where data is sparse or the observation system is unstable [12].

Second, MOS assumes that the statistical relationship between the predicted variables and the observed values remains stable. However, in climate change, the evolution of long-term climate trends or the update of the physical parameterization scheme of numerical models may undermine this assumption, resulting in a decline in model performance [13]. These problems limit the adaptability of MOS in dynamic environments and highlight the importance of introducing more flexible and nonlinear methods (such as machine learning) for bias correction.

In recent years, MOS has often been used as a baseline against which more advanced machine learning-based postprocessing techniques are compared. Despite its simplicity, MOS continues to be widely used in operational forecasting, particularly in producing calibrated, location-specific forecasts.

2.2.2 Kalman Filtering (KF)

Kalman Filtering (KF) [14] is a recursive statistical technique originally developed for linear dynamic systems, which has been widely adopted in numerical weather prediction (NWP) for real-time bias correction and data assimilation. In the context of weather forecasting, KF is primarily used to adjust model outputs by continuously updating predictions based on recent observations, effectively reducing short-term forecast errors.

The basic principle of the Kalman Filter includes two main steps: prediction and update. In the prediction stage, the filter makes a priori estimates of the state of the current variable (such as surface temperature or wind speed) based on the state transition model of the system and the estimate of the previous moment; in the update stage, the predicted value is corrected in combination with the newly obtained observation data. By covering both the uncertainty of observation data and that of model prediction, the Kalman filter gives the optimal weighted estimation of the state variable such that the mean square error (MSE) is minimized in a statistical sense.

Due to its recursive computational structure, the Kalman filter is particularly suitable for real-time or quasi-real-time forecast bias correction problems. Numerical weather prediction (NWP) can be used to update model outputs hour by hour, such as dynamically adjusting short-term fore-

casts of temperature, wind speed, or pressure. Compared with static statistical models, the Kalman filter has a stronger adaptive ability and can continuously correct the forecast results as the model error changes over time [15, 16].

However, the classical Kalman Filter is limited to linear systems and assumes that all errors follow a Gaussian distribution. These assumptions may not hold in highly nonlinear atmospheric systems or in the presence of non-Gaussian error characteristics. To overcome this problem, researchers have proposed variant algorithms such as the extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF), which can adapt to more complex dynamic systems.

One of the strengths of Kalman Filtering is its ability to adapt to changing error characteristics over time. In contrast to static methods such as Model Output Statistics (MOS), KF dynamically adjusts its bias correction coefficients as additional data are made available. This makes it especially suitable for correcting short-term forecasts and capturing time-varying biases, such as diurnal or seasonal changes.

In operational meteorological settings, KF is often implemented in a univariate form, where each variable at each location is corrected independently. More sophisticated variants, such as the EnKF, extend the framework to multivariate systems and are widely used in modern data assimilation techniques. Despite its limitations, Kalman Filtering remains a foundational method in operational forecasting, particularly in applications that require frequent updates and rapid bias correction.

2.3 Machine Learning-based Bias Correction Methods

2.3.1 Gradient Boosting Decision Trees (GBDT) and XGBoost

Gradient Boosting Decision Trees (GBDT) represent a robust ensemble learning method that constructs predictive models in a stage-wise manner by combining multiple weak learners, typically shallow decision trees. Each iteration of the model fits a tree to the residual errors of the previous model, gradually improving performance.

XGBoost is an optimized implementation of GBDT introduced by Chen and Guestrin [17], designed for speed and performance. It incorporates regularization techniques to prevent overfitting, handles missing values gracefully, and supports parallel and distributed computing. In the context of weather forecast bias correction, XGBoost has demonstrated success in capturing nonlinear relationships between numerical weather prediction (NWP) forecasts and observed values, while maintaining interpretability and high computational efficiency [18].

2.3.2 LightGBM

Light Gradient Boosting Machine (LightGBM) is a highly efficient implementation of GBDT, developed by Microsoft [19]. It uses histogram-based algorithms and leaf-wise tree growth to improve training speed and reduce memory usage. LightGBM also supports parallel and GPU-accelerated training, making it suitable for large-scale meteorological datasets. It has been used in post-processing of weather forecasts due to its ability to handle high-dimensional data and its built-in support for handling missing values and categorical features.

2.3.3 Deep Neural Networks (DNN)

Deep Neural Networks (DNNs) are composed of multiple hidden layers that enable learning complex, nonlinear mappings from inputs to outputs. They are trained using backpropagation and gradient-based optimizers. In bias correction tasks, DNNs can integrate various types of features such as spatial, temporal, and categorical variables, learning hierarchical representations that improve forecast accuracy [20].

Despite their power, DNNs require large volumes of labeled data, and they are sensitive to hyperparameters. They are also often regarded as black-box models due to limited interpretability. Nevertheless, DNNs serve as the foundation for more advanced models such as Convolutional Neural Networks (CNNs).

2.3.4 Convolutional Neural Networks (CNN)

CNNs are specifically designed for grid-type data structures, making them ideal for weather data arranged on spatial grids. CNNs use convolutional filters to extract spatial hierarchies of features and thus are capable of learning local and global spatial relationships. This makes them well suited for applications like bias correction, where spatial consistency is important [9].

CNNs have been applied to correct terrain-induced biases and to improve the representation of regional anomalies in gridded weather forecasts. Although computationally intensive, CNNs offer better spatial modeling capability compared to traditional DNNs.

2.3.5 U-Net Architecture

U-Net is a CNN-based architecture originally developed for biomedical image segmentation [21], but has shown promise in meteorological applications. Its U-shaped structure consists of a contracting path (encoder) and an expansive path (decoder), with skip connections between corresponding layers. These skip connections allow for better localization and reconstruction of spatial information.

In NWP bias correction, U-Net can model both fine-grained local details and broader spatial patterns. Its ability to handle multi-channel inputs makes it especially suitable for multivariate gridded datasets, such as temperature and wind fields across forecast lead times.

2.3.6 Comparison and Summary

Advanced machine learning techniques significantly improve upon classical statistical models as they are capable of learning nonlinear relationships and spatial correlations of forecast errors. Tree-based models like XGBoost are fast, interpretable, and perform well with limited data. Deep learning models, especially CNN-based architectures like U-Net, provide strong capabilities in spatial modeling but require substantial computational resources and training data.

While tree-based models are well suited for short-term, tabular data-driven correction, deep learning models such as U-Net excel in learning complex, structured errors in spatial forecasts.

This study aims to systematically compare these two approaches in correcting bias in GFS and GEFS forecasts, focusing on their accuracy, generalization capability, and operational applicability.

Chapter 3

Study Domains and Datasets

3.1 Introduction

This chapter provides a detailed overview of the study domains and datasets utilized in this research. Two geographically and climatologically distinct regions—the Continental United States (CONUS) and Germany—are selected to facilitate a comprehensive evaluation of the effectiveness and robustness of various bias correction methodologies across diverse environmental conditions. The primary datasets employed include forecast outputs from the Global Forecast System (GFS) and the Global Ensemble Forecast System (GEFS), along with the ERA5 reanalysis dataset, which serves as the reference benchmark for both model evaluation and training. In addition, this chapter describes the key forecast variables under investigation: 2-meter temperature (2m-T), 10-meter wind speed (10m-WS), and 100-meter wind speed (100m-WS). These variables are chosen due to their critical importance in a wide range of meteorological applications, including renewable energy forecasting, agricultural planning, and risk assessment for extreme weather events.

3.2 Study Domains

The study is conducted over two distinct geographical regions: the Continental United States (CONUS) and Germany. These two domains were selected to capture a wide range of climatic, topographic, and meteorological conditions [22], thereby enabling robust evaluation of bias correction methods under diverse environmental settings.

3.2.1 Continental United States (CONUS)

The first study domain is the Continental United States (CONUS), covering the region from $24.5°\mathrm{N}$ to $49.5°\mathrm{N}$ in latitude and from $-125.0°\mathrm{W}$ to $-66.75°\mathrm{W}$ in longitude. This spatial ex-

tent encompasses an area of approximately 2800 km $\times$ 5200 km, representing one of the most geographically and climatologically diverse regions in the world [22].

The climatic diversity of CONUS is driven by its large latitudinal extent and complex terrain. The southwestern region exhibits arid and semi-arid conditions, while the southeastern region experiences a humid subtropical climate [23]. The northeastern U.S. is dominated by a humid continental climate, whereas coastal areas are strongly influenced by maritime effects. The central Great Plains, known for their flat topography, are prone to convective storms and tornadoes, while mountainous areas are characterized by complex orographic-induced weather phenomena [24].

CONUS is a critical region for operational weather forecasting and climate modeling [25], given its exposure to rapidly changing weather systems, pronounced seasonal variability, and frequent extreme weather events. In addition, the region has important social and economic impacts in key sectors such as agriculture, energy, and disaster response [26], which further highlights the importance of improving the accuracy of weather forecasts through effective bias correction methods.

3.2.2 Germany

The second study domain is Germany, located in Central Europe, spanning from $47.2^{\circ}\mathrm{N}$ to $55.2^{\circ}\mathrm{N}$ in latitude and from $5.6^{\circ}\mathrm{E}$ to $15.2^{\circ}\mathrm{E}$ in longitude. This domain covers an area of approximately 888 km $\times$ 672 km and features varied topography, including coastal plains in the north, forested uplands in the center, and alpine terrain in the south [27, 28].

Germany has a variety of climate types, which are derived from its large latitudinal span and are significantly affected by its complex terrain. The northern region mainly presents the characteristics of a temperate oceanic climate, which is significantly regulated by the North Sea and the Atlantic Ocean, with relatively mild winters and cool summers. The central region gradually transitions to a temperate continental climate, with the typical characteristics of warm summers and cold winters. The southern region, especially the area near the Alps, is significantly affected by the

mountainous terrain and presents the characteristics of an Alpine climate, with complex weather phenomena and a high degree of locality [29, 30].

This domain is particularly well-suited for the analysis of wind-related variables due to its meteorological diversity, topographic complexity, and the availability of dense, high-quality observational and reanalysis datasets [9, 31, 32].

3.3 Data Description

3.3.1 GFS: Forecast Data for CONUS

The Global Forecast System (GFS) is a global numerical weather prediction model developed by the National Centers for Environmental Prediction (NCEP) [33]. Forecasts are issued four times daily at 00, 06, 12, and 18 UTC, with outputs provided at 3-hour intervals for the first 240 hours and at 12-hour intervals from 240 to 384 hours, resulting in a total of 93 forecast times [25, 34, 35].

This study uses GFS data over the CONUS domain from January 1, 2021, to September 30, 2024, at a spatial resolution of 0.25°. The selected time range captures seasonal and interannual variability and provides a robust basis for training and evaluation of the correction models.

3.3.2 GEFS: Forecast Data for Germany

For Germany, the Global Ensemble Forecast System (GEFS) is used. GEFS consists of 31 ensemble members (30 perturbed, 1 control) to represent forecast uncertainty. Forecasts are initialized at 00, 06, 12, and 18 UTC and extend up to 384 hours, with 3-hourly resolution for the first 240 hours [36–40].

This study uses GEFS forecasts from January 1, 2021, to December 31, 2023, with August and December 2022 selected as independent test periods.

3.3.3 Reference Dataset: ERA5 Reanalysis

ERA5 (ECMWF Reanalysis v5) is a global reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [41]. It integrates diverse observations using a

four-dimensional variational (4D-Var) assimilation scheme and provides hourly estimates at 0.25° resolution [42]. ERA5 includes key variables such as 2m temperature and wind components, which are used as the observational benchmark in this study.

3.4 Forecast Variables

3.4.1 2-meter Temperature (2m-T)

2m-T is directly output from GFS and GEFS. It is influenced by surface processes, boundary layer dynamics, and land-atmosphere interactions, and often exhibits systematic biases that must be corrected.

3.4.2 10-meter Wind Speed (10m-WS)

10m-WS is derived from the horizontal wind components at 10 meters:

$$10\text{WS} = \sqrt{u_{10}^2 + v_{10}^2} \tag{3.1}$$

3.4.3 100-meter Wind Speed (100m-WS)

100m-WS is also derived from wind components at 100 meters:

$$100\text{WS} = \sqrt{u_{100}^2 + v_{100}^2} \tag{3.2}$$

These wind speeds are key for wind energy applications but are subject to biases due to model resolution, interpolation methods, and vertical profile representation.

3.5 Data Preprocessing

To align the model outputs with ERA5:

- ERA5 remains at 0.25° for CONUS (matching GFS).
- ERA5 is interpolated to 0.4° for Germany (matching GEFS).

Wind speeds are computed from u and v components (3.1, 3.2) for all datasets. Forecast and reanalysis datasets are spatiotemporally aligned to enable consistent training and evaluation.

3.6 Conclusion

This chapter has described the study regions, datasets, forecast variables, and preprocessing steps. The combination of deterministic (GFS) and ensemble (GEFS) models with ERA5 reanalysis across two climatically diverse domains enables a robust framework for evaluating bias correction methods in subsequent chapters.

Chapter 4

Bias Correction Methodologies

This chapter presents the methodology developed for correcting biases in numerical weather prediction (NWP) outputs using machine learning. To illustrate the general framework, the correction of 2m-T forecasts is used as an example, as the procedures applied to other variables (e.g., wind speed) follow a similar structure.

Given a forecast $P_{t+\Delta_t}$ (e.g., 120 hours) issued at time t, and the corresponding observation $y_{t+\Delta_t}$, the objective is to learn a mapping function f such that:

$$f(P_{t+\Delta_t}) \approx y_{t+\Delta_t}$$

In practice, the observation y_t at the issue time is also included as an input to the model, based on the assumption that the current state of the atmosphere contributes to the evolution of future conditions. Therefore, the input to the model becomes a combination of $P_{t+\Delta_t}$ and y_t, and the functional relationship can be formulated as:

$$f(P_{t+\Delta_t}, y_t) \rightarrow y_{t+\Delta_t}$$

Figure 4.1 illustrates the overall workflow of the proposed bias correction method. During the training phase, historical 120-h GFS or GEFS forecasts of the 2m-T and the observation data (ERA5) at issue time t are used to train supervised learning models. The observation data (ERA5) at $t+120$ h is used as the ground truth during training. Both XGBoost (Extreme Gradient Boosting) and U-Net are implemented for comparative analysis. Once trained, the models are applied to new forecast data to generate bias-corrected outputs.

The remainder of this chapter is organized as follows: Section 4.1 introduces the XGBoost-based method as a baseline. Section 4.2 details the U-Net architecture and its application to spatial bias correction.

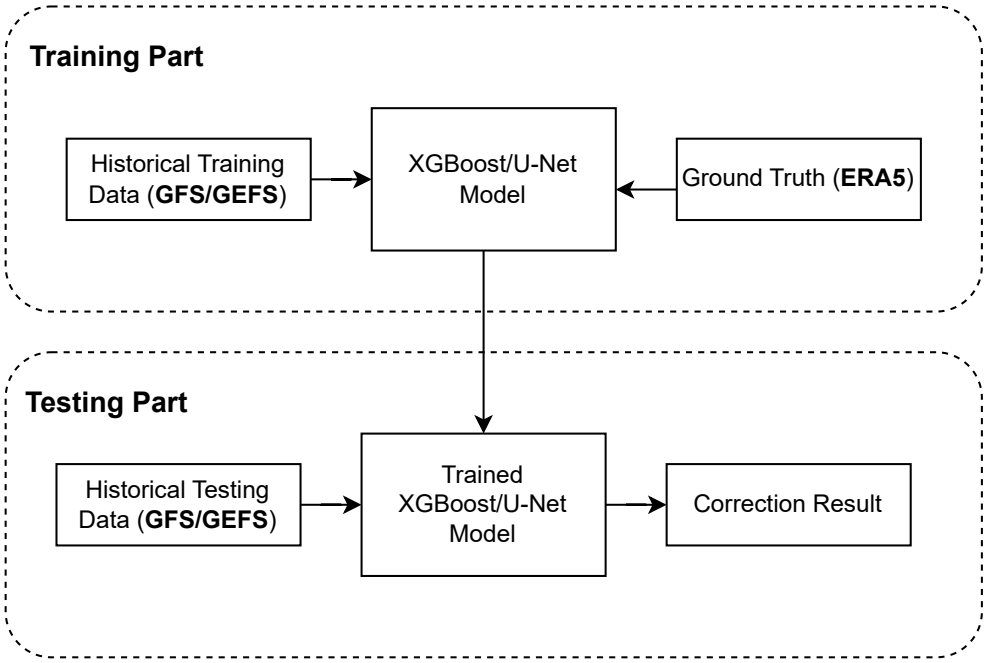

Figure 4.1: Flowchart of the deep learning correction method. Training: historical 120-h GFS/GEFS forecasts and the observation data (ERA5) at time t as inputs to the XGBoost/U-Net model. The ERA5 data at time $t + 120$ h is used as the ground truth. Testing: the trained XGBoost/U-Net model is applied to a testing dataset to evaluate its performance.

4.1 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a widely recognized machine learning algorithm proposed by Tianqi Chen in 2016 during his doctoral studies at the University of Washington [17]. Since its introduction, XGBoost has achieved remarkable success in numerous data science competitions and real-world applications across various domains. As a type of ensemble learning method, XGBoost is fundamentally based on the Gradient Boosting Decision Tree (GBDT) framework. It introduces multiple optimizations and enhancements over traditional GBDT, particularly improving performance when handling unstructured data, and is capable of solving both classification and regression problems with high efficiency.

In this study, the XGBoost model was trained using mean squared error (MSE) as the loss function, with a maximum tree depth of 6, 500 boosting rounds, and a learning rate of 0.015.

Similar to the U-Net model, XGBoost was applied to correct temperature and wind speed forecasts using paired GFS-ERA5 or GEFS-ERA5.

Unlike standard GBDT, which only utilizes first-order gradients in the loss function, XGBoost applies a second-order Taylor expansion to incorporate both the first and second derivatives of the loss. Additionally, it introduces a regularization term into the objective function to penalize model complexity, which helps mitigate overfitting and improves generalization performance. For decision tree construction, XGBoost adopts an exact greedy algorithm. It first performs a presorting of feature values and then enumerates all possible split points for each feature to identify the optimal split that minimizes the objective function. This allows for precise and efficient partitioning of the data. The objective function used in XGBoost is defined as follows:

$$Obj_{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + con \tag{4.1}$$

Here, $L(a, b) = (a - b)^2$ represents the squared loss function, con denotes a constant term, and y_i refers to the true value of the i-th sample, while $\hat{y}_i^t$ denotes the predicted value for the i-th sample after the t-th iteration of the model. The term $\Omega(f_t)$ is the regularization component, which is introduced to enhance the generalization ability of the model and prevent overfitting. It is defined as follows:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{T} w_j^2 \tag{4.2}$$

In this formulation, T represents the number of leaf nodes, and w_j denotes the weight of the j-th leaf. The parameter γ and λ are manually tunable regularization coefficients. Larger values of γ and λ lead to simpler tree structures by penalizing model complexity. By performing a second-order Taylor expansion of the objective function $Obj_{(t)}$ and setting its derivative to zero, the optimal leaf weight is $w_i = \frac{-G_j}{H_j+\lambda}$, then substituting this expression back into the objective function yields the simplified form:

$$Obj_{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T + c \tag{4.3}$$

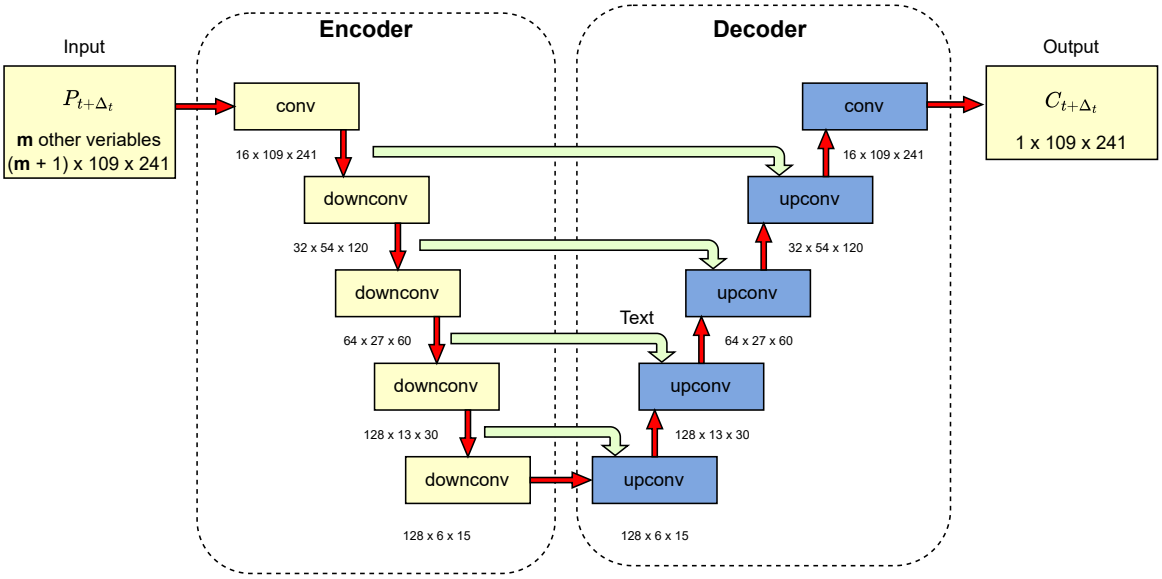

Figure 4.2: The U-Net architecture has a series of convolutional modules for downsampling (downconv) on the left, which facilitates the encoding stage, and an upsampling convolutional module (upconv) on the right, which is used for the decoding stage. Green arrows indicate skip connections that preserve fine details from earlier layers, thereby improving prediction accuracy. Data flows through the U-Net starting in the upper left corner, moving downward, and then rising on the right.

It can be seen that $Obj_{(t)}$ represents the sum of the scores across all leaf nodes. The next step involves enumerating all possible split points and selecting the optimal one. The gain from a potential split, also known as the split gain, is defined as:

$$Gain = \frac{1}{2}[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}] - \lambda \tag{4.4}$$

A larger gain indicates a greater reduction in the loss function after the split, leading to faster model convergence.

4.2 U-Net

Convolutional Neural Networks (CNNs) have achieved remarkable success across a wide range of domains, including applications such as image segmentation, style transfer, and video prediction—many of which fall under the category of image-to-image translation tasks. In these applications, CNN models are capable of learning complex mappings between the pixel values of input

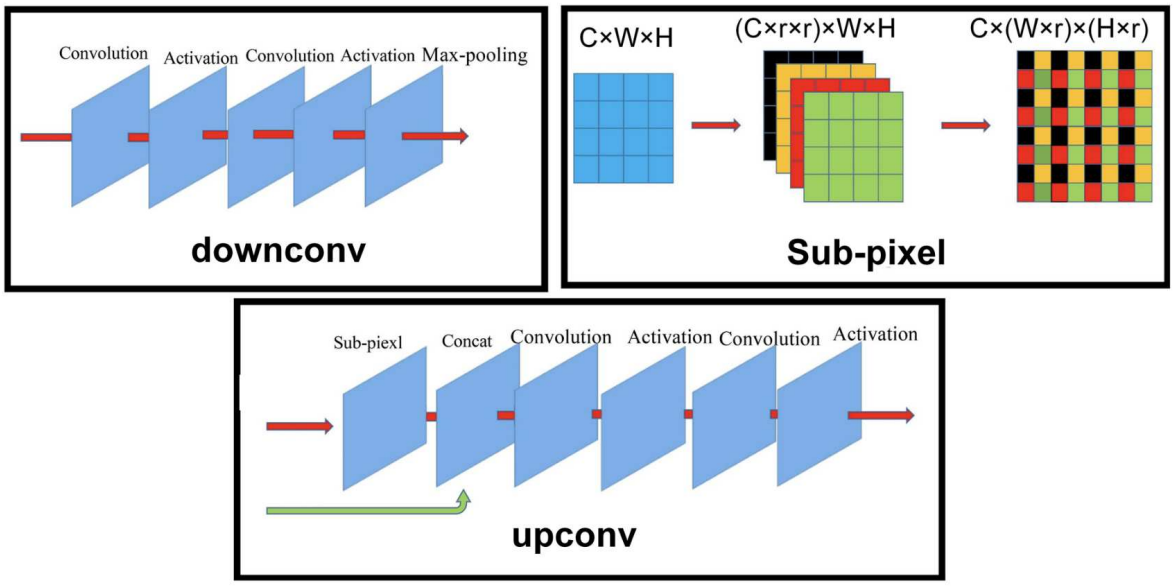

Figure 4.3: Illustration of the detailed structures of downconv, upconv, and sub-pixel. The concat layer appends the channels from the skip connection to channels produced by the sub-pixel operation.

and target images. In this study, the data is gridded fields, where each grid point can be interpreted analogously to a pixel. This allows the bias correction problem at the grid-point level to be reformulated as an image-to-image conversion problem, which is naturally suited to CNN-based architectures.

In this study, the widely used U-Net architecture, which was originally developed for image segmentation tasks. It is adopted to model the grid-point bias correction problem. The resulting model is referred to as Trained U-Net [21]. The network architecture of the trained U-Net model is illustrated in Fig. 4.2.

As shown in Fig. 4.2, taking GFS forecast dataset 2-meter temperature (2m-T) as an example, the input to the network consists of the observed 2m-T dataset at forecast initialization time t, denoted as R_t, and the 120-hour forecast dataset $P_{t+\Delta T}$. Similar to other variants of the U-Net architecture, the trained U-Net model is composed of multiple stacked modules that can be categorized into two main components: the encoder and the decoder.

The encoder part consists of a sequence of downconv modules, which perform spatial downsampling and feature extraction. These modules capture spatial correlations in the data, progres-

sively transforming low-level features into high-level representations by compressing and abstracting the input fields. The decoder part, in contrast, consists of a sequence of upconv modules that reconstruct the compressed feature representations back to the original spatial resolution.

The green arrows in Fig. 4.2 are skip connections, which allow features extracted during the encoding steps to directly inform and guide the decoding steps. This architecture allows for the reuse of spatial features in previous layers and thus enhancing reconstruction accuracy.

The data dimensions in Fig. 4.2 are expressed in the format C x W x H, in which C is the number of features, and W and H denote the spatial width and height, respectively. The internal structure of each individual model is shown in detail in Fig. 4.3.

As shown in Fig. 4.3, the downconv module consists of a convolutional layer followed by an activation layer, then another convolutional layer and activation layer, and finally a max-pooling layer. The stride of the max-pooling operation is set to 2. The spatial dimensions of the input remain unchanged after each convolutional operation, but are reduced by half after max pooling.

The use of max pooling is particularly well-suited to the characteristics of numerical weather prediction (NWP) model outputs. In many operational models, such as GFS, low-resolution forecast fields are often interpolated to produce higher-resolution outputs. Taking 2-meter temperature (2m-T) as an example, in some regions the values at high-resolution grid points may be derived through interpolation from coarser-resolution data. In such cases, max pooling can help capture the most salient features of the original signal prior to interpolation, thereby enabling the network to learn a more realistic and representative description of the underlying physical patterns.

The upconv module, as illustrated in Fig. 4.3, consists of a sub-pixel convolution layer, a concatenation (Concat) layer, followed by two convolutional layers and two activation layers. The primary function of the upconv module is to reconstruct high-resolution feature maps from low-resolution ones. Various strategies exist for solving the problem of mapping from low to high resolution. Common approaches include fixed upsampling kernels, such as bilinear or nearest-neighbor interpolation, and learnable techniques such as deconvolution (also known as transposed convolution) and unpooling, which are often used for network visualization in deep learning [43].

Traditional interpolation methods rely on static, non-trainable kernels and do not adapt during training. Although deconvolution and unpooling offer trainable alternatives, they may introduce artifacts or noise due to manual design choices. In this study, we adopt the sub-pixel convolution module proposed by Shi et al. [44] for upsampling, which effectively learns a data-driven interpolation function as part of the model training process.

As shown in Fig. 4.3, the input feature map to the sub-pixel layer has dimensions of C x W x H. After a convolutional operation, it is transformed into a feature map of size (C x r x r) x W x H, where r is the upsampling factor. These feature maps are then rearranged using a pixel-shuffling operation into a new feature map of size C x (W x r) x (H x r), thereby achieving spatial resolution enhancement by a factor of r in both width and height.

The Concat layer is a commonly used operation in deep neural networks for feature fusion. There are two widely adopted strategies for combining feature maps: one is the concatenation approach used in the trained U-Net architecture, and the other is the element-wise addition strategy employed in models such as ResNet [45].

Assuming that the two input feature maps entering the Concat layer, as illustrated in Fig. 4.3, have dimensions C_1 x W x H and C_2 x W x H, the output after concatenation becomes a single tensor with dimensions (C_1 + C_1) x W x H. This operation stacks the feature maps along the channel dimension without performing any arithmetic computation. Therefore, the Concat layer serves purely as a structural mechanism to combine information from different sources.

In contrast, the element-wise addition used in ResNet directly adds two feature maps of identical dimensions (e.g., C_1 x W x H and C_2 x W x H) on a pixel-by-pixel basis. This approach is primarily designed to facilitate gradient flow in deep networks, thereby mitigating the vanishing gradient problem and enabling more effective training of very deep architectures.

4.3 Conclusion

This chapter has presented two machine learning methods. One is Extreme Gradient Boosting (XGBoost), and the other is U-Net, which are for the postprocessing and bias correction of numer-

ical weather prediction (NWP) outputs. Each model offers distinct methodological advantages and is tailored to address different aspects of the bias correction task, both in terms of data structure and model complexity.

XGBoost, a highly efficient and interpretable ensemble learning algorithm, is grounded in the Gradient Boosting Decision Tree (GBDT) framework. By including second-order optimization, regularization methods, and an exact greedy algorithm-based tree construction, XGBoost provides a robust solution for structured, tabular data. The ability of XGBoost to deal with small datasets makes it well-suited for real-time correction. The formalism of its objective function and gain-based criterion of split ensures both predictive accuracy and computational efficiency.

By comparison, U-Net provides a powerful deep learning-based approach to learning complex spatial patterns. By solving bias correction as an image-to-image problem, U-Net uses encoder and decoder modules to learn spatial patterns between forecast datasets and observations. The use of sub-pixel convolution for learnable upsampling, combined with skip connections and channel-wise feature concatenation, allows the model to retain fine-grained information while reconstructing high-resolution corrections. This architecture is particularly effective in capturing terrain-related and spatially localized biases commonly found in high-resolution NWP data.

Together, these two models exemplify distinct but complementary paradigms in machine learning: XGBoost excels in lightweight, interpretable modeling with fast inference, while U-Net offers a data-driven, end-to-end learning approach capable of capturing spatiotemporal structure.

Chapter 5

Results and Discussion

5.1 Introduction

This chapter presents the evaluation results of bias correction methods applied to near-surface and upper-level meteorological forecasts, with a focus on comparing two representative approaches: the traditional machine learning algorithm Extreme Gradient Boosting (XGBoost) and the deep learning architecture U-Net, known for its strong spatial modeling capabilities. For the Global Forecast System (GFS), the evaluation covers 2-meter temperature (2m-T), 10-meter wind speed (10m-WS), and 100-meter wind speed (100m-WS). For the Global Ensemble Forecast System (GEFS), the focus is on correcting systematic biases in the 10-meter and 100-meter wind speed forecasts.

To comprehensively assess the performance of each method, the study combines spatial visualization comparisons with domain-averaged Root Mean Square Error (RMSE) metrics. The GFS evaluation includes two typical seasonal scenarios—summer and winter—and compares the correction effectiveness of both all-season and season-specific models. Similarly, the GEFS evaluation covers both seasons and uses season-specific models for bias correction. By comparing the spatial distributions of the original and corrected forecasts, residual maps with respect to ERA5, and the temporal evolution of RMSE, this study conducts an in-depth analysis of model performance across seasons, forecast lead times, and meteorological variables.

The analysis is organized by forecast source (GFS and GEFS), with a detailed examination of the performance of XGBoost and U-Net under various weather conditions. Finally, by synthesizing the spatial and quantitative results, the chapter explores the application scenarios, performance advantages, and potential for broader deployment of the two correction methods in different climatic environments.

5.2 Training Methods

This study develops two machine learning models (XGBoost and U-Net) to correct systematic biases in numerical weather prediction (NWP) data from GFS and GEFS. The target variables include 2-meter temperature (2m-T), 10-meter wind speed (10m-WS), and 100-meter wind speed (100m-WS). For model training, different forecast hour ranges are selected based on the characteristics and operational significance of each variable: 2m-T forecasts from 0-120 hours are used, while 10m-WS and 100m-WS forecasts from 0-48 hours are selected. These ranges are applied consistently to the input data for both XGBoost and U-Net models, ensuring the models are optimized within time windows most relevant to their respective variables.

For the XGBoost model, to improve the accuracy of pointwise predictions at each grid cell, spatial dependencies are captured by incorporating a 7×7 grid patch centered on the target point (i.e., a total of 49 neighboring grid cells). This approach enables the model to effectively learn local spatial dynamics in temperature and wind fields. Different training strategies are applied depending on the variable and region to enhance model performance.

For the U-Net model, the input data is treated as image-like fields, and the bias correction task is formulated as an image-to-image translation problem. Taking 2-meter temperature as an example, the input to the model consists of the ERA5 reanalysis at forecast initialization time t (denoted as R_t) and the corresponding 120-hour forecast field from GFS or GEFS ($P_{t+\Delta T}$). The U-Net architecture includes two main components: an encoder and a decoder. The encoder uses downsampling modules to extract high-level spatial features, while the decoder reconstructs the high-resolution output using sub-pixel convolution and skip connections to retain spatial detail. This structure is well-suited for fine-grained bias correction of gridded forecast data.

For the GFS model over the Continental United States (CONUS), the variables of interest include 2m-T, 10m-WS, and 100m-WS. Both the XGBoost and U-Net models adopt three training strategies:

- Summer model: Trained using data from April to September during the years 2021–2024, to capture the characteristics and challenges of summer weather.

- Winter model: Trained using data from October to March of the same years, allowing the model to learn the distinct features of winter conditions.
- All-season model: Trained on data from January to December to build a generalized model capable of correcting forecasts across all seasons.

The evaluation is conducted using two withheld test datasets:

- The summer model is tested on August 2022 data.
- The winter model is tested on December 2022 data.
- The all-season model is evaluated on both summer and winter conditions to assess robustness and adaptability.

All models are evaluated over a 10-day forecast horizon (0–240 hours) to assess performance in medium-range forecasting and to explore practical applications in operational settings.

For the GEFS model over the Germany domain, the focus is on correcting 10m-WS and 100m-WS forecasts. The training strategies mirror those of the GFS-CONUS setup:

- Summer model: Trained using data from April to September during the years 2021 through 2023, to capture the characteristics and challenges of summer weather.
- Winter model: Trained using data from October to March of the same years, allowing the model to learn the distinct features of winter conditions.

The evaluation is conducted using two withheld test datasets:

- The summer model is tested on August 2022 data.
- The winter model is tested on December 2022 data.

For GEFS, all models are evaluated over a 16-day forecast range (0–384 hours), allowing assessment of long-range bias correction performance and exploration of their applicability in extended-range forecasting systems.

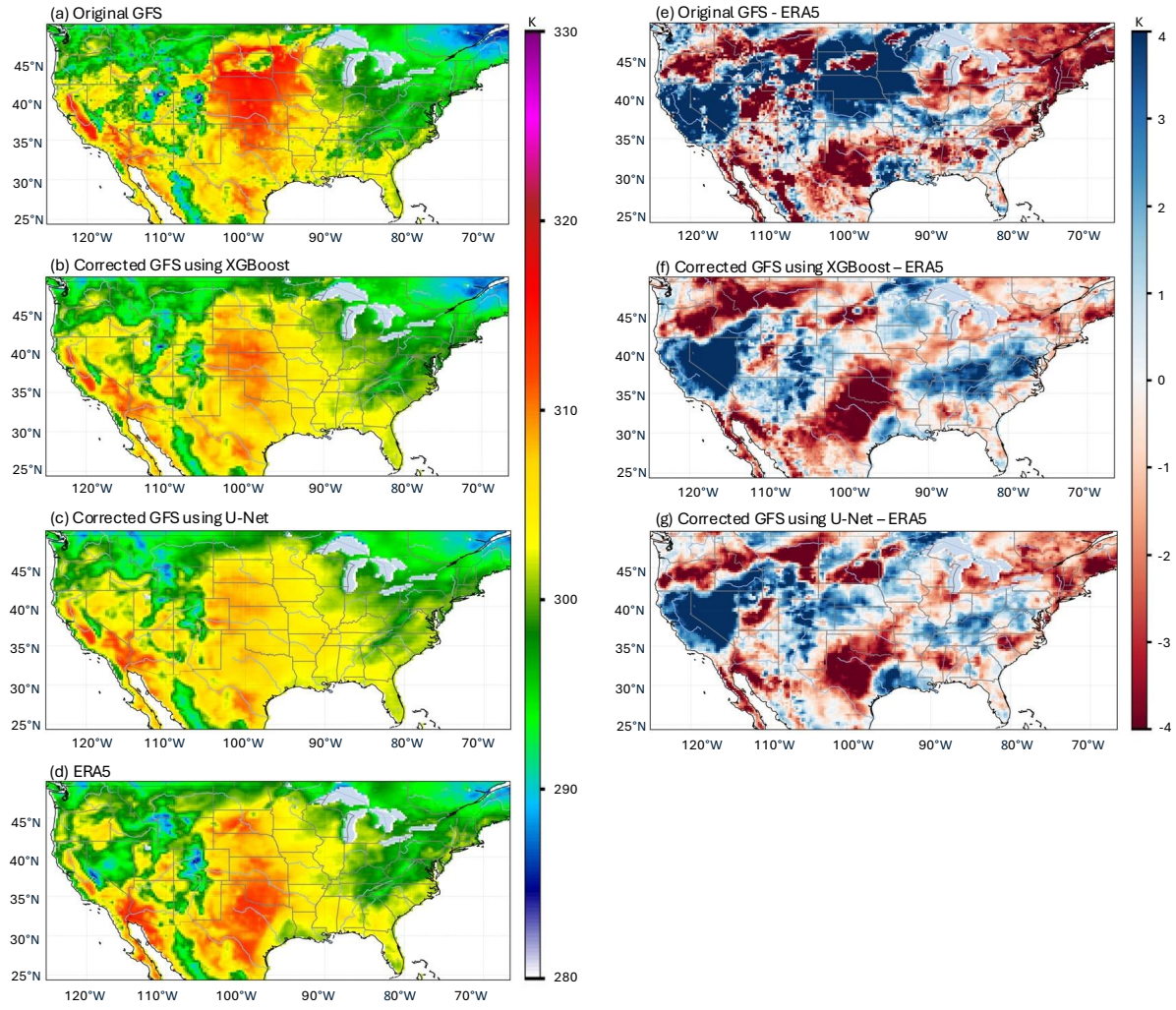

Figure 5.1: Illustration of the all-season model 120 h 2m-T forecast on 1 August 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

5.3 Testing Results and Evaluation

5.3.1 GFS: CONUS Region

Figures 5.1 and 5.2 present examples of 120-hour 2m-T forecasts for summer (August 1, 2022) using both the all-season model and the summer models. Similarly, Figs. 5.3 and 5.4 present examples of 120-hour 2m-T forecasts for winter (December 1, 2022) using both the all-season model and the winter model. In all cases, the original GFS forecasts (Figs. 5.1 (e), 5.2 (e), 5.3

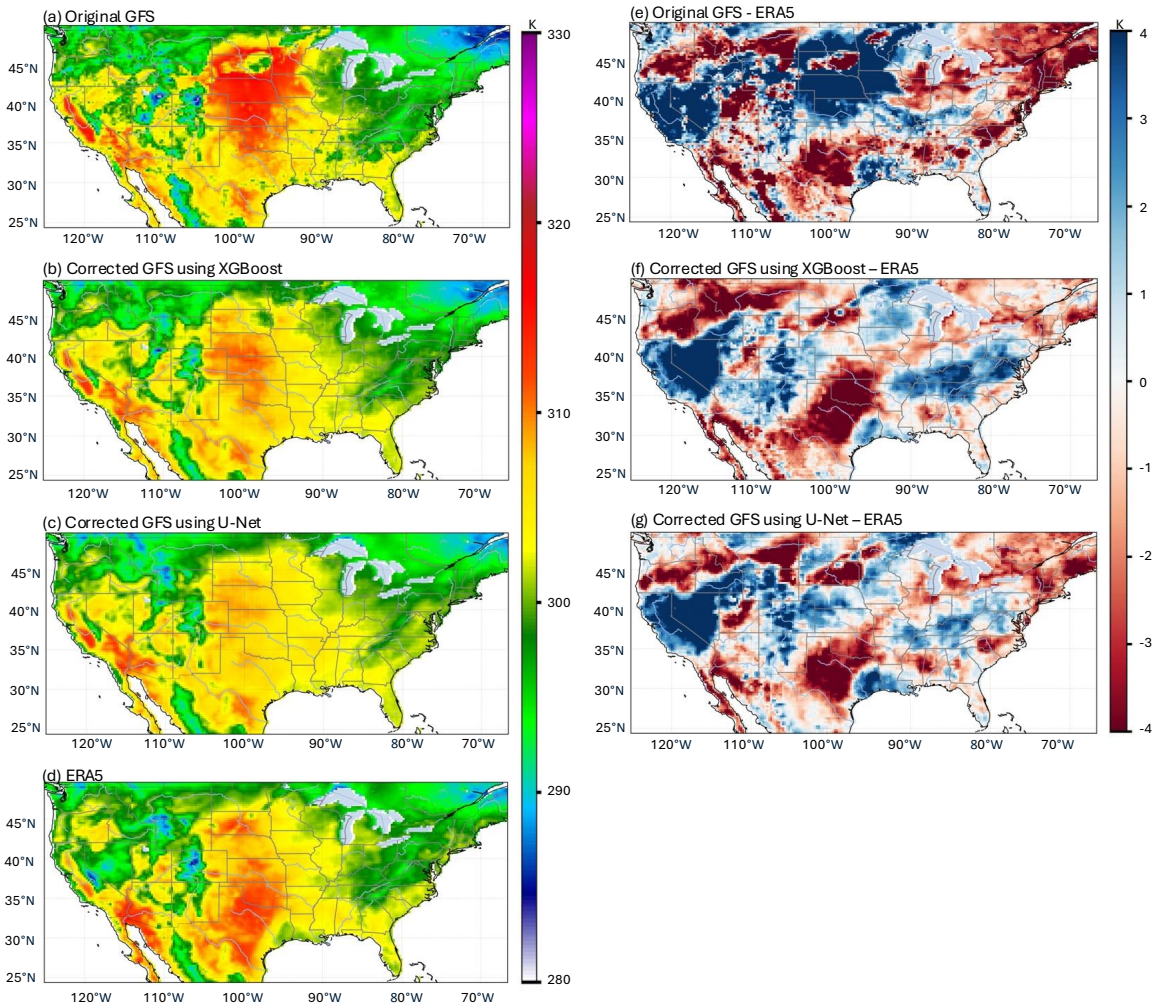

Figure 5.2: Illustration of the summer model 120 h 2m-T forecast on 1 August 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

(e), and 5.4 (e)) exhibit systematic biases. In these maps, blue indicates overestimation, and red indicates underestimation.

In the summer cases (Figs. 5.1 (e) and 5.2 (e)), the original GFS forecast tends to overestimate temperatures across the western mountainous regions of the U.S. and underestimate them in the central and southeastern regions. After applying XGBoost, the overestimation in the west is partially reduced, with noticeable improvements in the central and northeastern regions. U-Net further mitigates both overestimation and underestimation, showing significant correction particularly across the central and southeastern U.S.

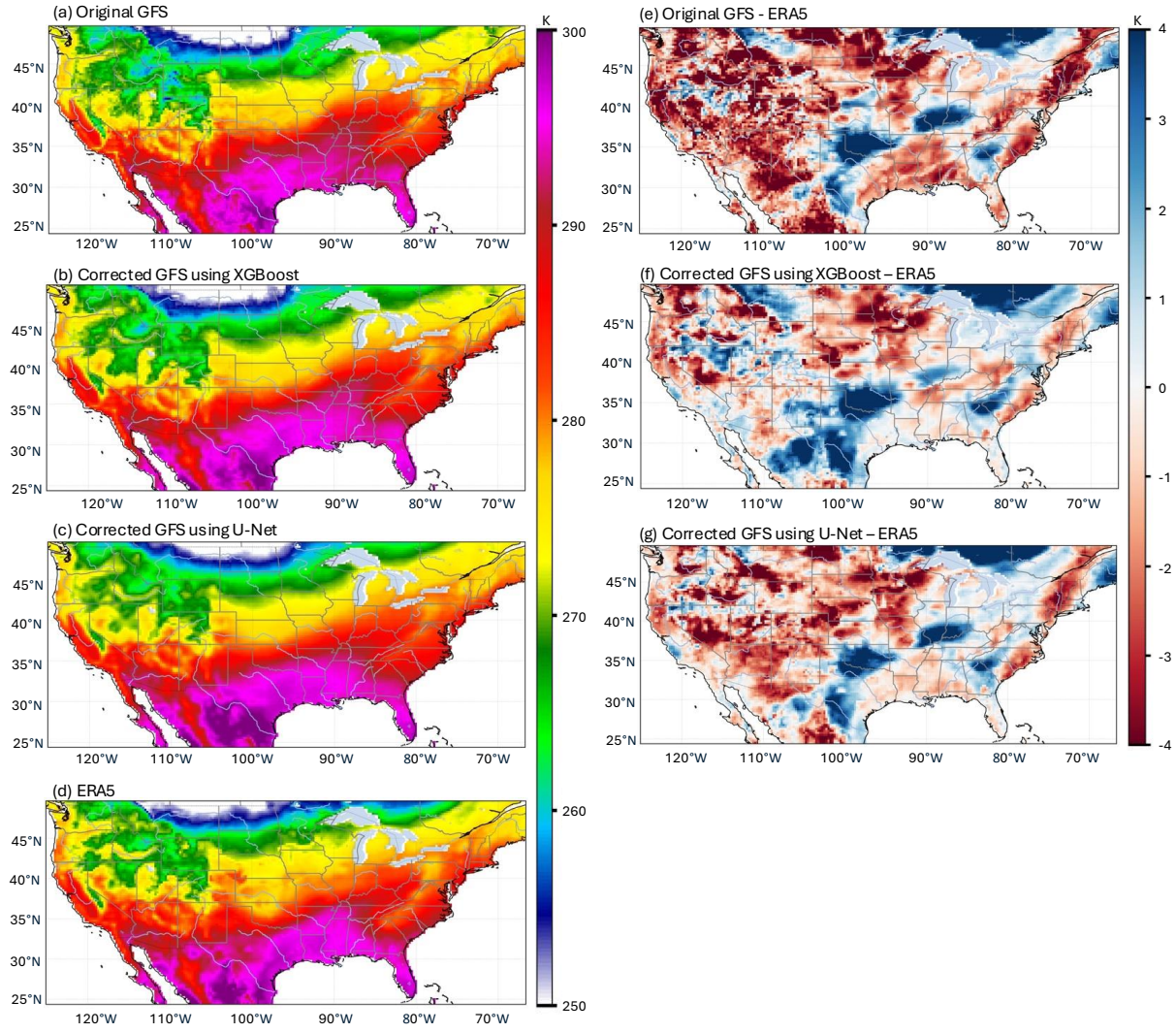

Figure 5.3: Illustration of the all-season model 120 h 2m-T forecast on 1 December 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

In the winter cases (Figs. 5.3 (e) and 5.4 (e)), the original GFS forecast underestimates temperatures across the western U.S. and the southeastern regions, where widespread warm biases are indicated by red shading. After applying XGBoost, the overall bias is reduced, particularly in the western mountainous areas. U-Net further corrects biases in both the western and southeastern U.S., showing a more spatially consistent improvement.

Figures 5.6 and 5.7 present examples of 48-hour 10m-WS forecasts for summer (August 1, 2022) using both the all-season model and the summer models. Similarly, Figs. 5.8 and 5.9 present

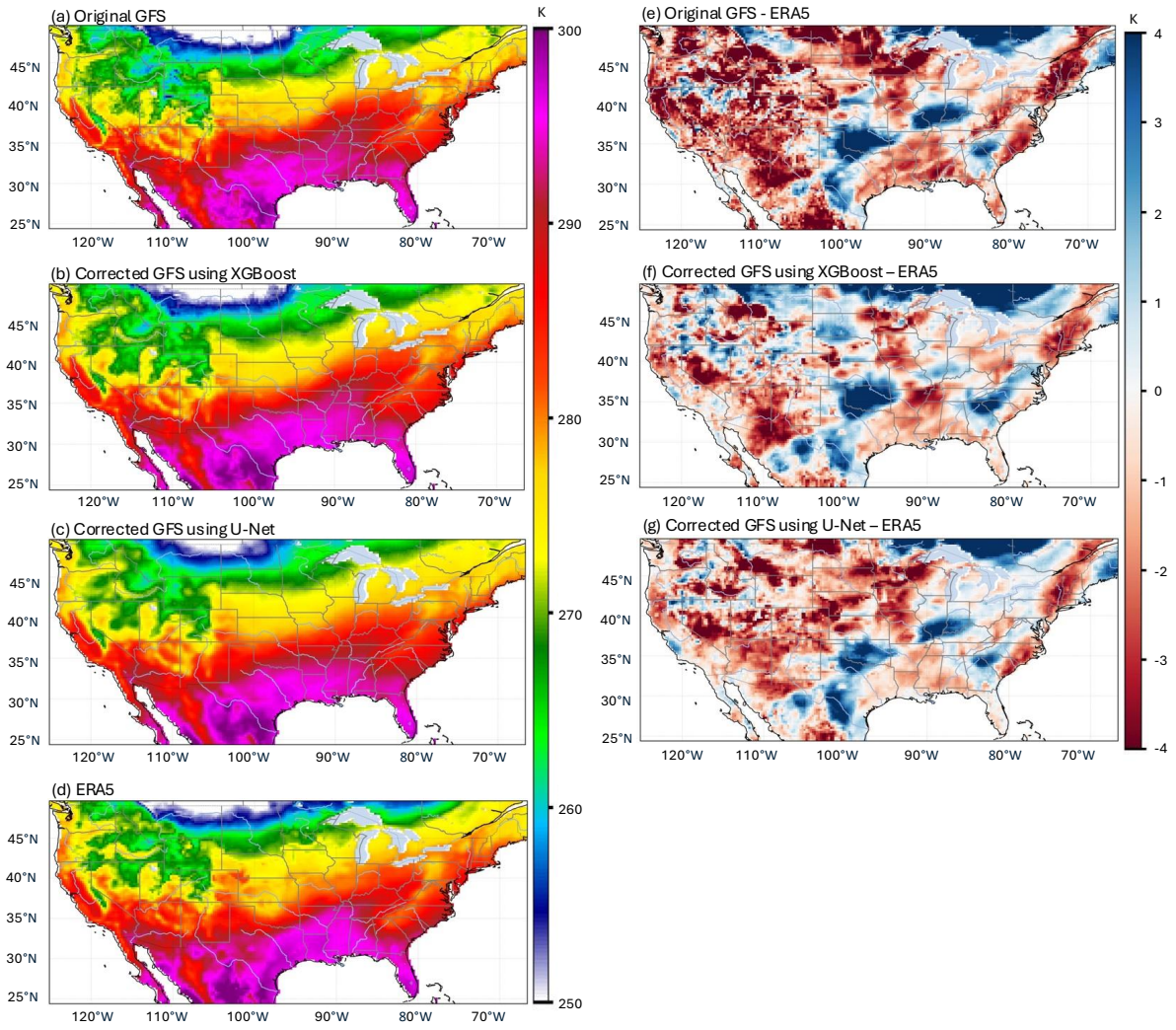

Figure 5.4: Illustration of the winter model 120 h 2m-T forecast on 1 December 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

examples of 48-hour 10m-WS forecasts for winter (December 1, 2022) using both the all-season model and the winter model. In all cases, the original GFS forecasts exhibit significant systematic errors compared to ERA5.

In the summer cases (Figs. 5.6 (e) and 5.7 (e)), the original GFS overestimates wind speed across many regions, especially in the central and western U.S. After applying XGBoost, the bias is reduced in these regions, with improvement over the central region. With U-Net the system-

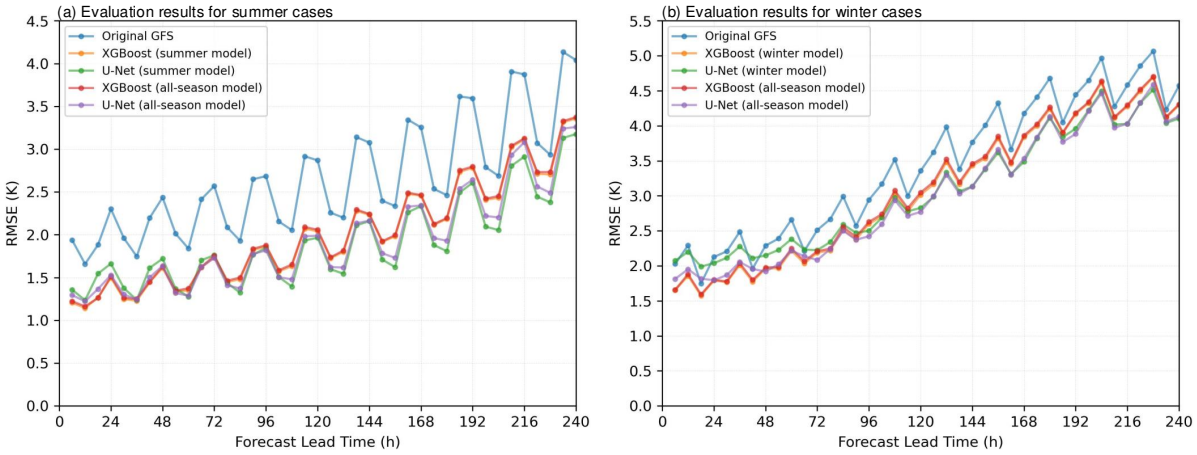

Figure 5.5: Domain-averaged 2m-T (a) RMSE for the 00Z cycle, averaged over August 2022 (b) RMSE for the 00Z cycle, averaged over December 2022

atic errors are reduced as well, particularly over the southeastern U.S., showing a more effective correction.

In the winter cases (Figs. 5.8 (e) and 5.9 (e)), the original GFS overestimates wind speed in the central U.S. After applying XGBoost and U-Net, the overestimation in this region is somewhat reduced. It shows a stronger ability to correct systematic errors.

Figures 5.11 and 5.12 show examples of 48-hour 100m-WS forecasts for summer (August 1, 2022) using both the all-season model and the summer models. Similarly, Figs. 5.13 and 5.14 present examples of 48-hour 100m-WS forecasts for winter (December 1, 2022) using both the all-season model and the winter model.

In the summer cases (Figs. 5.11 and 5.12), the original GFS forecasts show an overestimation of wind speed over the central U.S. After applying XGBoost, the bias in this region is noticeably reduced. U-Net further improves the forecast by reducing biases across a broader area, particularly in the southern U.S.

In the winter cases, a similar overestimation pattern occurred over the central U.S. and along the East Coast. After applying both XGBoost and U-Net corrections, the bias in these regions is effectively reduced.

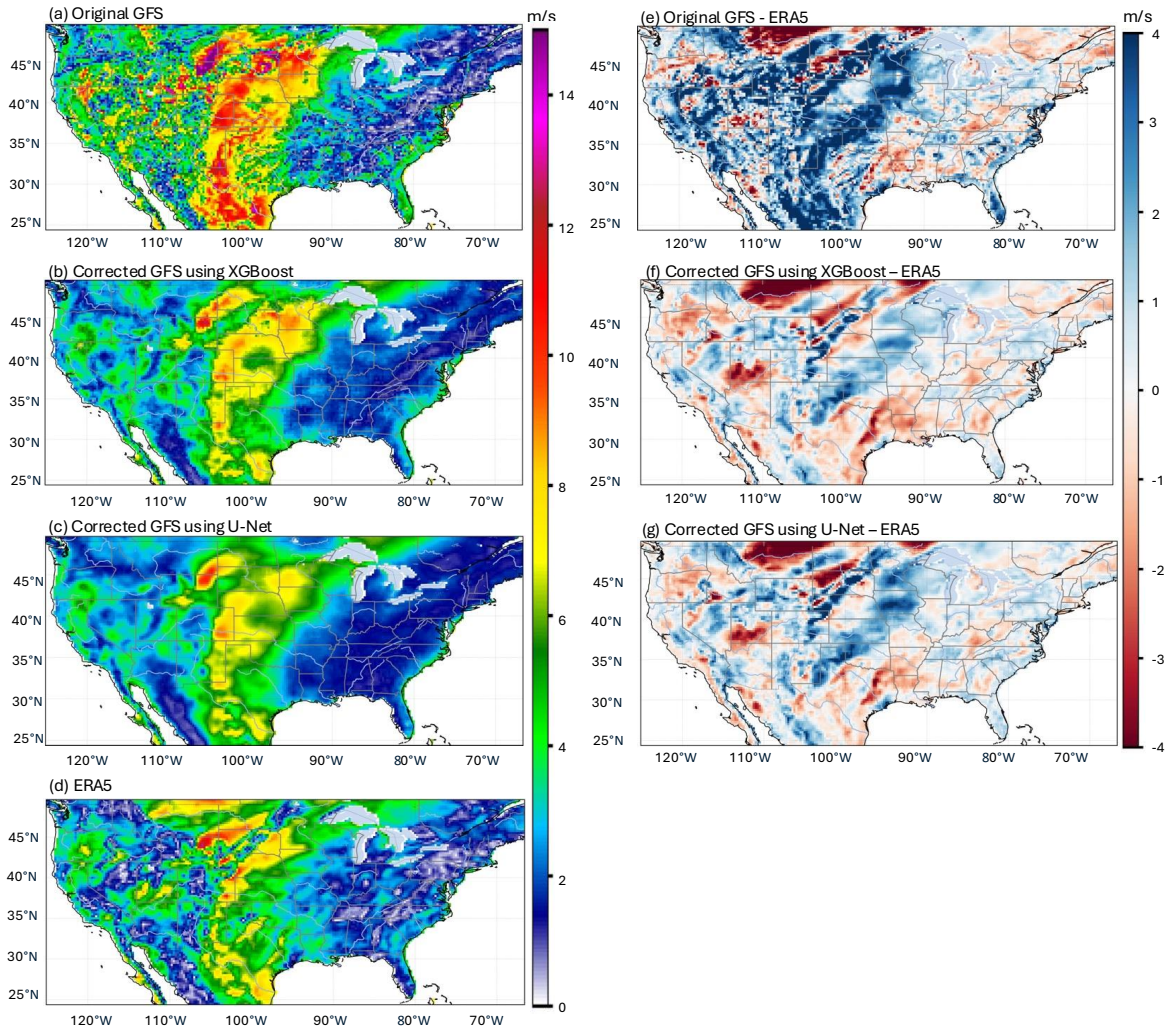

Figure 5.6: Illustration of the all-season model 48 h 10m-wind speed forecast on 1 August 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

For determining bias correction efficiency, RMSE has been calculated both for pre- and after-bias correction forecast values. The RMSE according to Eq. (5.1) is the square root of the mean square errors of forecast and observed values. It is computed over a region spanning 24.5°N to 49.5°N in latitude and -125.0°W to -66.75°W in longitude, extending beyond just the CONUS domain.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{5.1}$$

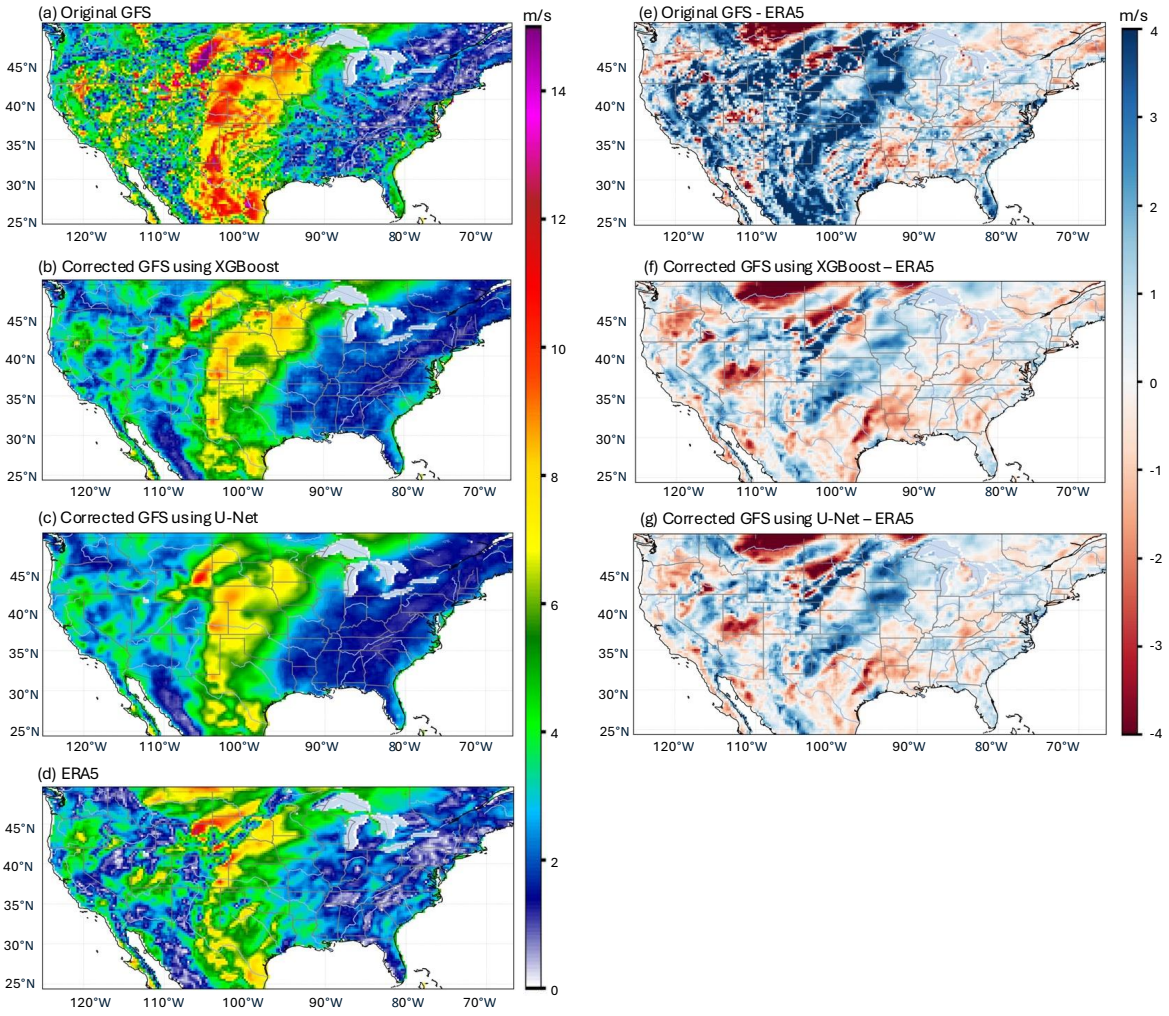

Figure 5.7: Illustration of the summer model 48 h 10m-wind speed forecast on 1 August 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

Where $\hat{y}$ is the predicted value (original or bias-corrected GFS forecasts), y_i is the actual observed value (ERA5 data), and n is the total number of data points.

The 2m-T bias in the original GFS forecast increases with forecast time, with a significant rise in RMSE beyond 100 hours. The error in winter (December 2022) is higher than in summer (August 2022) (Fig. 5.5 (a)), indicating greater GFS 2m-T forecast bias in winter (Fig. 5.5 (b)).

In the summer case, the XGBoost all-season model (red line) and the XGBoost summer model (orange line) show nearly identical performance. During the first two days of the forecast period, both models achieve the lowest RMSE values among all methods, indicating strong effectiveness

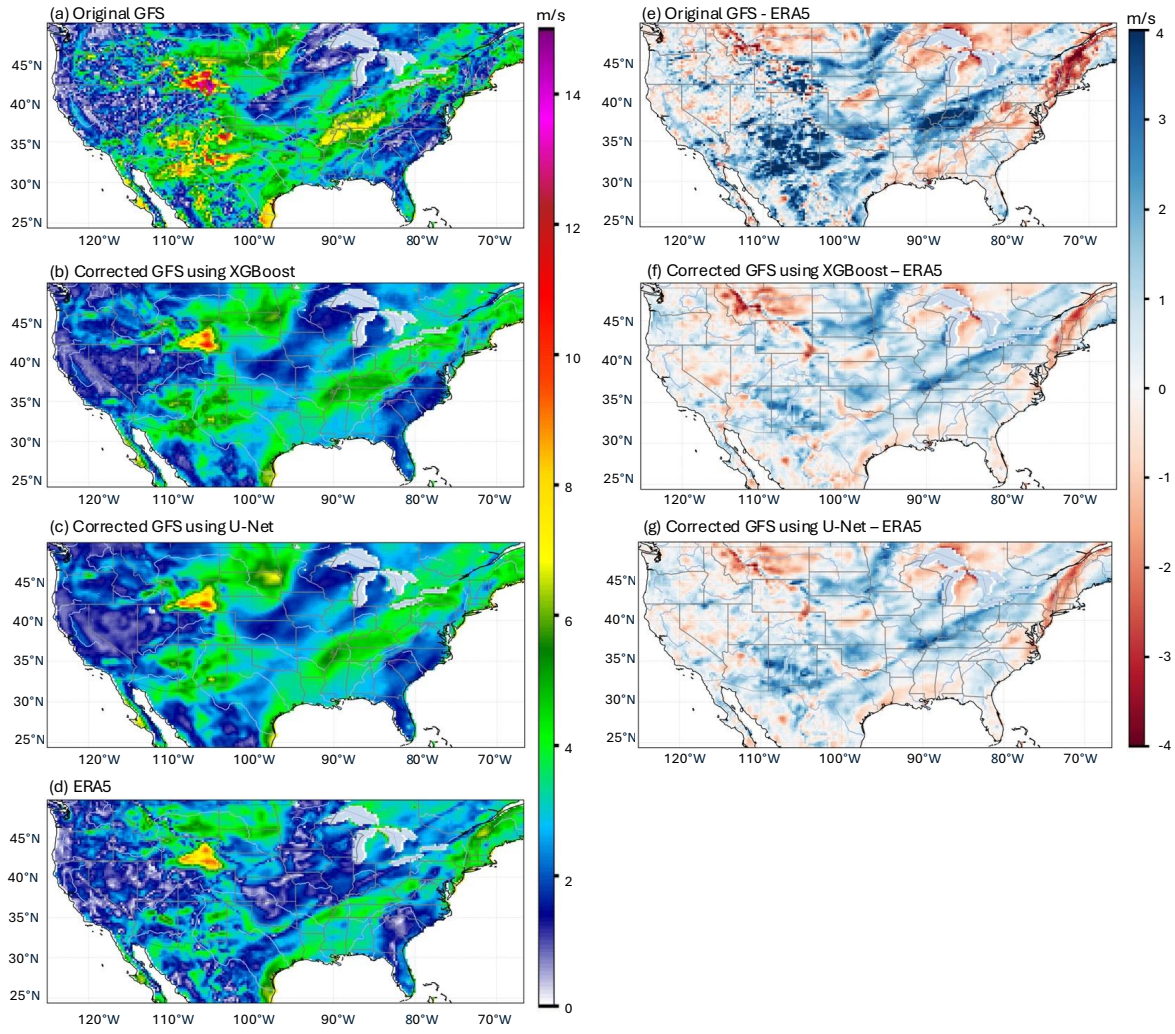

Figure 5.8: Illustration of the all-season model 48 h 10m-wind speed forecast on 1 December 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

in short-term bias correction. However, beyond Day 2, the performance trend shifts. The U-Net summer model (green line) begins to outperform XGBoost. From Day 2 through Day 10, the U-Net summer model consistently achieves the lowest RMSE, showing superior capability in capturing complex bias patterns and maintaining accuracy over extended forecast lead time during the summer season.

In the winter case, a similar pattern emerges during the early forecast period: the XGBoost all-season model (red line) and the XGBoost winter model (orange line) show nearly identical

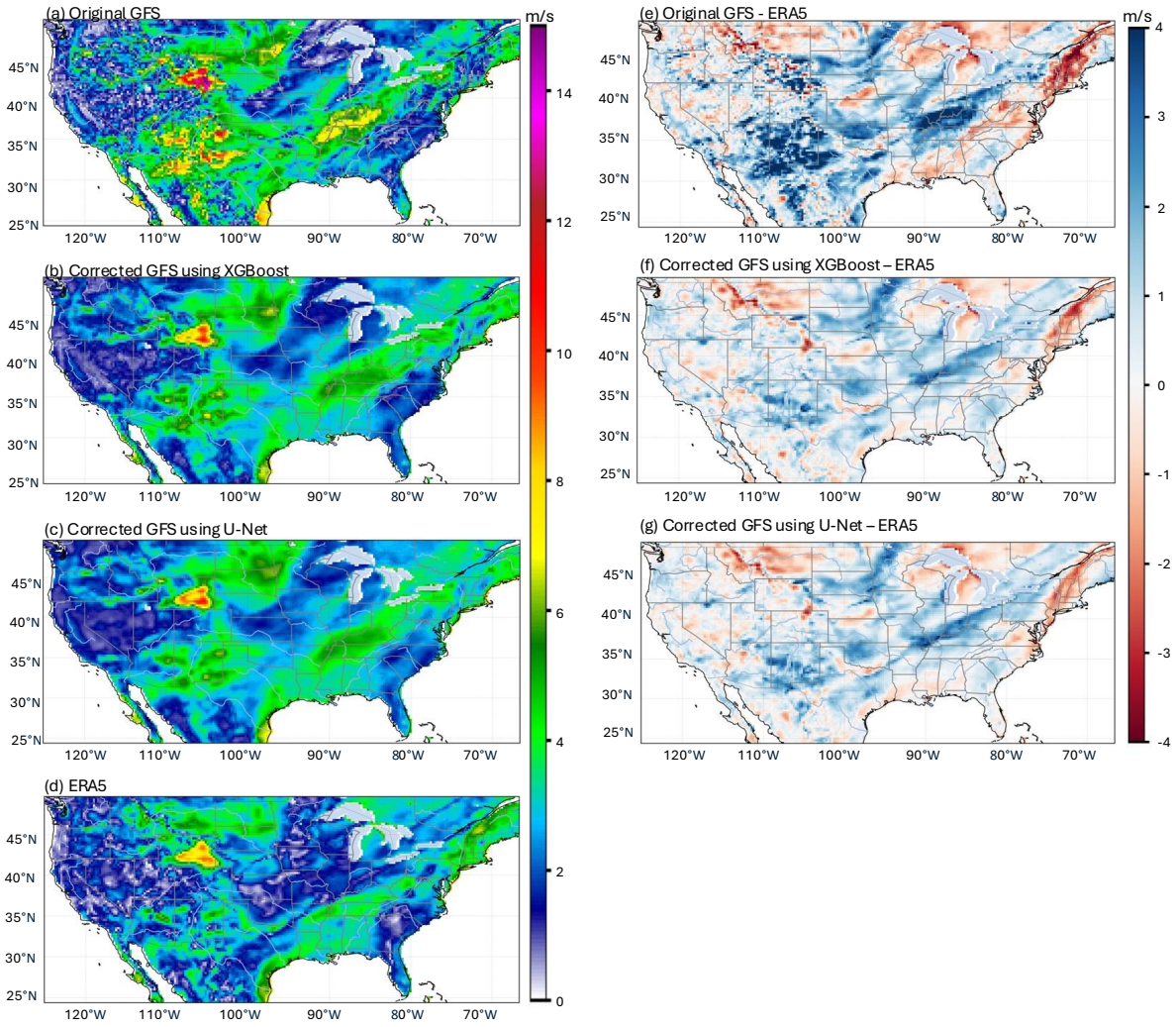

Figure 5.9: Illustration of the winter model 48 h 10m-wind speed forecast on 1 December 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

performance and achieve the lowest RMSE values during the first three days. However, after Day 3, the advantage transitions to the U-Net all-season model (purple line), which maintains consistently lower RMSE than all other models for the remainder of the forecast horizon. Interestingly, the U-Net all-season model performs better than the U-Net winter model, indicating that in winter, training on a more diverse all-season dataset may improve the model's ability to generalize across the complex weather experienced in the cold season.

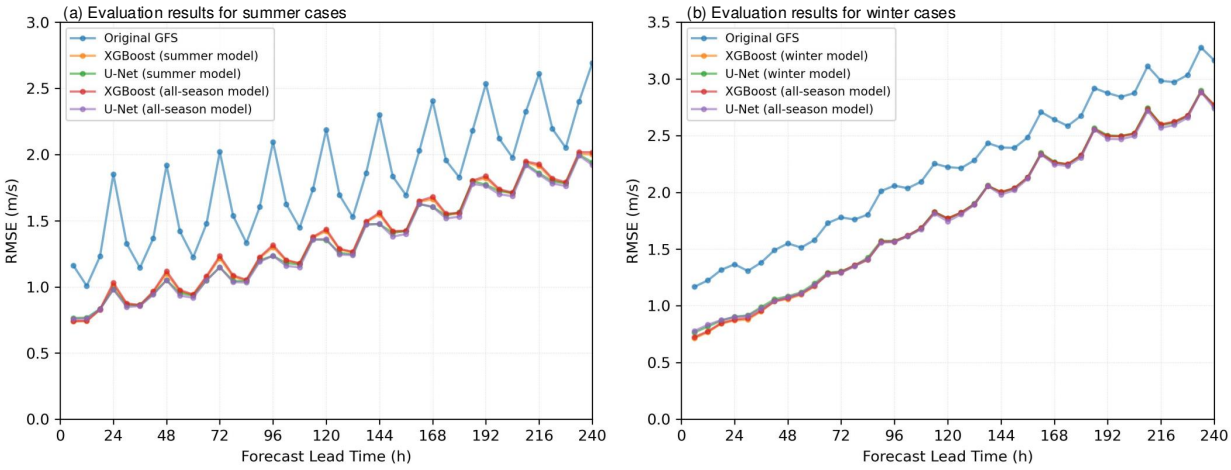

Figure 5.10: Domain-averaged 10m-wind speed (a) RMSE for the 00Z cycle, averaged over August 2022 (b) RMSE for the 00Z cycle, averaged over December 2022

Figure 5.10 presents RMSE for 10m-wind speed forecasts in summer (Fig. 5.10(a)) and winter (Fig. 5.10(b)). The original GFS (blue line) shows the highest RMSE at all lead times, with increasing RMSE over longer forecasts, particularly in winter.

In the summer cases, the RMSE curves of different models show a pattern similar to that observed in the 2-meter temperature forecasts. The XGBoost all-season model (red line) and the XGBoost summer model (orange line) perform almost identically throughout the forecast period. However, both the U-Net summer and U-Net all-season models yield lower RMSEs compared to the XGBoost models, indicating that U-Net provides more effective bias correction during the summer season.

In the winter cases, during the first 24 hours, the XGBoost all-season model (red line) and the XGBoost winter model (orange line) again show nearly identical performance and outperform the U-Net models in the short-term forecast range. After the first day, the RMSE curves of all models converge, showing similar performance. Nevertheless, the U-Net all-season model maintains a slight advantage in the longer forecast lead times, suggesting better long-term bias correction capabilities in winter.

Figure 5.15 shows RMSE for 100m-wind speed forecasts in summer (Fig. 5.15(a)) and winter (Fig. 5.15(b)), averaged over August 2022 and December 2022. The original GFS forecast has the

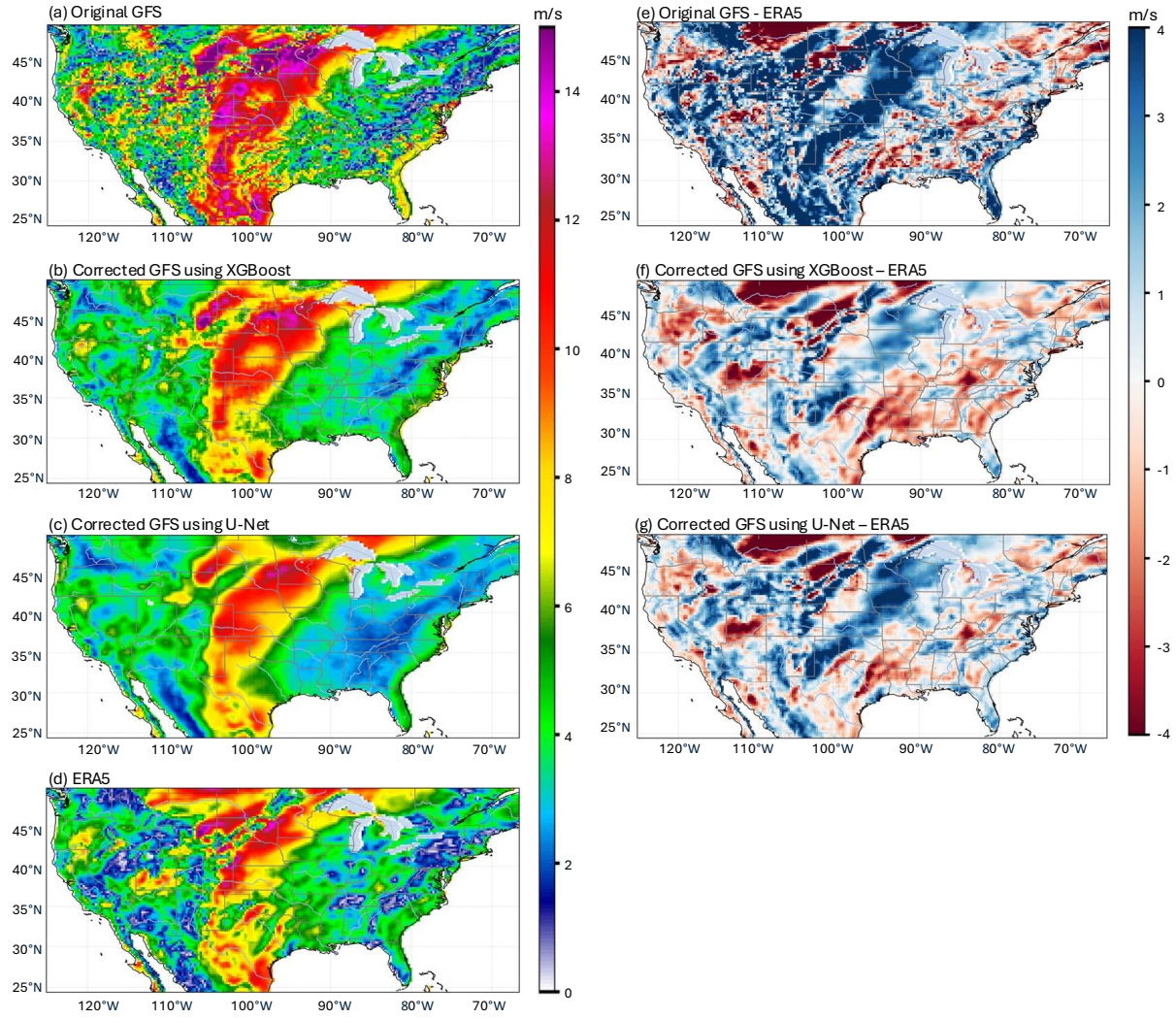

Figure 5.11: Illustration of the all-season model 48 h 100m-wind speed forecast on 1 August 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

highest RMSE, with errors increasing over time. Winter RMSE is significantly higher than summer RMSE, particularly at longer lead times, suggesting greater difficulty in winter 100-m wind speed forecasting.

In the summer case, the overall RMSE results indicate that the U-Net all-season model delivers the best performance. Beyond 120 forecast hours, it achieves lower RMSE values than the U-Net summer model, as evidenced by the separation between the purple and green lines in the figure.

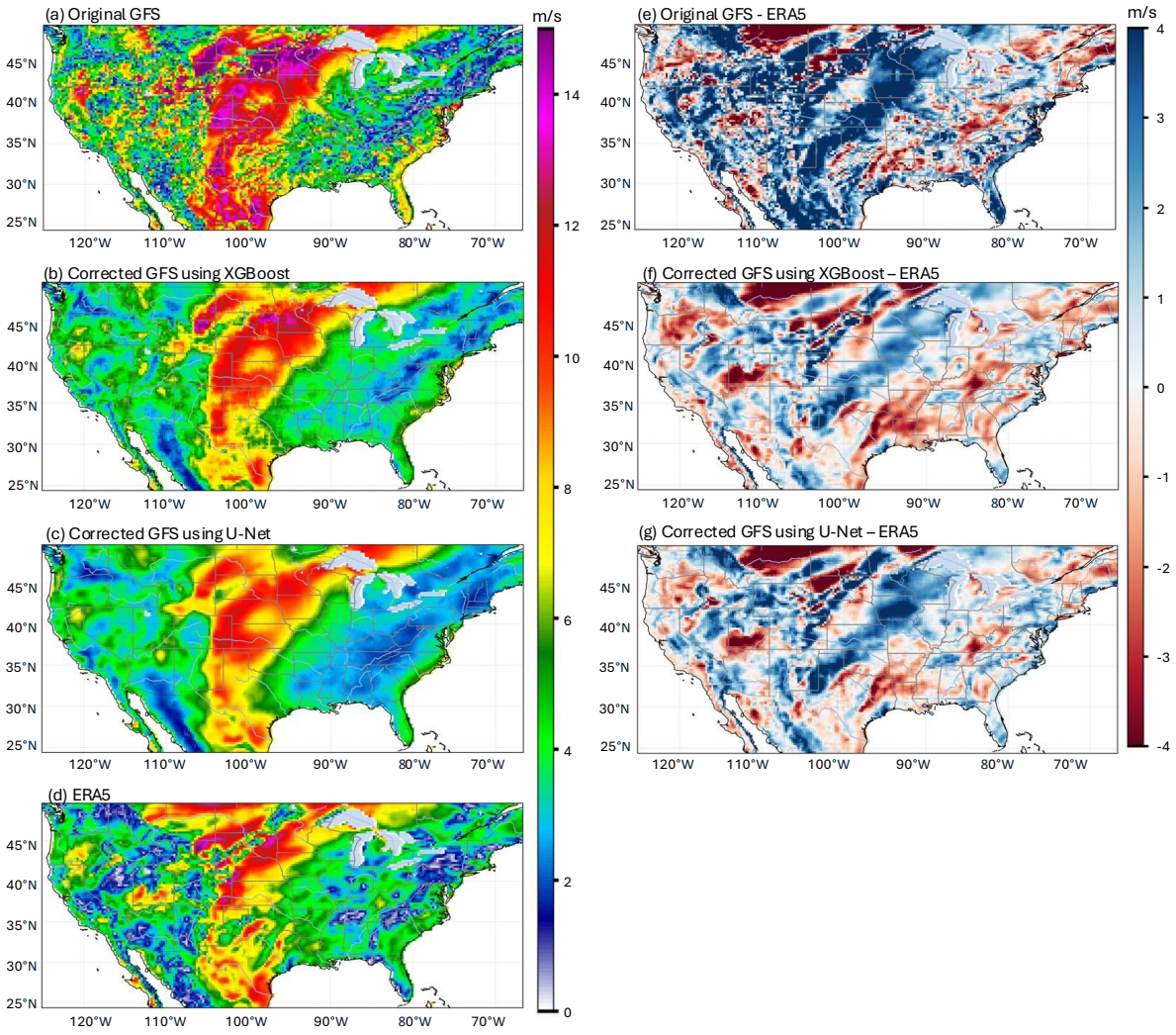

Figure 5.12: Illustration of the summer model 48 h 100m-wind speed forecast on 1 August 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

In the winter case, the XGBoost models outperform U-Net during the initial 24 hours, demonstrating stronger short-term correction capability. Between 24 and 144 hours, all models show similar performance, with their RMSE curves largely overlapping. However, after 144 hours, the U-Net all-season model begins to show a distinct advantage, achieving the lowest RMSE among all models and highlighting its efficiency in long-range bias correction throughout the winter season.

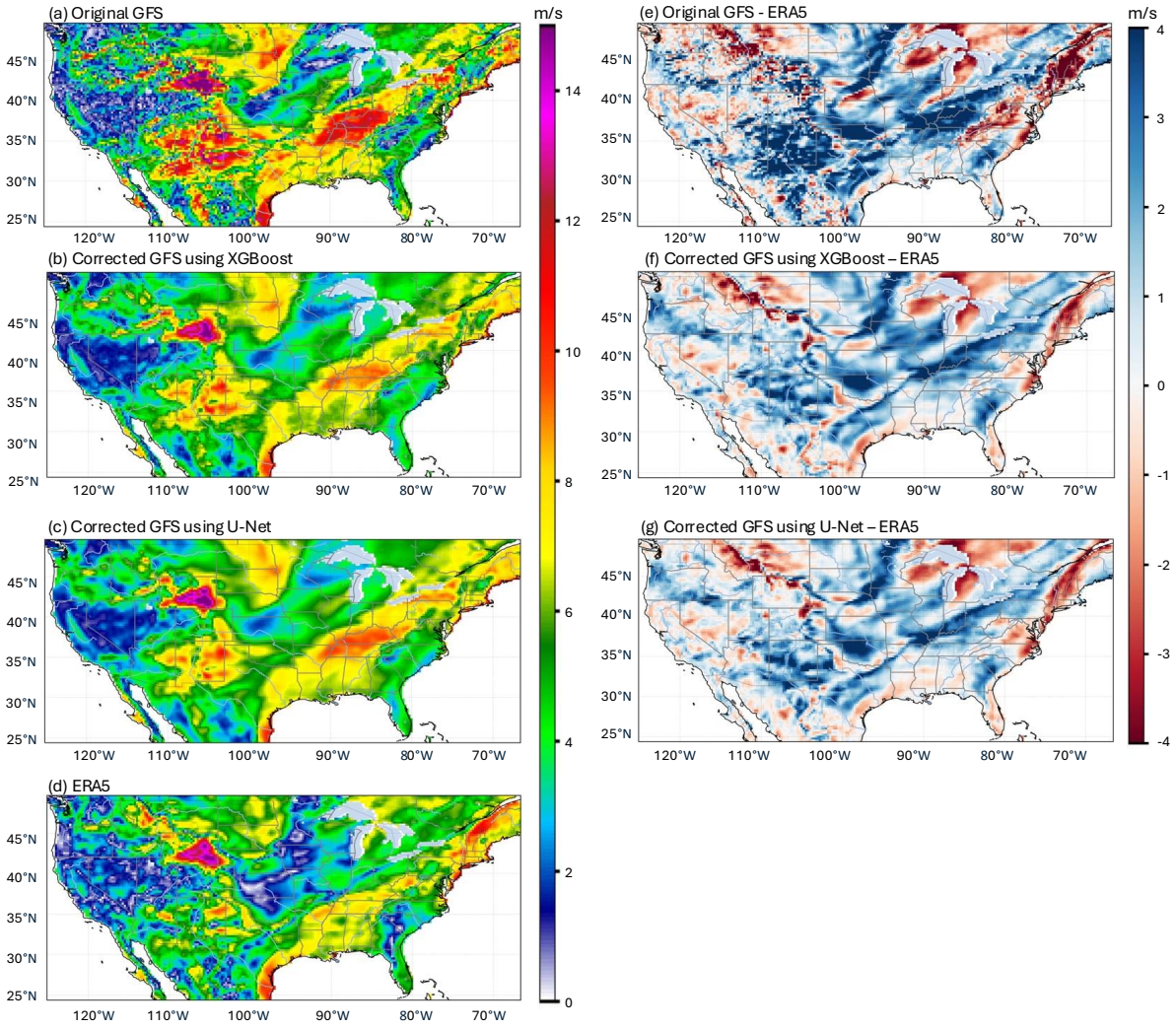

Figure 5.13: Illustration of the all-season model 48 h 100m-wind speed forecast on 1 December 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

5.3.2 GEFS: Germany Region

Figures 5.16, 5.17, 5.18, and 5.19 present the GEFS ensemble mean forecasts of 10-m and 100-m wind speeds (GEFS_EM) over Germany, along with the results after bias correction using XGBoost and U-Net, and the corresponding bias distributions with respect to ERA5. The figures illustrate the results for summer (Figs. 5.16 and 5.17) and winter (Figs. 5.18 and 5.19). In each figure, subplots (e), (f), and (g) respectively show the residuals between the original GEFS_EM forecast and ERA5, the XGBoost-corrected forecast and ERA5, and the U-Net-corrected forecast

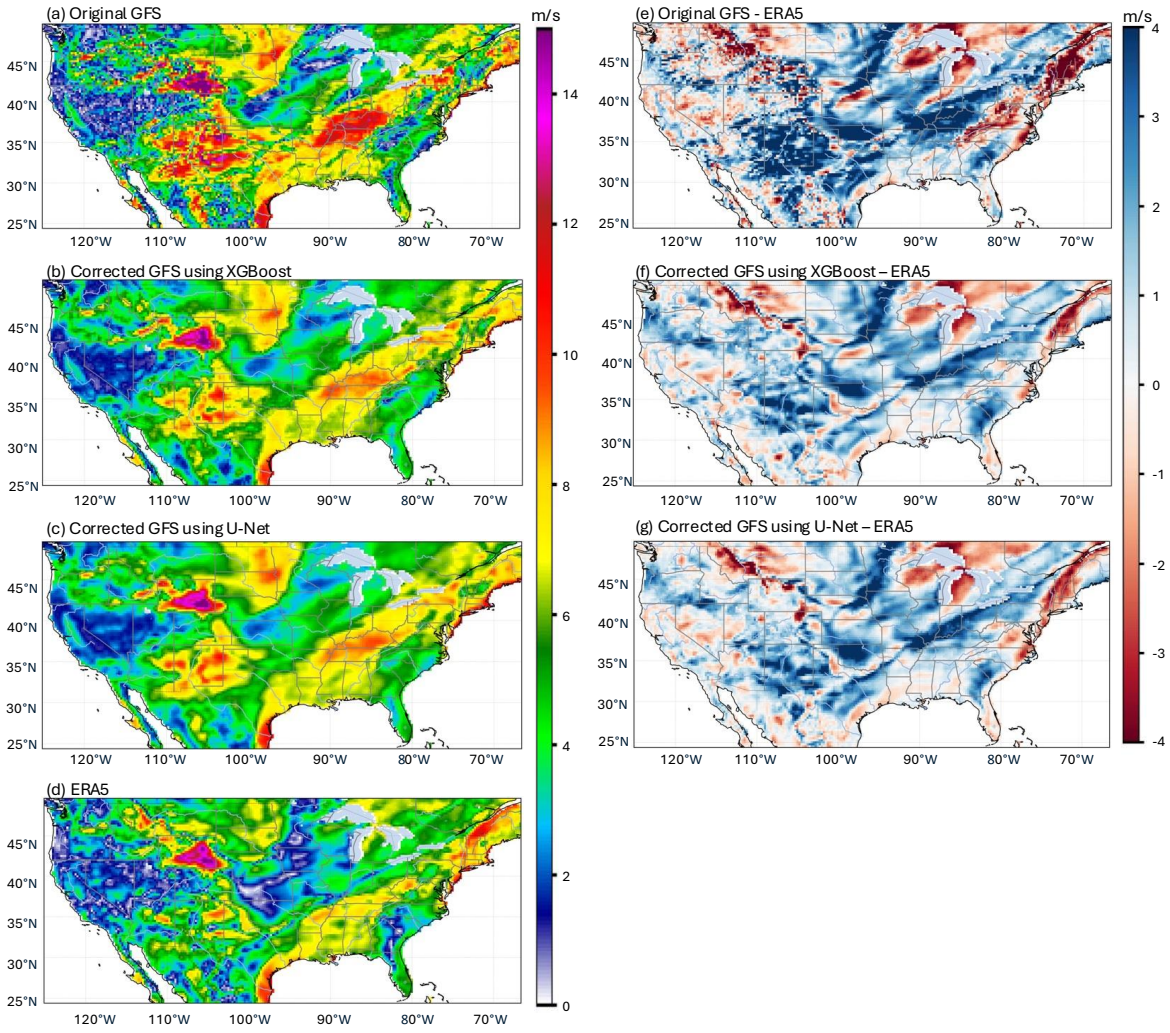

Figure 5.14: Illustration of the winter model 48 h 100m-wind speed forecast on 1 December 2022: (a) Original GFS; (b) Corrected GFS using XGBoost; (c) Corrected GFS using U-Net; (d) ERA5; (e) Original GFS - ERA5; (f) Corrected GFS using XGBoost - ERA5; (g) Corrected GFS using U-Net - ERA5

and ERA5, which are used to visually assess the spatial improvement achieved by each correction method.

During summer, the original GEFS_EM wind speed forecasts (subplot (e) in Figs. 5.16 and 5.17) exhibit significant systematic overestimation. The 10-meter wind speed is generally overestimated by 2–3 m/s over the North German Plain and central hilly regions, while the bias for the 100-meter wind speed is even more severe, with localized areas exceeding 3.5 m/s. The spatial

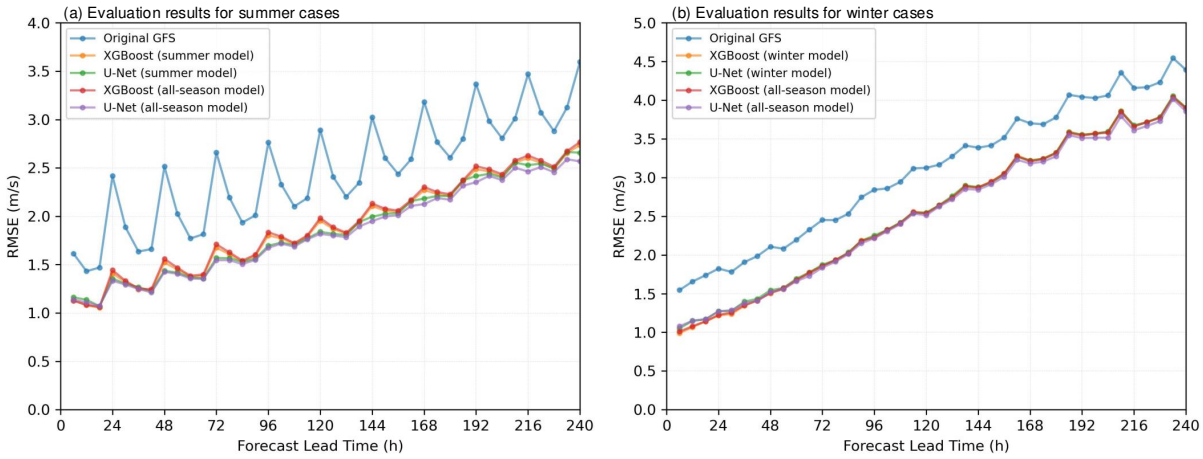

Figure 5.15: Domain-averaged 100m-wind speed (a) RMSE for the 00Z cycle, averaged over August 2022 (b) RMSE for the 00Z cycle, averaged over December 2022

distribution of the bias is relatively concentrated and shows geographic continuity, indicating the existence of learnable bias patterns.

After correction with XGBoost (subplot (f) in Figs. 5.16 and 5.17), the overall bias is reduced, especially in areas with strong overestimation, such as northern Germany. However, although the magnitude of the bias is mitigated, the residuals remain spatially fragmented and lack continuity.

The U-Net correction results (subplot (g) in Figs. 5.16 and 5.17) demonstrate stronger spatial modeling capabilities. The residuals appear smoother, with more continuous spatial transitions, especially in hilly and forested areas where the patterns look more natural. For the 10-meter wind speed, U-Net significantly improves overestimation, particularly in the southern hilly regions, indicating that it effectively captures the complex spatial relationship between wind and geographic background. However, in the summer correction of 100-meter wind speed, U-Net does not fully eliminate the bias, suggesting it still faces challenges in modeling higher-altitude wind features.

During winter, the bias distribution (subplot (e) in Fig. 5.18 and 5.19) becomes more pronounced compared to summer. The original GEFS_EM forecasts exhibit significant overestimation across most of Germany, especially in the central and northern regions and hilly areas. The 10-meter wind speed bias ranges from 2.5 to 3.5 m/s, while the 100-meter bias can reach up to 4 m/s. The spatial distribution is more irregular, with more localized extreme bias regions.

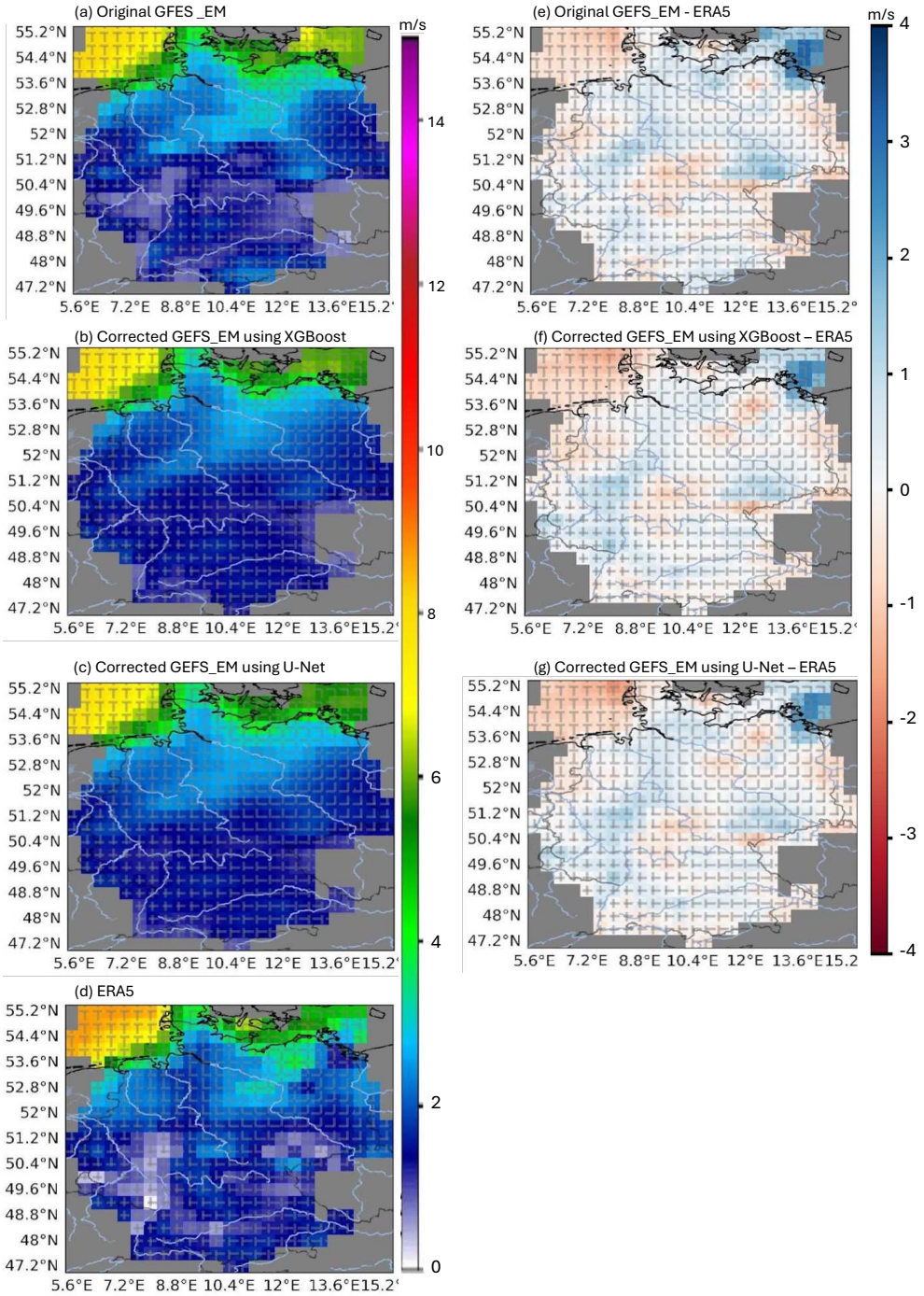

Figure 5.16: Illustration of the summer model 48h 10m-wind speed forecast on 1 August 2022: (a) Original GEFS_EM; (b) Corrected GEFS_EM using XGBoost; (c) Corrected GEFS_EM using U-Net; (d) ERA5; (e) Original GEFS_EM - ERA5; (f) Corrected GEFS_EM using XGBoost - ERA5; (g) Corrected GEFS_EM using U-Net - ERA5

The XGBoost-corrected residuals (subplot (f) in Fig. 5.18 and 5.19) still reduce the overall bias in winter, but the lack of spatial continuity becomes even more apparent. In complex terrain, the correction is obviously not sufficient, so XGBoost, despite being able to tune the size of the bias, is unable to model spatial coherence well.

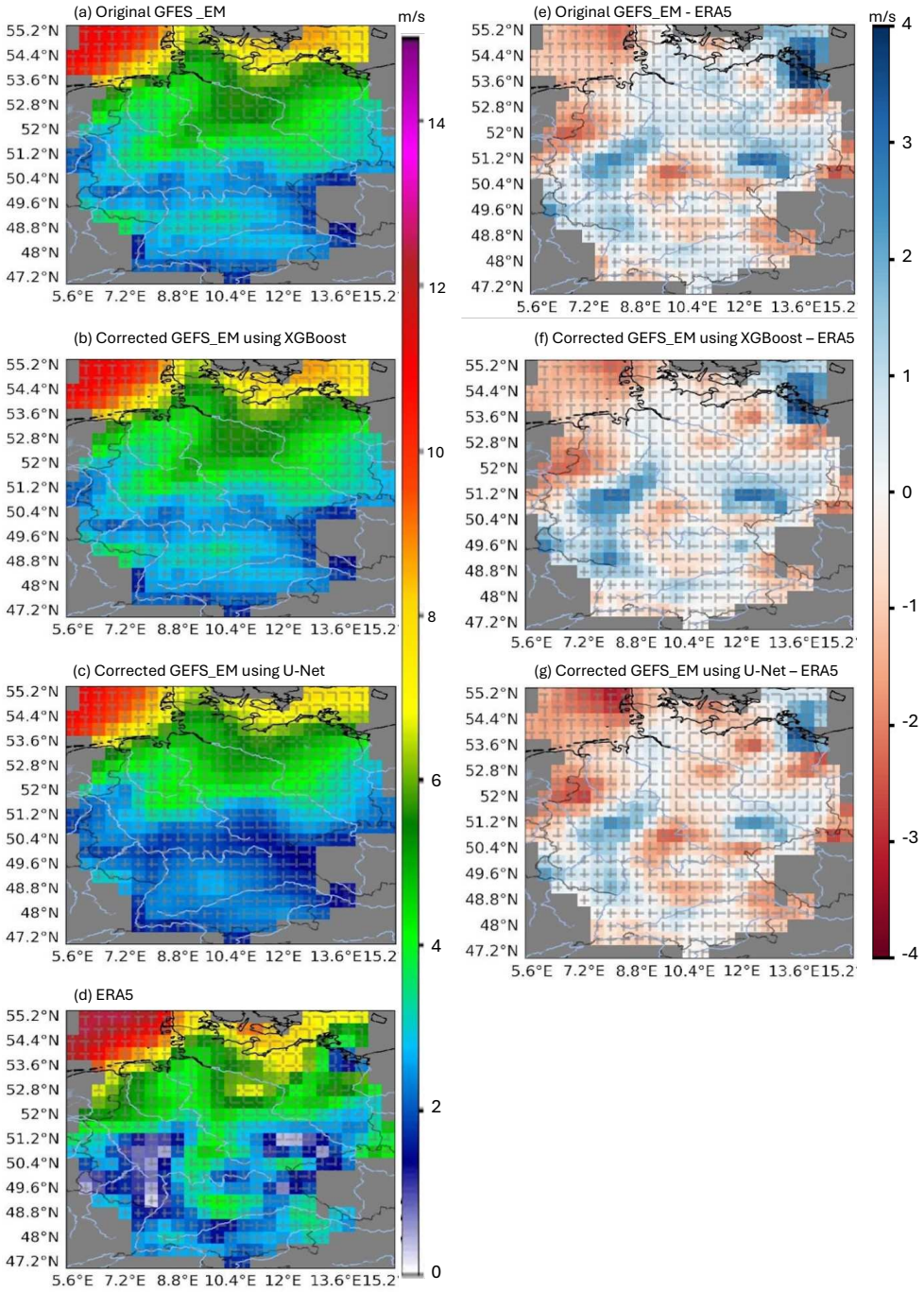

Figure 5.17: Illustration of the summer model 48 h 100m-wind speed forecast on 1 August 2022: (a) Original GEFS_EM; (b) Corrected GEFS_EM using XGBoost; (c) Corrected GEFS_EM using U-Net; (d) ERA5; (e) Original GEFS_EM - ERA5; (f) Corrected GEFS_EM using XGBoost - ERA5; (g) Corrected GEFS_EM using U-Net - ERA5

In contrast, U-Net performs more stably in winter correction. Its residual maps (subplot (g) in Figs. 5.18 and 5.19) show significant spatial consistency, enabling smooth bias transitions across larger scales. It performs especially well in regions such as the southern hills and eastern forested areas of Germany. U-Net successfully reconstructs the spatial structure of the ERA5 wind field and

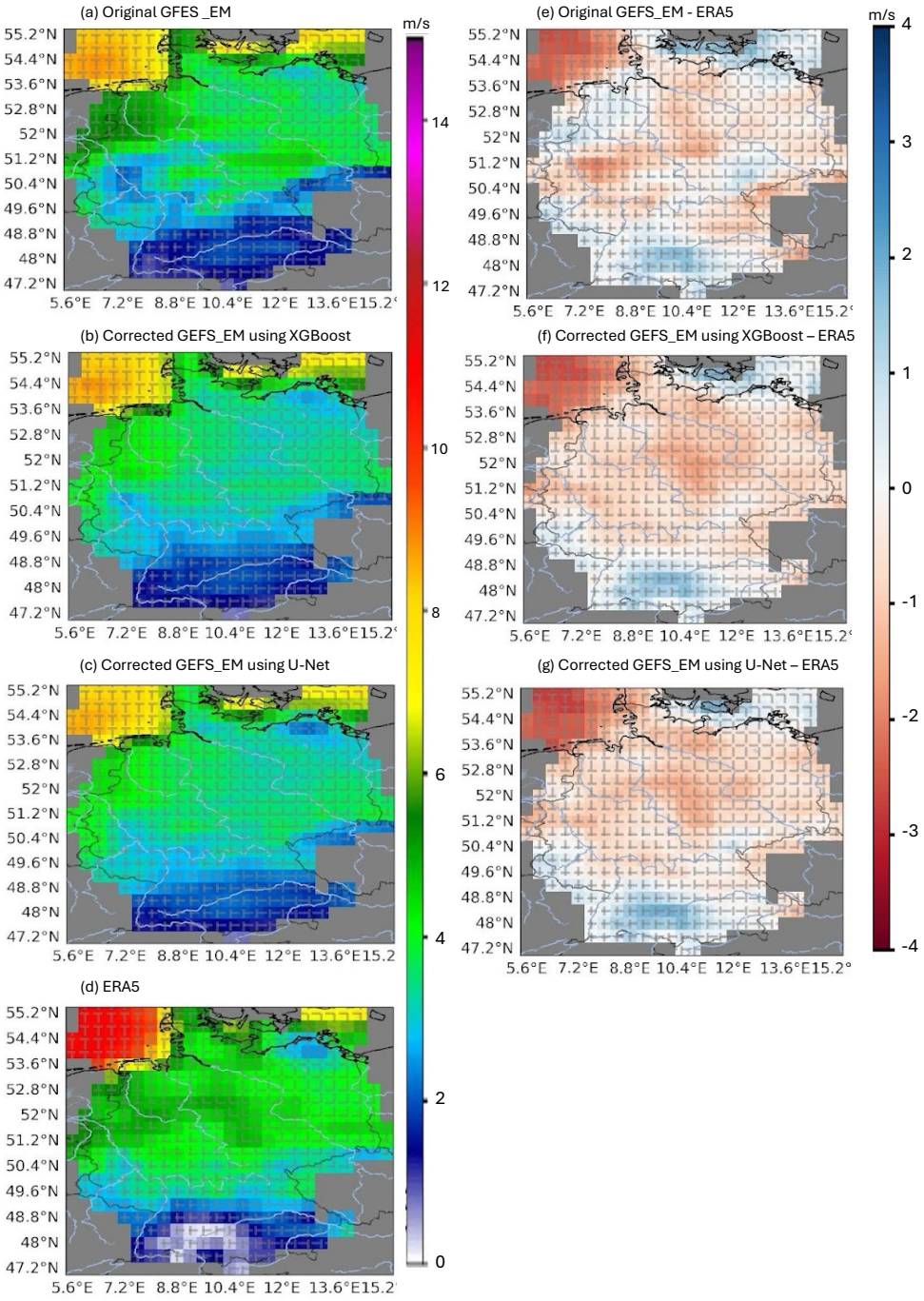

Figure 5.18: Illustration of the winter model 36h 10m-wind speed forecast on 1 December 2022: (a) Original GEFS_EM; (b) Corrected GEFS_EM using XGBoost; (c) Corrected GEFS_EM using U-Net; (d) ERA5; (e) Original GEFS_EM - ERA5; (f) Corrected GEFS_EM using XGBoost - ERA5; (g) Corrected GEFS_EM using U-Net - ERA5

demonstrates strong spatial modeling capabilities for both 10-meter and 100-meter wind speeds. This indicates that even under the highly variable and terrain-complex conditions of winter, U-Net maintains good spatial generalization ability.

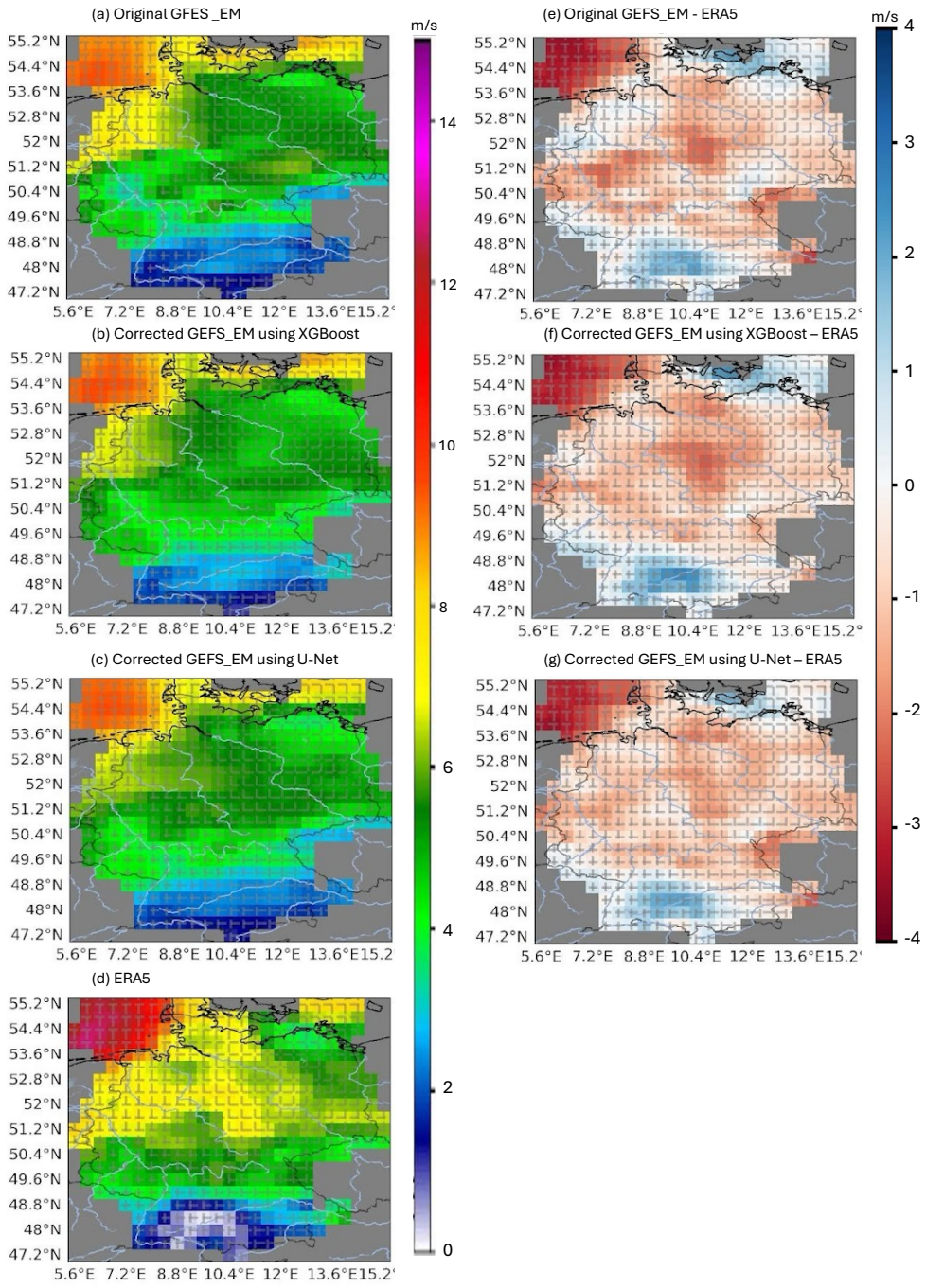

Figure 5.19: Illustration of the winter model 36 h 100m-wind speed forecast on 1 December 2022: (a) Original GEFS_EM; (b) Corrected GEFS_EM using XGBoost; (c) Corrected GEFS_EM using U-Net; (d) ERA5; (e) Original GEFS_EM - ERA5; (f) Corrected GEFS_EM using XGBoost - ERA5; (g) Corrected GEFS_EM using U-Net - ERA5

To quantitatively evaluate the performance of the XGBoost and U-Net correction methods, Figs. 5.20 and 5.21 illustrate the RMSE trends of 10-meter and 100-meter wind speeds over forecast lead times (0–48 hours and 0-36 hours) for summer and winter seasons. The evaluation

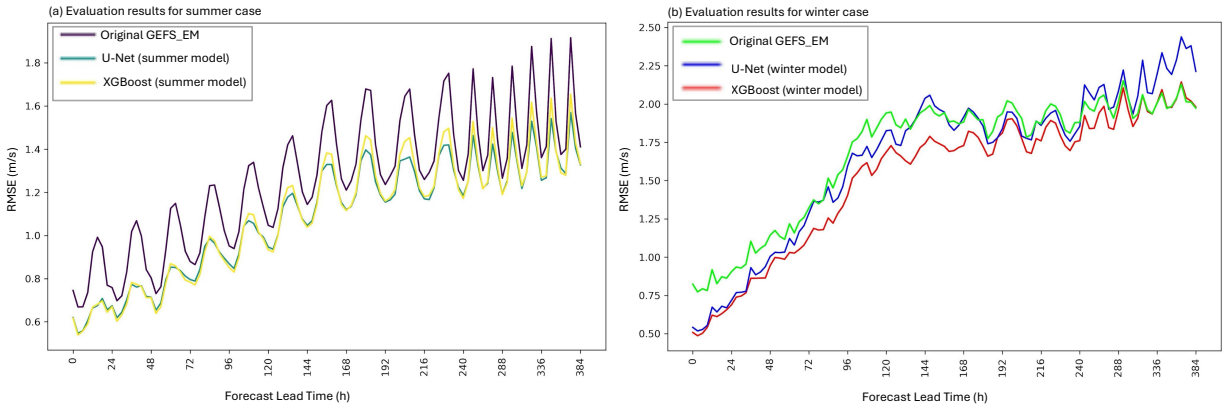

Figure 5.20: Domain-averaged 10m-wind speed (a) RMSE for the 00Z cycle, averaged over August 2022 (b) RMSE for the 00Z cycle, averaged over December 2022

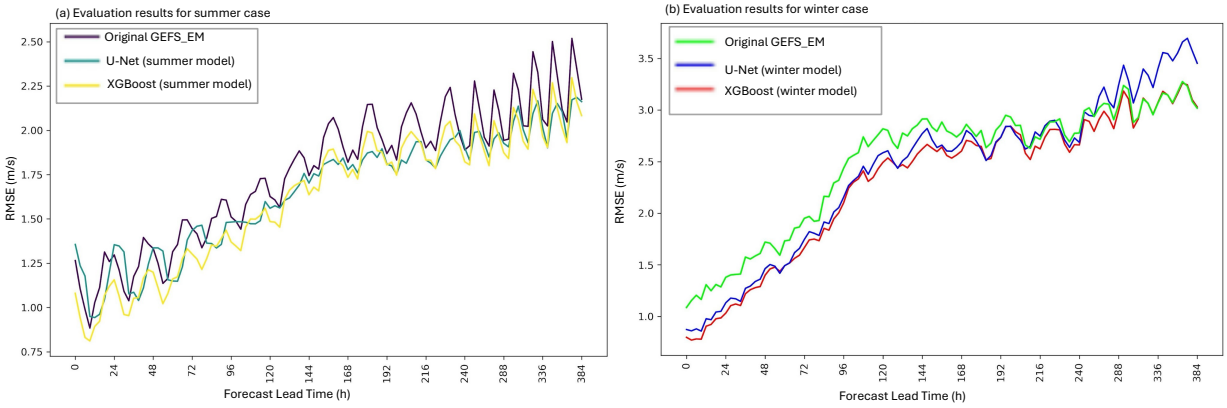

Figure 5.21: Domain-averaged 100m-wind speed (a) RMSE for the 00Z cycle, averaged over August 2022 (b) RMSE for the 00Z cycle, averaged over December 2022

includes the original GEFS ensemble mean forecast (Original GEFS_EM), the forecasts corrected by XGBoost, and those corrected by U-Net. All models are trained using season-specific datasets.

For 10-meter wind speed in summer, the original GEFS exhibits a systematic bias with RMSE ranging from approximately 2.5 to 2.9 m/s across all lead times. XGBoost effectively reduces the RMSE by about 0.3–0.5 m/s and maintains stable performance. U-Net achieves the lowest RMSE in the first 24 hours—reaching as low as 1.9 m/s—demonstrating strong short-term correction capability. Overall, U-Net shows clear advantages in the early lead times, while XGBoost performs more consistently in the mid-to-late range (e.g., 36–48 hours).

For 10-meter wind speed in winter, the original RMSE is higher—around 3.0 m/s—indicating greater forecasting difficulty in winter. XGBoost reduces the RMSE to approximately 2.6–2.8 m/s, though with less spatial consistency. U-Net outperforms XGBoost in the first 30 hours, maintaining RMSE below 2.0 m/s, reflecting its strong modeling capability under complex winter conditions.

For 100-meter wind speed in summer, the original GEFS RMSE remains high, around 3.0–3.2 m/s. XGBoost outperforms U-Net across all lead times, reducing RMSE to around 2.6–2.8 m/s. U-Net performs poorly in this case, with RMSE often comparable to or even exceeding the original forecast, indicating its difficulty in correcting high-altitude wind forecasts during summer. Possible reasons include insufficient training data quality, weaker spatial structure of high-altitude winds, or model complexity mismatch.

For 100-meter wind speed in winter, the original GEFS errors are more severe, with RMSE peaking at 3.5 m/s. XGBoost reduces the RMSE to around 2.8 m/s, showing solid correction performance. Notably, U-Net exhibits significant improvement in winter, with RMSE falling below 2.3 m/s across multiple lead times—especially between 6 and 30 hours—highlighting its superior ability to capture and generalize under complex climatic conditions.

5.4 Discussion

The results presented in this study highlight the comparative strengths and limitations of XGBoost and U-Net in correcting systematic biases in numerical weather prediction (NWP) forecasts. Through evaluations across multiple meteorological variables, seasons, and forecast ranges, several important conclusions can be drawn regarding the effectiveness and applicability of each method.

First, U-Net consistently demonstrates superior spatial correction capabilities, especially for near-surface variables like 2-meter temperature and 10-meter wind speed. Its encoder–decoder architecture with skip connections enables it to preserve fine-scale spatial structures while capturing high-level semantic relationships. This capability proves particularly beneficial in winter conditions and in regions with complex terrain, where spatially correlated forecast errors are more prevalent. U-Net's residual maps show more continuous, smooth corrections compared to XG-

Boost, especially when addressing cold biases in southeastern areas or wind overestimations across topographically diverse regions in Germany.

In contrast, XGBoost offers more stable and interpretable results across forecast ranges and is especially effective for variables with less pronounced spatial complexity, such as 100-meter wind speed during summer. Its performance is generally more robust across mid-to-long forecast lead times. XGBoost achieves moderate correction improvements, particularly in scenarios where systematic biases dominate but where spatial dependencies are less critical. However, its inability to fully capture nonlinear or spatially contiguous error patterns limits its performance in more dynamically complex environments, especially in winter.

When comparing the seasonal models (summer or winter models) to the all-season models, the results showed that the seasonal models were slightly more accurate within their respective time periods. However, the full-season models showed better generalization across seasons and were more consistent over longer forecast horizons. This finding suggests that there is a trade-off between seasonal specificity and overall robustness. While seasonal models may be preferred for short-term, high-accuracy forecasts within known regimes, all-season models are more practical in operational settings with limited training resources or mixed-season applications.

The results also confirm that bias correction is most effective at shorter lead times (0–72 hours), with the benefits tapering off beyond this range. U-Net, in particular, maintains lower RMSE in long-range temperature forecasts compared to XGBoost, indicating better temporal generalization. However, both models face growing difficulty in correcting 100-meter wind speed at extended lead times, reflecting the inherent challenges in modeling upper-atmospheric dynamics.

Overall, this study suggests that U-Net is better suited for spatially complex, surface-level variables and short-term corrections, while XGBoost remains a valuable option for faster training, simpler bias patterns, or ensemble postprocessing tasks. A hybrid approach that leverages the complementary strengths of both models—using XGBoost for rapid, scalable corrections and U-Net for detailed spatial refinement—could offer a promising direction for operational bias correction systems.

5.5 Conclusion

This chapter provided a thorough comprehensive evaluation of two bias correction methods—XGBoost and U-Net—applied to near-surface and upper-level temperature and wind forecasts from GFS and GEFS. The assessment, conducted across different seasons, variables, and forecast lead times, reveals the distinct strengths and limitations of each approach.

U-Net, with its ability to model complex spatial relationships, consistently outperforms XGBoost in correcting 2-meter temperature and 10-meter wind speed, particularly under winter conditions and in regions with complex terrain. Its spatially coherent corrections lead to lower RMSE values and smoother residual fields, showing a better ability to resolve fine-scale atmospheric structures.

XGBoost, however, presents more stable and computationally efficient performance in scenarios with less spatial complexity, such as 100-meter wind speed forecasts. While its corrections are less spatially refined, it remains a valuable tool for rapid deployment and general bias reduction in ensemble and deterministic forecasts.

Season-specific models generally provide better accuracy within their respective seasons, whereas all-season models show more robust generalization across lead times and climatological conditions. Short-range forecasts benefit the most from bias correction, though improvements are still notable in extended forecasts.

Overall, using XGBoost and U-Net for bias correction significantly improves the accuracy of GFS and GEFS_EM forecasts, with U-Net performing better in the forecasts.

Chapter 6

Summary and Future Work

Building on the promising results obtained from XGBoost and U-Net in this study, several directions for future work are identified to further enhance the performance, generalization capability, and operational value of bias correction models.

First, model performance can be improved by expanding the input feature space to include additional atmospheric and geographic variables such as relative humidity, surface pressure, and terrain elevation. For example, adding elevation information may help capture orographic influences.

Second, the U-Net model performs well in spatially complex areas, such as mountainous terrain or coastal environments. This strength is due to its convolutional architecture, which uses feature extraction and skip connections such that it preserves fine-scale spatial details. However, the performance of U-Net can be further improved by transitioning to spatiotemporal architectures such as Transformer, which has shown excellent ability to learn long-range dependencies and temporal dynamics in other domains.

In contrast, XGBoost performs well in data-limited scenarios where samples are limited. Its tree structure enables strong and explainable learning with minimal hyperparameter to tune. XGBoost is particularly valuable when the situation is limited in terms of observation coverage or when we are concerned with high computation costs.

In summary, this study identifies several potential directions to further enhance bias correction models. Expanding the input feature space to include atmospheric and geographic variables (e.g., elevation, humidity) may improve model accuracy. U-Net shows strengths in spatially heterogeneous regions, and its performance could be advanced by adopting spatio-temporal architectures like Transformers. Meanwhile, XGBoost has advantages in data limits due to its efficiency, making it well suited for operational environments with limited observations or computational resources.

Bibliography

[1] A. G. Patt, L. Ogallo, and M. Hellmuth. Sustainability: Learning from 10 years of climate outlook forums in africa. *Science*, 318:49–50, 2007.

[2] N. E. Breuer, C. W. Fraisse, and V. E. Cabrera. The cooperative extension service as a boundary organization for diffusion of climate forecasts: A 5-year study. *Journal of Extension*, 48(4):4rb7, 2010.

[3] A. S. Mase and L. S. Prokopy. Unrealized potential: A review of perceptions and use of weather and climate information in agricultural decision making. *Weather, Climate, and Society*, 6:47–61, 2014.

[4] R. Pandya et al. Using weather forecasts to help manage meningitis in the west african sahel. *Bulletin of the American Meteorological Society*, 96:103–115, 2015.

[5] S. Alexander, E. Atsebeha, S. Negatu, K. Kirksey, D. Brossard, E. Holzer, and P. Block. Development of an interdisciplinary, multi-method approach to seasonal climate forecast communication at the local scale. *Climatic Change*, 162:2021–2042, 2020.

[6] U.S. Department of Energy, Energy Efficiency and Renewable Energy. 20% wind energy by 2030: Increasing wind energy's contribution to u.s. electricity supply. Technical report, U.S. Department of Energy, Energy Efficiency and Renewable Energy, Washington, D.C., n.d. Federal technical report.

[7] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, 29(3):1033–1044, 2014.

[8] Harrison Dreves. How extreme weather and system aging affect the us photovoltaic fleet. https://www.nrel.gov/news/detail/program/2024/

how-extreme-weather-and-system-aging-affect-the-us-photovoltaic-fleet, 2024. National Renewable Energy Laboratory, Jan. 24, 2024. Accessed: 2025-07-01.

[9] S. Rasp and S. Lerch. Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, 146:3885–3900, 2018.

[10] T. Jung and A. M. Tompkins. Systematic errors in the ecmwf forecasting system. Technical Report Technical Memorandum No. 442, ECMWF, Shinfield Park, Reading RG 29AX, U.K., 2003.

[11] F. Zheng, J. Zhu, H. Wang, and R.-H. Zhang. Ensemble hindcasts of enso events over the past 120 years using a large number of ensembles. *Advances in Atmospheric Sciences*, 26:359–372, 2009.

[12] H. R. Glahn and D. A. Lowry. The use of model output statistics (mos) in objective weather forecasting. *Journal of Applied Meteorology and Climatology*, 11:1203–1211, 1972.

[13] Daniel S. Wilks. *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series. Elsevier, 4th edition, 2019.

[14] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82:35–45, 1960.

[15] Eugenia Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2002.

[16] Geir Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer, Berlin, 2009.

[17] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. pages 785–794, August 2016.

[18] J. Fan et al. Evaluation of svm, elm and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of china. *Agricultural and Forest Meteorology*, 263:225–241, 2018.

[19] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, pages 3149–3157, 2017.

[20] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[21] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*, 2015.

[22] A. R. Crimmins and et al., editors. *Fifth National Climate Assessment*. U.S. Global Change Research Program, 2023.

[23] M. C. Peel, B. L. Finlayson, and T. A. McMahon. Updated world map of the köppen-geiger climate classification. *Hydrology and Earth System Sciences*, 11(5):1633–1644, 2007.

[24] C. D. Whiteman. Mountain climates of north america. In *Mountain Meteorology: Fundamentals and Applications*. Oxford University Press, 2000.

[25] NOAA/NCEP Environmental Modeling Center. Global Forecast System (GFS). https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php, 2025. Accessed: 2025-07-01.

[26] National Institute of Environmental Health Sciences. Fifth national climate assessment released. Environmental Factor, Dec. 2023, Dec 2023. Accessed: 2025-07-01.

[27] P. E. Bett, H. E. Thornton, and R. T. Clark. European wind variability over 140 yr. *Advances in Science and Research*, 10(1):51–58, April 2013.

[28] Deutscher Wetterdienst (DWD). Dwd — deutscher wetterdienst. https://www.dwd.de/EN/Home/home_node.html, 2025. Accessed: 2025-07-01.

[29] K. Chapman. Climate zones of germany: Different climate regions of germany. 2024. Accessed: Dec. 20, 2024.

[30] Qianya Zhu and Haonan Chen. Bias correction of wind forecasts from the noaa global ensemble forecast system (gefs) using machine learning. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2025. Accepted for presentation.

[31] Sara Pryor and R. Barthelmie. Climate change impacts on wind energy: A review. *Renewable and Sustainable Energy Reviews*, 14:430–437, 01 2010.

[32] Philip E. Bett, Hazel E. Thornton, and Robin T. Clark. Using the Twentieth Century Reanalysis to assess climate variability for the European wind industry. *Theoretical and Applied Climatology*, 127(1-2):61–80, January 2017.

[33] NOAA National Centers for Environmental Prediction. Global forecast system (gfs), 2025.

[34] NCEP. Ncep gfs 0.25 degree global forecast grids historical archive. https://doi.org/10.5065/D65D8PWK, 2015. Accessed Jan. 28, 2025.

[35] NOAA/NCEP Office Note. NCEP PMB (Production, Maintenance, and Backup) Changes. https://www.nco.ncep.noaa.gov/pmb/changes/, 2025. Accessed: 2025-07-01.

[36] National Centers for Environmental Information (NCEI). Global ensemble forecast system (gefs). 2025. Accessed: Jan. 6, 2025.

[37] X. Zhou, Y. Zhu, D. Hou, B. Fu, W. Li, H. Guan, E. Sinsky, W. Kolczynski, X. Xue, Y. Luo, J. Peng, B. Yang, V. Tallapragada, and P. Pegion. The development of the ncep global ensemble forecast system version 12. *Weather and Forecasting*, 37(6):1069–1084, 2022.

[38] NOAA Environmental Modeling Center / AWS Registry of Open Data. NOAA Global Ensemble Forecast System (GEFS). https://registry.opendata.aws/noaa-gefs/, Jul 2025. Accessed: 2025-07-01.

[39] A. J. Clark, K. A. Hoogewind, A. J. Hill, E. D. Loken, and M. J. Hosek. Extended range machine-learning severe weather guidance based on the operational gefs. *Weather and Forecasting*, 2025. Published online ahead of print.

[40] M. Leutbecher and T. N. Palmer. Ensemble forecasting. 227(7), 2008.

[41] H. Hersbach et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146:1999–2049, 2020.

[42] D. X. He, Z. M. Zhou, Z. P. Kang, and L. Liu. Numerical studies on forecast error correction of grapes model with variational approach. *Advances in Meteorology*, page 2856289, 2019.

[43] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013.

[44] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *CoRR*, abs/1609.05158, 2016.

[45] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.