Integrating Deep Learning with Numerical Models to Improve Weather and Air Quality Forecasts

by
Alqamah Sayeed

A dissertation submitted to the Department of Earth and Atmospheric Sciences,
College of Natural Sciences and Mathematics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Atmospheric Sciences

Chair of Committee: Yunsoo Choi

Committee Member: Xun Jiang

Committee Member: Bernhard Rappenglueck

Committee Member: Kyo Lee

University of Houston
August, 2021

DEDICATION

This dissertation is dedicated to my parents, QamarJahan and Abdul Sayeed; their unyielding love, support, and encouragement have enriched my soul and inspired me to pursue and complete this research.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Yunsoo Choi, for the continuous support of my Ph.D. course work and related research, for his patience, motivation, and immense knowledge. His guidance has helped me in improving my research and writing skills. I would never have imagined having a better advisor and mentor for my Ph.D. studies.

Besides my advisor, I would like to thank the rest of my thesis committee members: Dr. Bernhard Rappenglueck, Dr. Xun Jiang, and Dr. Kyo Lee, for their insightful comments and encouragement that enabled me to widen my research from various perspectives.

I especially thank my parents, QamarJahan and Abdul Sayeed, brother, Alqamar, and sister, Talha; for their continued support and faith in me. My hard-working parents have sacrificed their lives for us and provided unconditional love and care. I would not have made it this far without them. I especially thank my brother and sister for taking care of our parents when I was away pursuing my dream; without them, this thesis would not have been possible.

I also thank my friend and confidant, Manjusmriti, who was there for me at my weakest and the strongest. She has been a true and great supporter and; unconditionally stood by me during my good and bad times. She has been non-judgmental of me and instrumental in instilling confidence. She has always had faith in me and my intellect, even when I did not have faith in myself. Completing my graduate studies would not have been possible without her never-ending support.

I thank my fellow lab-mates for all the work we did together before deadlines and for all the fun we had in the last four years. In particular, I am grateful to my AI team members, Dr. Ebrahim Eslami, Yannic Lops, and Ahmed Khan Salman, for the countless stimulating discussions we had regarding various exciting research topics in the field of machine learning and atmospheric science. Additionally, I would like to thank Dr. Anirban Roy, Jia Jung, Bavand Sadeghi, and Arman Pouyaei for working together on various projects. I express my thanks to the other EAS faculty members and graduate students for discussions on different interesting research topics.

Last but not least, I would like to thank my friends for supporting me spiritually throughout writing this thesis and my life in general. This dissertation would not have been possible without their warm love, continued patience, and endless support.

ABSTRACT

Numerical models are excellent tools for forecasting future weather events or air quality. The scientific and computational advancements in numerical models provided us with more accurate forecasts while promoting an understanding of various physical and chemical processes in the atmosphere. However, due to the simplified implementation of such processes, the understanding of modeling uncertainties has been limited. In addition to this, these numerical models require significant computational resources and time to have a quality forecast. This thesis tries to mitigate the uncertainties which lead to large biases, using advanced deep neural network (DNN) models and integrate them into the existing numerical model to have better and faster weather and air quality forecasts.

In this study, a long-term forecasting system based on a deep Convolutional Neural Network (CNN) was developed for air pollutants such as ozone, NO_2, $PM_{2.5,}$ and PM_{10} for up to two weeks. An optimized deep learning algorithm was used to develop species-specific and location-neutral models. These models used the simulation outputs of the Weather Research and Forecasting (WRF) model and Community Multiscale Air Quality (CMAQ) model at the first step, and then trained a deep neural network (DNN) model for each air pollutant in the second step. Once trained, these models forecasted the next 24-hour in advance for all species across the geographical domain for up to two weeks.

In the final task, a deep CNN system was developed to bias-correct and reduce systematic uncertainties in the simulation of meteorology, such as wind speed and direction, surface pressure, temperature, humidity, etc., in the WRF model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The air quality of any region is strongly influenced by the weather conditions of the region. The interaction of emitted or residing chemical species or particulates with the prevailing weather condition decides the region's air quality. Meteorological processes in chemical transport models (CTMs) like Community Multiscale Air Quality (CMAQ) models are derived from the partial differential equation of mass and energy conservation. The process also involves parameterization of the various process to have simplified physics.

The main objective of this dissertation was to develop techniques that can mitigate the uncertainties in weather and CTM models. In this dissertation, several artificial intelligence models based on the deep convolutional neural network (CNN) were developed to forecast and bias-correct hourly air quality and meteorological parameters. The models developed to bias-correct and forecast hourly ozone, NO_2, $PM_{2.5,}$ and PM_{10} for up to two weeks in advance were discussed in Chapters 3, 4, and 5. CMAQ air-quality and Weather Research and Forecasting model's (WRF) meteorological parameters were used to develop these CNN-based models. In Chapter 6, the models based on CNN to estimate the 24-hour meteorological parameters (such as temperature, pressure, precipitation, wind speed and direction, relative humidity, etc.) with better accuracy over the spatial domain using the WRF simulations were discussed.

1.1. Air quality

The severe implications of high concentrations of air pollutants such as ozone $PM_{2.5}$, PM_{10}, NO_x, etc., to human health and the environment; necessitate prior reporting of such concentrations. Numerical air quality modeling, also referred to as chemical transport modeling (CTM), was often used for this purpose. The most widely used air pollution chemical transport model was the

Community Multiscale Air Quality (CMAQ) model, developed by USEPA (Byun and Schere, 2006). These models cover large spatial domains and have reasonable accuracy for a one-day ahead forecast (Chai et al., 2013).

Numerical models such as CMAQ are excellent tools to forecast the air quality of a region with considerable accuracy. CMAQ was a three-dimensional model that provides the estimation across the domain with coarse to fine spatial and temporal resolutions. It has been used as a primary dynamical model in regional air pollution studies. However, some concerns remained in CMAQ modeling. CMAQ has uncertainties in estimating ozone, leading to large overestimations (Li et al., 2016; Liu et al., 2010).

The process to forecast air quality with numerical modeling requires significant computational time - even with simplified physics. The computation time increases in two-way modeling (online coupling of WRF and CMAQ) processes. Despite their greater domain coverage and reasonable results, the CTM models consume significant computational resources and time (Zhang et al., 2012). This necessitates the use of statistical and/or artificial intelligence techniques, such as deep neural networks (DNN), which are considerably more efficient and consume fewer computational resources and time (Fernando et al., 2012).

Generally, such a model uses an artificial neural network (ANN) that can be trained from historical events that form the input and output set (perceived outcome based on given inputs). Figure 1.1 shows the schematic of the general process in a neural network. Since all atmospheric phenomena are interrelated, the DNN then predicts a future event(s) based on a given set of new inputs (or unseen inputs) (Bengio, 2009; Marr, 1976). Machine learning (ML), on the other hand, can be trained to forecast one hourly output using certain inputs. Another computational benefit of machine learning was the model only needs to be trained once. While the machine learning model has higher accuracy and faster processing speed, they are very localized (station-specific) and have a large under-prediction of the daily maximum ozone concentrations (Eslami et al., 2019a; Sayeed et al., 2020b).

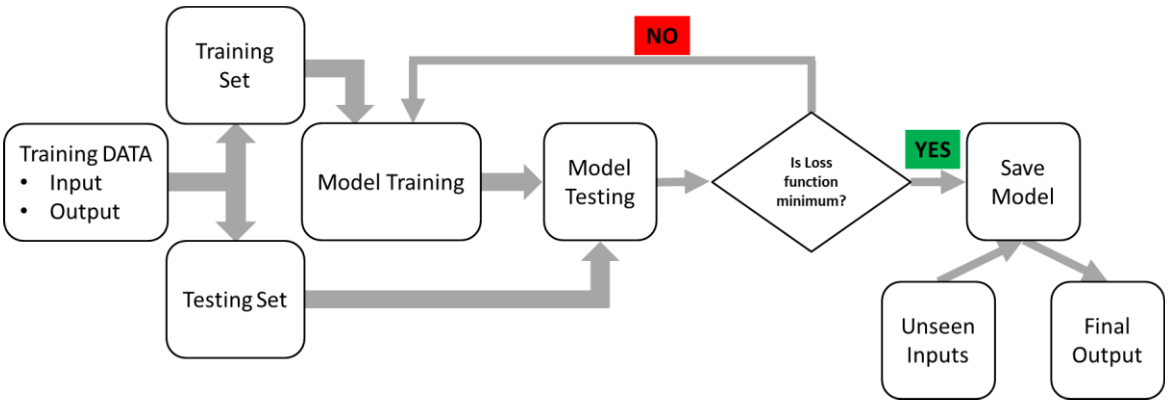


Figure 1.1. *Process Flow Diagram of a Machine Learning Model.*

Although the ANN models are fast, they have the following issues: i) Most predict a single value (single hour of the day, the daily mean, the daily maximum, etc.), ii) their predictions are sometimes not comparable with numerical models, and/or iii) have an only temporal or spatial resolution. Hence, there was a need to develop a fast, stable, and accurate model covering both spatial and temporal domains. To develop such a model, the deep architecture of Convolutional Neural Network (CNN) (Krizhevsky et al., 2017; Lawrence et al., 1997; Lecun and Bengio, 1995) can be used. CNN was suitable for this study because of three qualities: i) it was capable of understanding the complex features of input variables by applying a convolution using multiple filters and kernels (Lecun and Bengio, 1995); ii) since data (both meteorological and pollutant data) exhibit temporal coherence that CNN preserves by convolution on adjacent inputs only (Lawrence et al., 1997), it provides the desired accuracy in the results; and iii) it can cover both spatial and temporal domain.

The objective of using this technique was to enhance the CMAQ modeling results by taking advantage of ML as a computationally efficient system to recognize the uncertainties in the CMAQ model, as well as computing the measured chemical variables along with fine temporal and spatial resolutions. This approach aimed to use the best of both numerical modeling and machine learning to design a robust and stable algorithm to forecast air quality with better accuracy in predicting concentration and to cover a larger domain, both spatially and temporally.

1.2. Meteorology

Many academic studies have been devoted to the problem of forecasting difficult-to-retrieve weather events and their associated uncertainties. Methods used for weather forecasting can be classified into dynamical and statistical groups. Dynamical (physical) models such as WRF use meteorological and topological information to determine the weather parameters of a specific region. Statistical methods mainly use historical meteorological data to forecast the future state of the weather. Despite the major progress of numerical weather prediction (NWP) in the last several decades, meteorological models are unable to provide fully reliable weather forecasts, especially in topographically complex regions, because of shortcomings in horizontal resolution, physical parameterizations, and initial and boundary conditions (Cassola and Burlando, 2012). They are also computationally expensive, particularly with regard to fine-resolution forecasting. In addition, because of the misrepresentation of unresolved small-scale features or neglected physical processes, parts of numerical models have to be represented by empirical sub-models or parameterizations. Despite their unreliability for longer-term forecasting, statistical methods are popular because they are easily implemented and less computationally intensive than NWPs. Owing to the chaotic nature of the weather system, errors in weather forecasting are unavoidable but quite often significant regardless of the implemented modeling approach.

An alternative approach was be applied to estimate meteorology, which was a fully data-driven framework that combines deep neural networks and physical models that simulate the dynamics of a complex weather system. Weather-AI as a gridded real-time weather forecasting model was developed that reduces the model-measurement error of the WRF model. The system post-processes WRF simulation output based on an observation network 24 hours ahead in real-time. It uses a convolutional neural network algorithm (Krizhevsky et al., 2017) that bias-corrects the WRF outputs at each grid linked to each station location.

CHAPTER 2

METHODOLOGY

2.1. Coupled CMAQ and WRF

To take advantage of numerical modeling, air quality and meteorological parameters were obtained by the CMAQ v5.2 (Byun and Schere, 2006) and the Weather Research and Forecasting (WRF) v3.8, covering the eastern part of China, the Korean Peninsula, and Japan, with a 27 km spatial resolution. The detailed configurations of the CMAQ and WRF models are available in Jung et al. (2019).

WRF v3.8 covers the eastern part of China, the Korean Peninsula, and Japan, with a 27 km horizontal grid spacing for the years 2014 to 2018. Detailed configurations of the WRF model are available in (Jung et al., 2019). The WRF simulation used for this study was conducted in hindcast mode by using one-degree by one-degree National Centers for Environmental Prediction FNL (final) operational global analysis data as initial and boundary condition, as well as 0.5-degree real-time global sea surface temperature (RTG SST) for a reasonable sea surface temperature. Thus, a four-dimensional data assimilation (FDDA) option every 6 hours for the temperature, the water vapor mixing ratio, and wind components was applied in conjunction with the indirect soil moisture and temperature nudging technique (Pleim and Gilliam, 2009; Pleim and Xiu, 2003).

2.2. Deep Convolutional Neural Network

As introduced by Lecun and Bengio (1995), the CNN was generally used for image classification. However, since its inception, CNN has improved significantly and has been used in various applications, from speech recognition to object detection (LeCun and Bengio, 1995). Deep CNN consists of several stacked neuron layers: 1) a convolutional layer, 2) a pooling layer, and 3) fully connected layers (Krizhevsky et al., 2012). These layers build a hierarchal and distributed network. The convolutional layer consists of neurons stacked together and responds only to the overlapping region instead of the whole signal. Mathematically, convolution was integral measuring the extent to which two functions overlap as one passes over the other. The pooling layer eliminates features with similar attributes, thus reducing the computational burden. In this layer, features are either averaged (i.e., average pooling) or maximum (i.e., MaxPooling) over one section of a signal region to increase the system's robustness by decreasing the number of features. The fully connected layer translates input features into output predictions. Figure 2.1 shows the schematic diagram of the CNN model architecture used in this study.

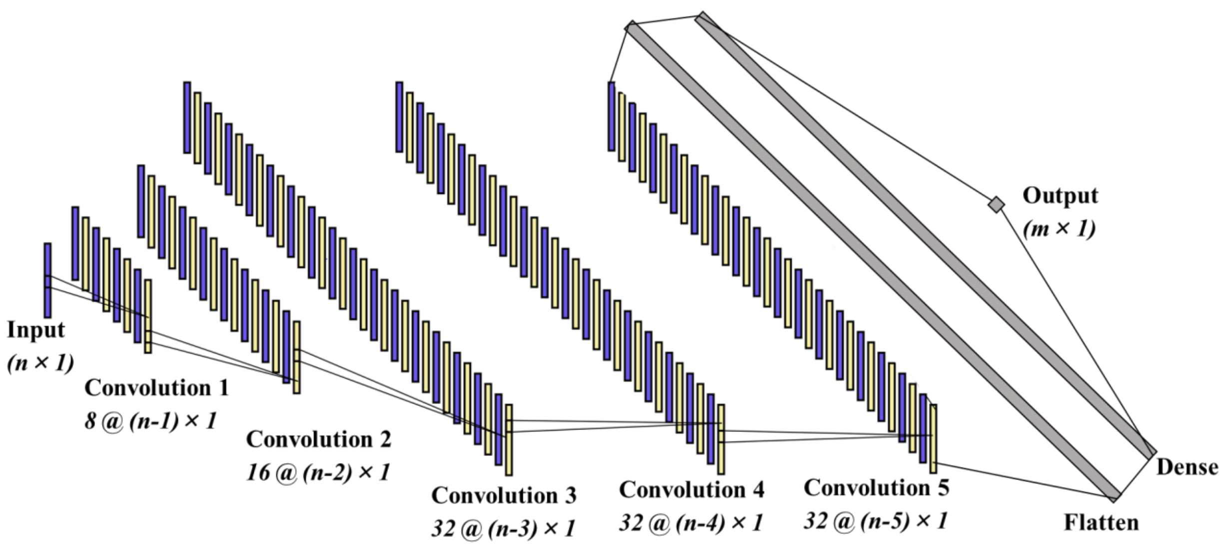


Figure 2.1: *Schematic representation of the architecture of the CNN model.*

Pooling can potentially remove important features sensitive to the output (e.g., a sudden change in peak due to the absence of favorable chemistry in the atmosphere) (Eslami et al., 2019a, 2019b). Also, in contrast to an image or speech signal, the number of features (meteorology and air pollution) in this study was limited (Eslami et al., 2019a, 2019b). The representation of each feature was essential; therefore, to avoid loss of information, the pooling layer (Hinton, 2014) was not used. This was confirmed during the test runs – where the addition of a pooling layer reduced the model's performance.

The depth of any neural network depends on the type of problem it needs to solve. Ozone concentration was highly dependent on the availability of favorable meteorological conditions (e.g., sunlight) and the presence of certain chemical species (e.g., concentrations of volatile organic compounds and NOx concentrations) in the atmosphere. Additionally, the transport of pollutants from the surrounding region can lead to a change in their concentration. These factors add high non-linearity in the time-series of pollutants concentrations. Since the problem here was highly non-linear and dependent upon the external factors, the architecture was tested with more than one layer of CNN with various configurations (e.g., number of layers and kernel size). Then the best configuration was selected with the least mean squared error and/or index of agreement (IOA) on the cross-validation set and maximum IOA on the test set. Note: the test set throughout the study was never used for training to avoid data leakage.

The CNN model consists of five convolutional layers and one fully connected layer, as shown in Figure 2.1. A convolution was applied to the input features and the elements of the kernel. The final feature map obtained at the end of the first layer of the CNN acts as input for the second layer. Similarly, the output feature map of the second layer was input for the third layer, and so on. In this way, the model has a five-layer CNN, each layer with 32 filters (activation by ReLU), each with a size two kernel randomly initialized by some value for the first iteration. After determining the last feature maps in the last convolutional layer, the fully connected hidden layer with 264 nodes provides the 24-hour output of target pollutant or meteorology. The algorithm was implemented in the Keras environment with a TensorFlow backend. (Abadi et al., 2016; Chollet and others, 2015)

CHAPTER 3

USING A DEEP CONVOLUTIONAL NEURAL NETWORK TO PREDICT 2017 OZONE CONCENTRATIONS, 24 HOURS IN ADVANCE[1]

3.1. Introduction

Ozone was a criteria pollutant generated by the photochemical reaction between nitrogen oxides (NO_x) and volatile organic compounds (VOCs) in the atmosphere (US-EPA, 2006). Depending on its concentrations, ozone can have serious health implications. In general, it was advisable that hourly ozone concentration limits should not exceed 80 ppbv and/or 50-60 ppbv for a maximum daily eight-hour average (MDA-8) (Ayres et al., 2006; Taylan, 2017). In urban areas, short-term exposure to ozone can cause headaches, chest pains, sore throats, coughing, and decreased lung function, depending on the ozone concentration (usually > 150-300 ppbv). At elevated ozone concentrations above 100 ppbv, living beings are more susceptible to bacterial infections (Jacobson, 2005). According to Bell et al. (2006), exposure to high concentrations of ozone and long-term exposure to low concentrations of ozone adversely affect human health. Similarly, Mills et al. (2007) claimed that long-term exposure to more than 40 ppbv of ozone could damage crops and ecosystems.

The human health hazards related to ozone exposure lead to the development of several models to estimate it. Byun et al. (2007) performed CMAQ simulation for base and HRVOC (a highly reactive volatile organic compound) emissions in the Houston-Galveston-Brazoria region (HGB), Texas, USA, during a five-day summer period of TexAQS-2000 and found that HRVOC produced better results (slope, a = 0.81 ppbv; intercept, b = 5.48 ppbv, r^2 = 0.76). Misenis and

[1]Published: Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., Jung, J., 2020b. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. Neural Networks.

Zhang (2010) used the WRF/Chem model during the same period and region. They ran simulations in various configurations and reported an hourly ozone prediction mean bias range between 3.8 ppbv and 11.0 ppbv. Using the CMAQ modeling system, the NAQFC (National Air Quality Forecast Capability) project (Chai et al., 2013) generated operational and experimental real-time ozone predictions for the United States in 2010 and found that the model overestimates by 5.6 ppbv annually for the contiguous US (CONUS) and that the RMSE was 15.4 ppbv. They also found out that among the six regions of the study, the southeastern US had a maximum bias of 10.5 ppbv, and the lower middle US (Texas, Oklahoma, Arkansas, Louisiana) had a minimum bias of 3.7 ppbv. Czader et al. (2015) developed another chemical transport model, STOPS (v1.0), a hybrid Eulerian-Lagrangian-based modeling tool, and compared its results with those of the CMAQ for August 25, 28, and 30, in the year 2000, for the Houston region. They found that the mean bias for surface ozone mixing ratios varied between -0.03 and -0.78 ppbv, and the slope varied between 0.99 and 1.01 for several configurations. Another study by Pan et al. (2017) simulated NO_x and ozone for the Houston region for September 5-14, 2013, with a spatial resolution of 4 km and 1 km and found correlations for ozone concentrations in the range of 0.79 to 0.87 and an index of agreement (IOA) of 0.74 to 0.86. The model, however, generated stronger correlations and IOA mean bias (MB): 10 to 14.9 ppbv and a mean absolute error (MAE) of 10.7 to 15.2 ppbv. Despite their greater domain coverage and reasonable results, the above models consume significant computational resources and time (Zhang et al., 2012). To overcome the computational burden and improve the performance, alternative statistical and machine learning (ML) approaches were used.

These ML models are trained on a historical event having certain inputs and outputs. As all atmospheric phenomena are interrelated, the ML then predicts a future event(s) based on a given set of new inputs (or unseen inputs) (Bengio, 2009; Marr, 1976). Hoshyaripour et al. (2016) developed an FS-ANN model, compared it with WRF-Chem, and evaluated it in two stations in Sao Paulo (Brazil) between August 5 and 20, 2012. They found that while the daily mean IOA ranged from 0.39 to 0.59 (at various locations) from FS-ANN and 0.54 to 0.68 from WRF-Chem, the daily peak value was slightly higher (0.53 to 0.78 from FS-ANN and 0.63 to 0.67 from WRF-Chem). These results also showed a mean daily bias of between -3.49 and 1.60 ppbv from FS-ANN and between -8.05 and 4.25 ppbv from WRF-Chem. They concluded that WRF/Chem produces better results than FS-ANN, but the latter was computationally faster and cheaper.

Prasad et al. (2016) developed an adaptive neuro-fuzzy inference system (ANFIS) for Howrah City, India, for the years 2009 to 2011, reported an IOA of 0.81 and R^2 of 0.51 for one-day advance forecasting of ozone concentrations. Biancofiore et al. (2015) used a recurrent neural network (RNN) to predict one-, three-, six-, 12-, 24-, and 48-hour ozone concentrations at an observation station in Pescara, Italy, in 2005. Although their model performed reasonably, with comparatively stronger correlation coefficients for the one-, three-, 24-, and 48-hour concentrations (i.e., 0.97, 0.89, 0.86, and 0.83 respectively), it yielded poor correlation coefficients for the six- and 12-hour concentrations (i.e., 0.78 and 0.77, respectively). The main reason for the poor six- and 12-hour correlations was that both meteorological and pollutant criteria differ significantly (or reverse in the case of 12-hour) from the 0^{th} hour while the stronger 24- and 48-hour correlation coefficients occur at the same time of day. The other caveat in their model was that they predicted one specific hour of the day per iteration.

Although the above models are fast, they have either one or both of the following issues: i) Most predict a single value (single hour of day, daily the mean, the daily maximum, or MDA-8) as in Biancofiore et al. (2015), and/or, ii) their predictions are inaccurate, as in Hoshyaripour et al.

(2016). A fast, stable, and accurate model can overcome these problems. To develop such a model, the deep architecture of CNN (LeCun and Bengio., 1995; Lawrence et al., 1997; Krizhevsky et al., 2012) was used. CNN was suitable for this study because of two qualities: i) it was capable of understanding the complex features of input variables by applying a convolution using multiple filters and kernels (LeCun and Bengio, 1995); and ii) since data (both meteorological and pollutant data) exhibit temporal coherence that CNN preserves by convolution on adjacent inputs only (Lawrence et al., 1997), it provides the desired accuracy in the results. Therefore, in this study, an artificial intelligence model based on a deep convolutional neural network to predict next-day 24-hour ozone concentrations at a station was developed and then the model was evaluated based on daily, weekly, monthly and seasonal values for the year 2017. (Note: The model was trained for the years 2014 to 2016).

3.2. Material and Methods

A five-layer deep CNN architecture model was used to estimate real-time 24-hour predictions of ozone concentrations. The main purpose of using a multilayer CNN was to increase computational efficiency over traditional numerical models and various deep learning models. When compared with multilayer perceptron (MLP), it involves more thorough and complex calculations to preserve nonlinear characteristics of the input and output features. Several tests were performed to determine the optimal number of layers (in this case, five) and achieve the lowest mean squared error (MSE). This technique involved site selection, observations, and model definition, training, and prediction. These steps are explained in the following subsection.

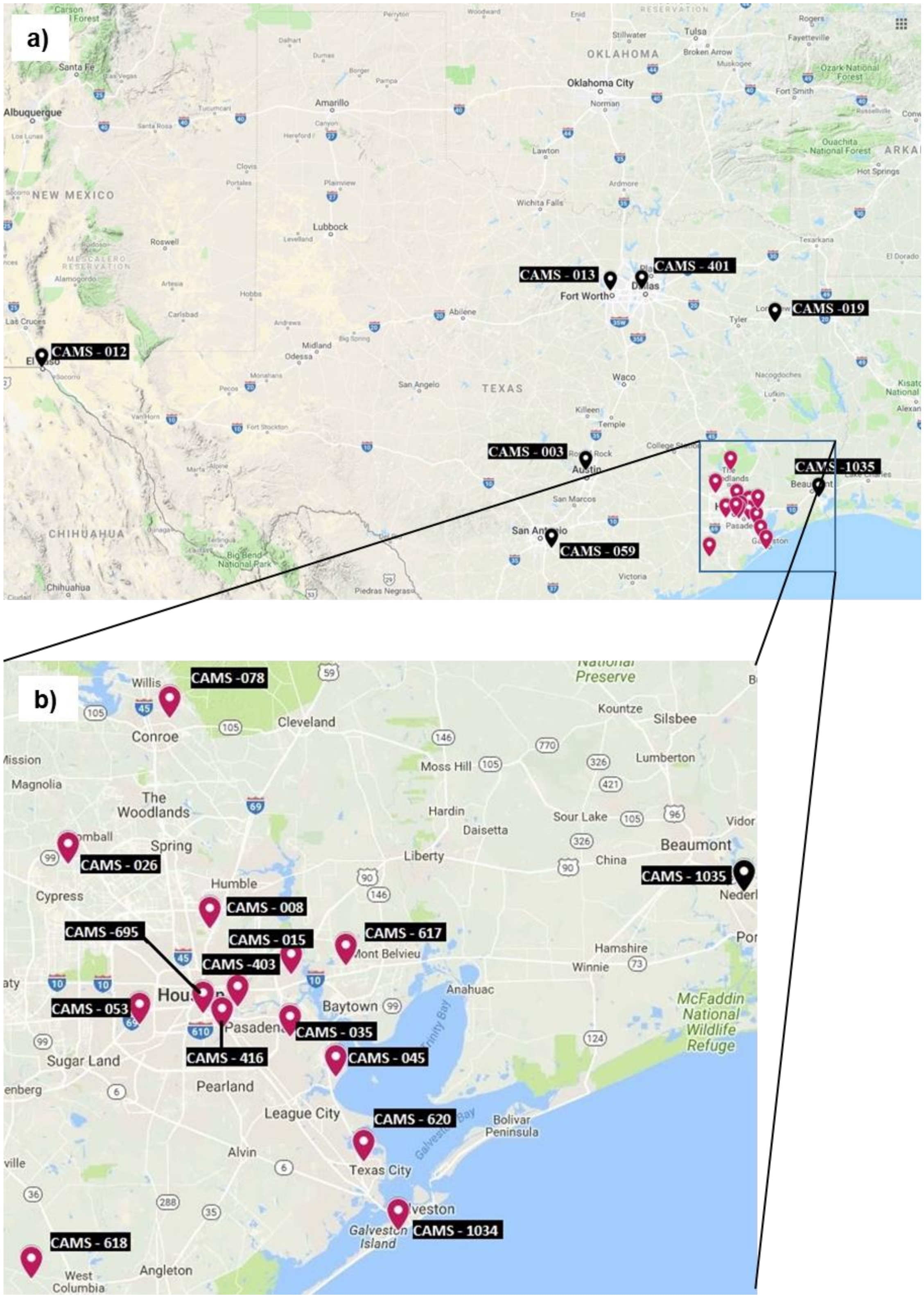


Figure 3.1. *Map of Texas (a) indicating locations of all stations used for the study. Map of Houston (b) shows the individual stations used in Houston for the study.*

3.2.1. Site Selection and Observation

Anthropogenic precursors, along with meteorological conditions, significantly affect regional air quality (Baklanov et al., 2008; Damo and Icka, 2012.; Taylan, 2013). Thus, only those stations were selected in Houston with both ozone and NOx data available and at least three meteorology measurements monitored for at least four years (i.e., 2014 to 2017). Figure 3.1 shows the station location (see Table-T1 in the appendix for station details). In addition, seven stations were selected from other cities in Texas to evaluate the model performance outside of Houston. A four-year (2014-2017) observational data were obtained from the Texas Commission on Environmental Quality (TCEQ) website, which provides a variety of meteorological (e.g., wind speed, wind direction, temperature, pressure, precipitation, dew point temperature, relative humidity, solar radiation) and air pollutant (i.e., NO_x and ozone) data.

3.2.2. Model Definition

To build the deep CNN model, five convolutional layers and one fully connected layer were used (Figure 3.2). A convolution was applied to the input features and the elements of the kernel (Figure 3.3). Since the kernel size (convolution window) here was 2×1, the convolution of two successive hours of input features takes place in the first layer. The results of the convolution operation were then passed to the activation function. The final features are the activation function (ReLU) applied to the output of the convolution (i.e., the three-dimensional tensor).

For any neural network to achieve efficient optimization within a weight matrix while preserving nonlinearity, it needs to have an activation function (Nair and Hinton, 2010). For this model, ReLU was used as the activation function, defined by equation 3.1 as follows:

$$f\ (x) = \max(0, x). \tag{3.1}$$

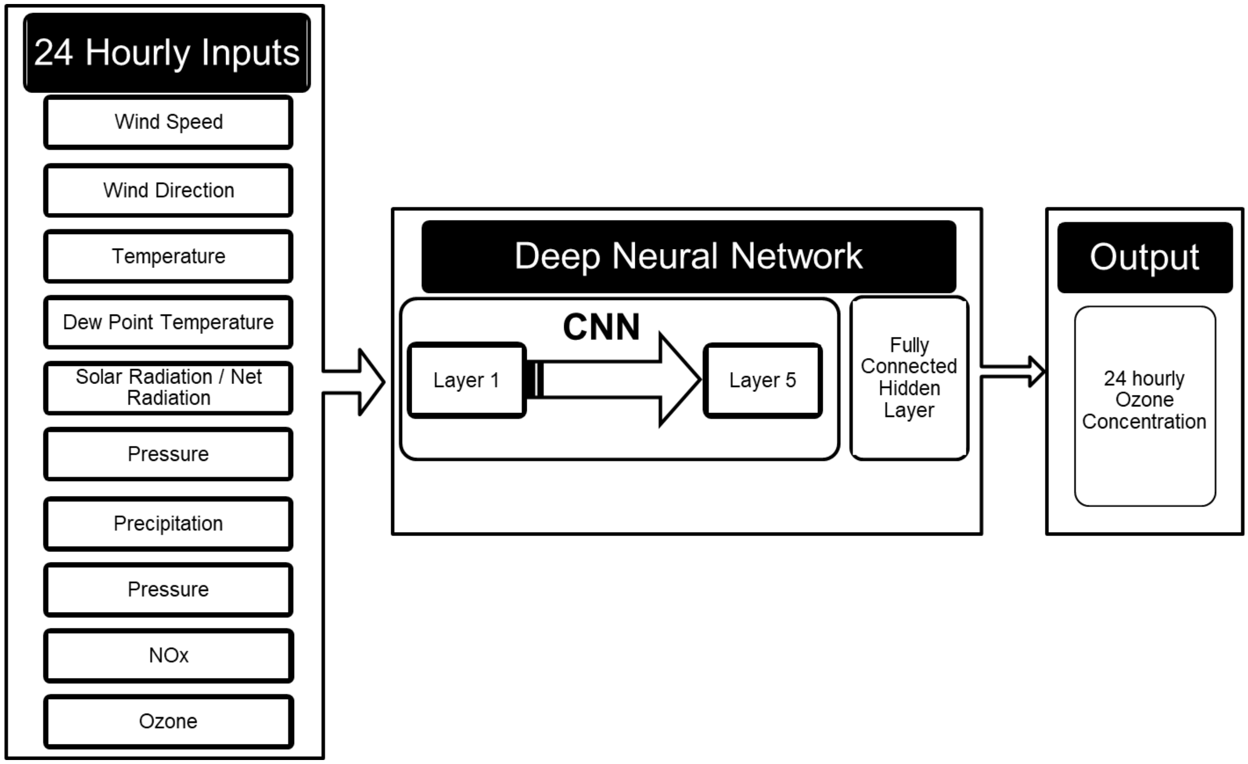


Figure 3.2. *Model Architecture: Detailed process flow of the deep CNN model.*

The final feature map obtained at the end of the first layer of the CNN acts as input for the second layer. Similarly, the output feature map of the second layer was the input for the third layer, and so on. In this way, the model has a five-layer CNN, each layer having 32 filters (activation by ReLU), each with two kernels randomly initialized with some value for the first iteration. After determining the feature maps in the last convolutional layer, a fully connected hidden layer with 264 nodes, gives the final output of the model. The algorithm was implemented in the Keras environment with a TensorFlow backend (Abadi et al., 2016; Chollet and others, 2015).

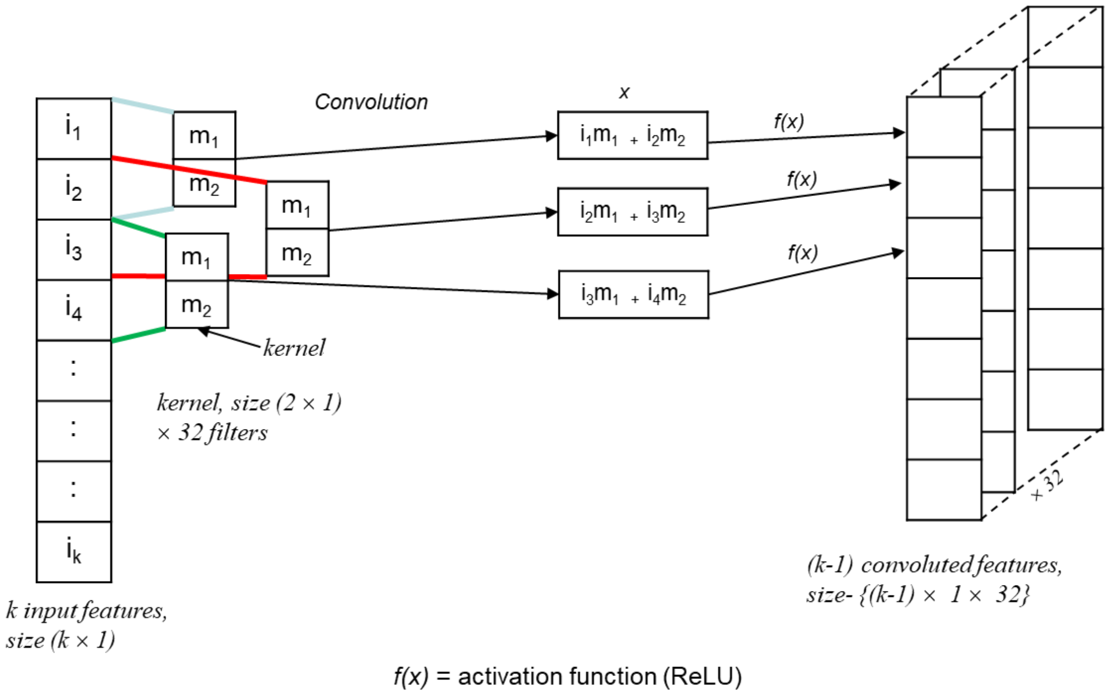


Figure 3.3. *Expanded view of the first CNN layer (the architecture has four more layers with a similar structure): Operations performed in a convolutional layer of CNN. (k was the total number of input features, whereas "i" was the input feature. m1 and m2 are the elements of each kernel that are randomly initialized).*

3.2.3. Model Training and Prediction

Once the model architecture was defined, it required a training set consisting of input and output features from the previous day. For example: to predict the $(n+1)^{th}$ day, the model was trained until the n^{th} day with the input feature of the $(n-1)^{th}$ day and the output target of the n^{th} day. Therefore, to predict the $(n+1)^{th}$ day, the model has n training examples (for details, see Section C - Experimental set-up in the appendix). The model was then trained by the greedy layer-wise algorithm (Bengio et al., 2006). Instead of optimizing the model in a single step, it was divided into several stages. The algorithm trains each model layer by layer. Initially, all the layers are frozen, and only the first layer was trained. Then the model assigns a weight to each input feature and computes the output, which was then compared to the actual observations and the MSE was calculated. Depending on the MSE, the model changes the weight of the input feature and computes a new output. Again, this output was compared with actual observations and the MSE was calculated. Once the MSE was minimized, the model was said to be trained. Once the first layer was trained, the second layer was added, and weights are preserved from the previous layer. The entire process was repeated once again on the two-layer network, which was then applied to the third layer, again preserving the weights from the two-layer model, and so on. This process optimizes the weights in a computationally efficient manner.

3.3. Results and Discussion

The observed values of the 21 stations were obtained from the TCEQ CAMS stations for the year 2014-2017. The model was then trained for 2014-2016 (keeping 20% data for cross-validation for each deep learning and machine learning model). The model was then used to make predictions for the entire year, updating each day (i.e., adding the previous day in the training set). For example, initially, the model was trained until December 31, 2016, and predicted ozone concentrations for January 1, 2017. Then the observations from January 1, 2017, were updated to the training set and again trained the model and then predicted ozone concentrations for January 2, 2017, from the model. This process was repeated for 365 days. Once predicted ozone concentrations for all days of 2017, each station was evaluated for the hourly and maximum daily eight-hour average (MDA-8) values.

3.3.1. Evaluation Based on Hourly Values

Figures 3.4 and 3.5 show the daily mean of the ozone concentrations over all stations. The results indicate the model slightly under-predicted ozone concentrations for most of the year. The maximum under- and over-prediction of the model were -15.5 ppbv and 12.9 ppbv, respectively. The monthly mean biases were in a range of -2.43 to 2.10 ppbv. The model underpredicted monthly mean biases in January, February, March, April, May, and September but overpredicted them in the remaining months. The IOA (Willmott et al. 1981) of the ozone concentration was 0.89, and the correlation was 0.81 for all stations.

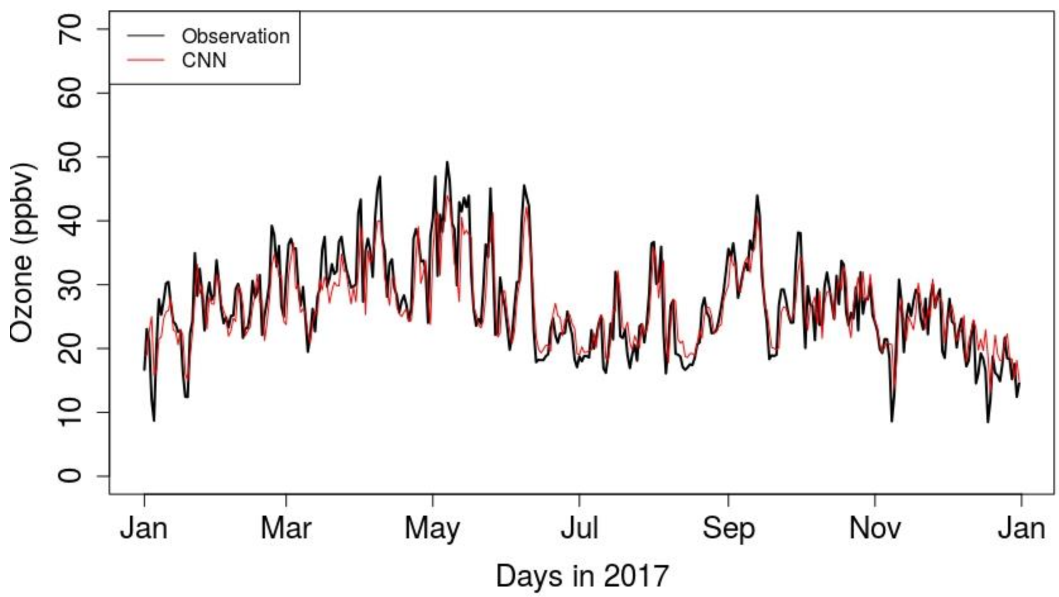


Figure 3.4. *Model-measurement comparisons for a daily mean of concentrations averaged over all stations mean for the year 2017. Y-axis represents daily mean ozone concentration, and the x-axis was days of the year. Black solid lines are observed ozone concentration (in ppb), and red lines are predicted ozone concentration (ppb).*

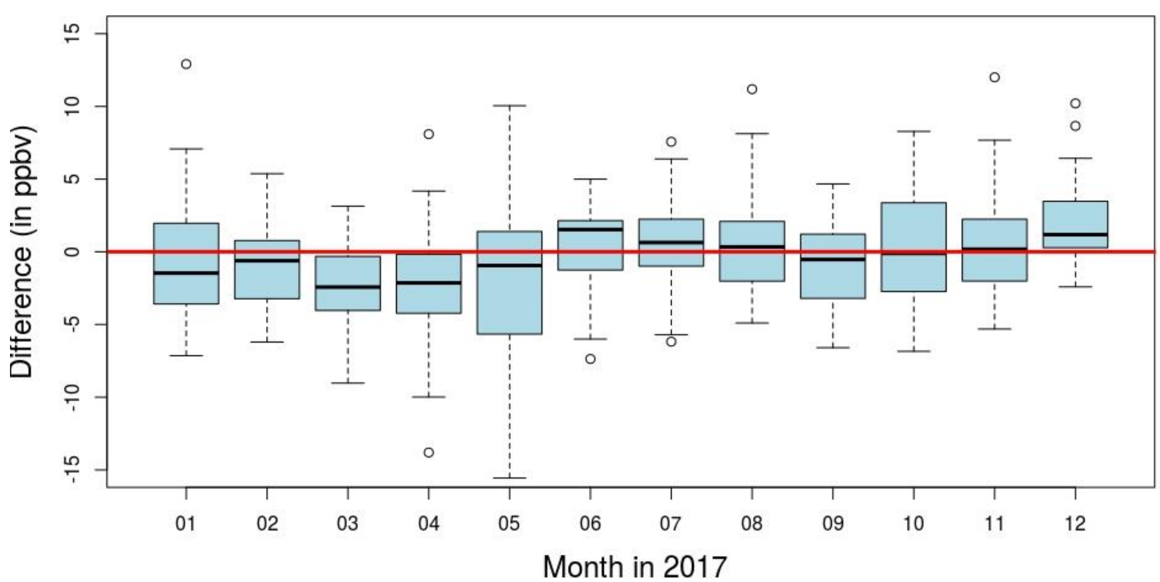


Figure 3.5. *Box and whisker plot for differences between the daily mean of predicted and observed ozone concentrations averages over all stations for 2017. The x-axis represents the month of the year; the y-axis the difference between predicted and observed daily mean ozone concentrations in ppb. Upper and lower ends of the box indicate 25th and 75th percentiles, respectively; the center of the box represents median and whiskers represents the maximum and minimum. Dot circle represents outliers or a single value.*

The warmer months of the year (June, July, and August) showed minimum prediction bias and a median ozone concentration close to zero (Figure 3.5). Figure 3.6 shows the monthly mean bias followed a positive trend to the observed ozone concentration, indicating that from January to May when the ozone concentration was increasing, the model bias (under-prediction) also increased. In June, July, and August, when the observed ozone concentration showed a decreasing trend, the model bias increased (over-predicted). This trend of under- and over- prediction also occurred in subsequent months. In winter, the intensity of sunlight (due to clouds and solar zenith angle) reaching the troposphere was comparatively lower, which leads to low variability in ozone formation. During the warm months, the meteorological conditions are stable, which leads to the efficient formation of ozone (Eslami et al., 2019a, 2019b). Due to this reason, the hourly ozone concentration had a more uniform daily trend in the summer months compared to the winter months. Since the trend was more uniform in summer, the model was more effective at predicting ozone concentrations during summer.

In general, the monthly mean ozone concentration decreased during the summer months (June, July, and August), with the lowest in July (after a yearly high in May). Figure F1 in the appendix shows this monthly trend. This trend occurred in all stations except CAMS-012 (El Paso) and CAMS-013 (Fort Worth). At station CAMS-012, instead of decreasing after May, the ozone level peaked in June and then decreased until December. Since the model was station-specific, it recognized general trends in ozone concentrations and predicted them with an IOA of 0.89 and a correlation of 0.81. (See Tables T2 and T3 in the appendix for monthly IOA and correlation values).

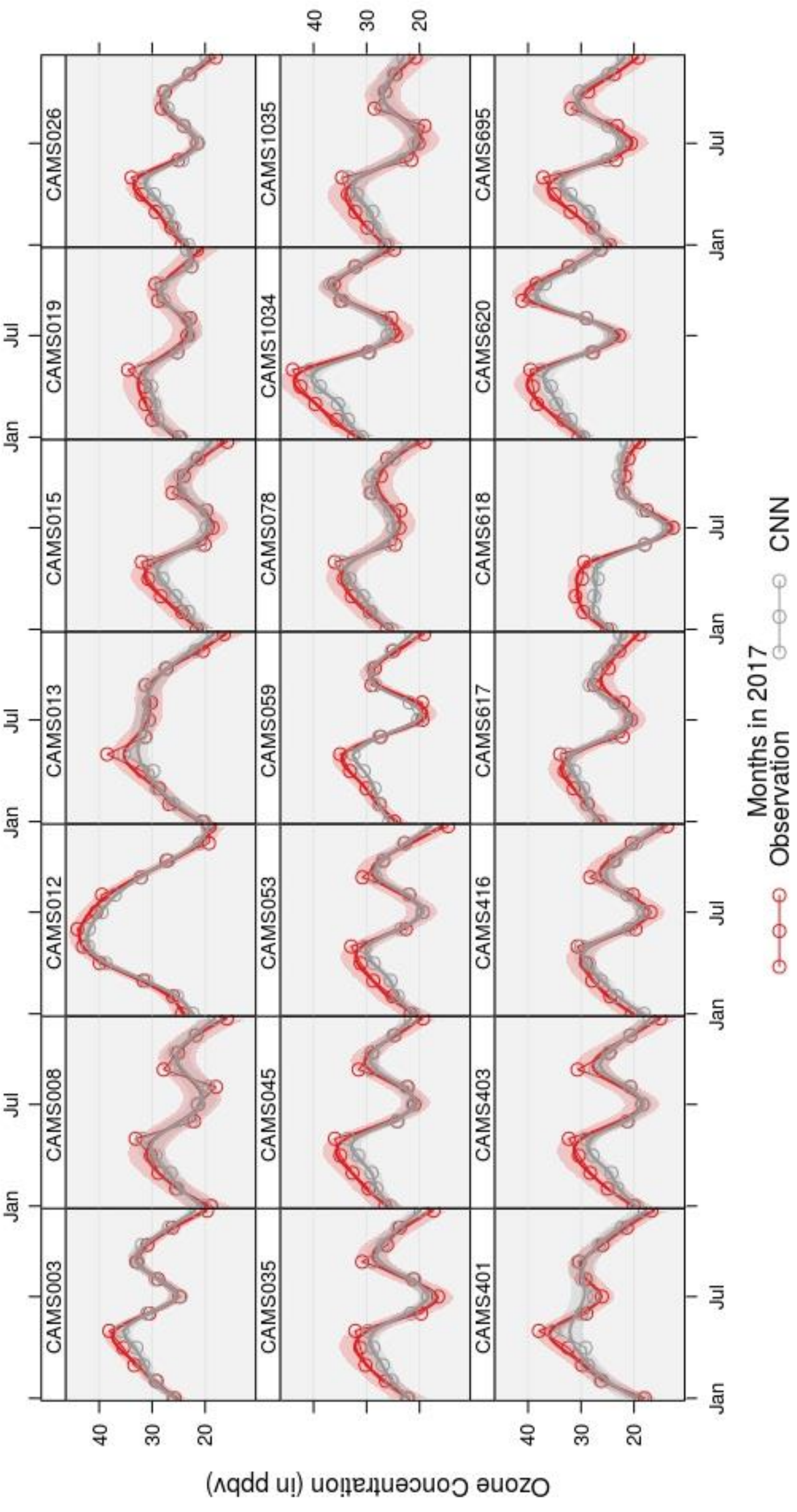


Figure 3.6. *The smooth trend of model and observation: Circles indicate monthly mean with the shaded region showing a 95% confidence interval.*

Table 3.1 lists the annually averaged ozone mean bias and RMSE for 2017 at each station. The maximum mean bias, which was greater than 1 ppbv, occurred at stations CAMS-620 and CAMS-403. The bias stemmed from large variations in ozone concentrations because of the proximity of these stations to petrochemical plants, which increased their sensitivity to NO_x (Pan et al., 2017). As a result, ozone concentrations varied from day to night as well as from weekday to weekday. The other stations showed a bias of 0.5 and -1 ppbv. The RMSE of stations CAMS-012, 403, 008, 015, 301, 695, 053, 045, and 013 was between 9-10 ppbv.

Table 3.1. *Discreet Statistics of AI Model: MB-Mean Bias, NMB- Normalized Mean Bias, RMSE- Root Mean Square Error and NME- Normalized Mean Bias of the average of ozone concentration of all 21 stations*

Station Name	Region	Discreet			
		MB (ppbv)	**NMB (%)**	**RMSE (ppbv)**	**NME (%)**
CAMS-003	Austin-Round Rock	-0.146	-0.49	7.831	20.02
CAMS-008	Houston-Galveston-Brazoria	-0.028	-0.12	9.428	29.17
CAMS-012	El Paso-Juarez	-0.769	-2.38	9.647	23.19
CAMS-013	Dallas-Fort Worth	-0.131	-0.47	9.082	25.05
CAMS-015	Houston-Galveston-Brazoria	-0.355	-1.51	9.411	30.22
CAMS-019	Tyler-Longview-Marshall	-0.554	-2.04	8.468	24.22
CAMS-026	Houston-Galveston-Brazoria	-0.719	-2.76	8.605	24.62
CAMS-035	Houston-Galveston-Brazoria	-0.358	-1.45	8.531	26.10
CAMS-045	Houston-Galveston-Brazoria	-0.945	-3.42	9.152	24.85
CAMS-053	Houston-Galveston-Brazoria	-0.651	-2.62	9.303	28.02
CAMS-059	San Antonio	0.052	0.20	7.996	22.41
CAMS-078	Houston-Galveston-Brazoria	0.159	0.57	8.504	23.48
CAMS-401	Dallas-Fort Worth	0.002	0.01	9.372	26.68
CAMS-403	Houston-Galveston-Brazoria	-1.143	-4.78	9.524	29.64
CAMS-416	Houston-Galveston-Brazoria	-0.385	-1.68	8.684	27.99
CAMS-617	Houston-Galveston-Brazoria	0.523	2.01	8.501	24.57
CAMS-618	Houston-Galveston-Brazoria	-0.430	-1.87	8.111	26.23
CAMS-620	Houston-Galveston-Brazoria	-1.188	-3.58	8.822	19.98
CAMS-695	Houston-Galveston-Brazoria	0.005	0.02	9.350	25.62
CAMS-1034	Houston-Galveston-Brazoria	-0.837	-2.50	8.751	19.22
CAMS-1035	Beaumont-Port Arthur	-0.213	-0.80	7.877	22.38
Average	-	**-0.39**	**-1.14**	**8.81**	**27.94**

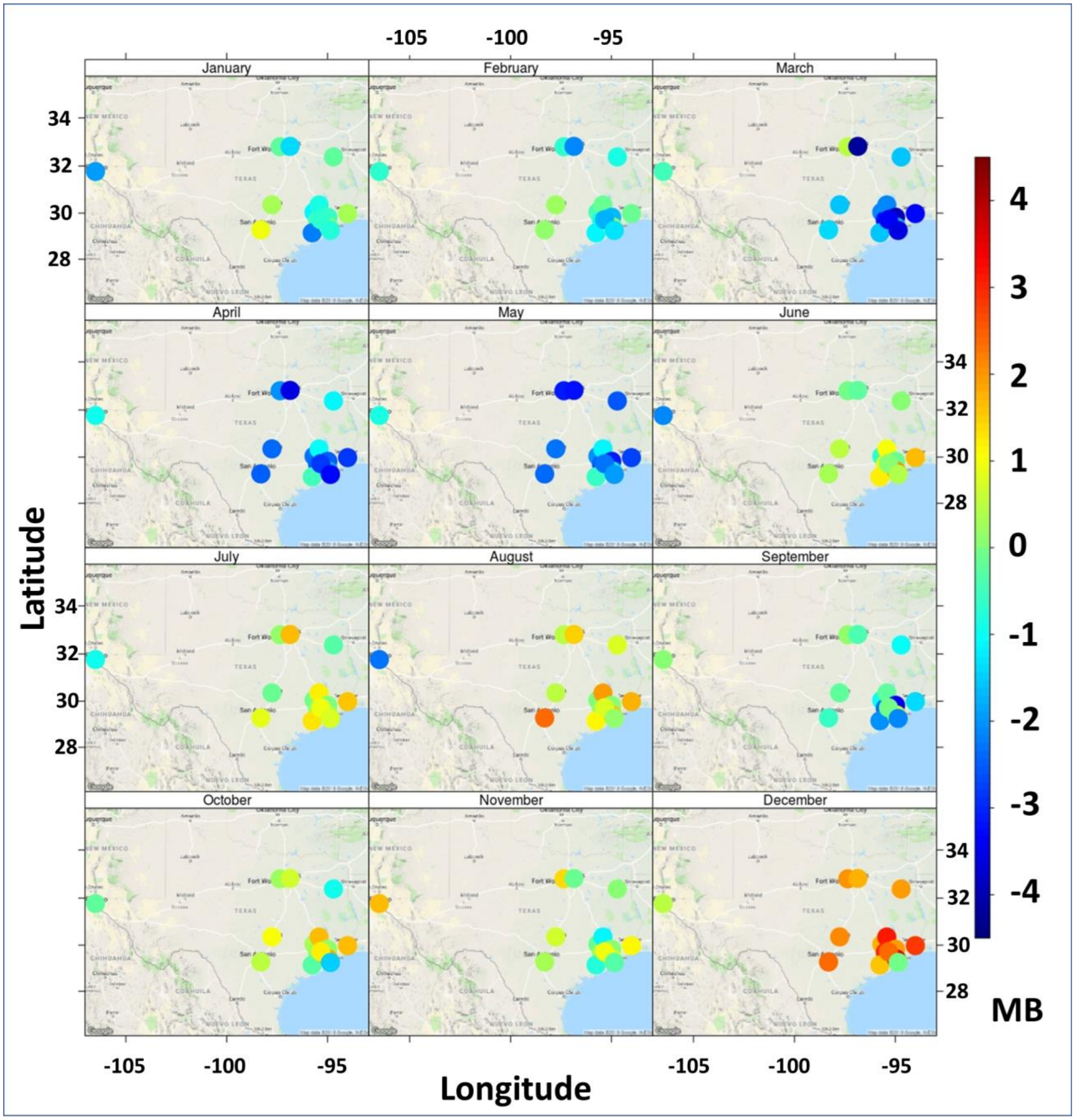


Figure 3.7. *Mean Bias (Monthly MB of each Station): Each dot shows location-wise MB for each month of the Year 2017.*

A detailed analysis of each station suggests under-predictions for January to May and September at most of the stations (Figure 3.7). As ozone concentrations steadily increased from January to May, the model tried to follow it and thus lagged, leading to under-predictions during these months. In June, July, and August, ozone concentrations steadily decreased, and the delay in

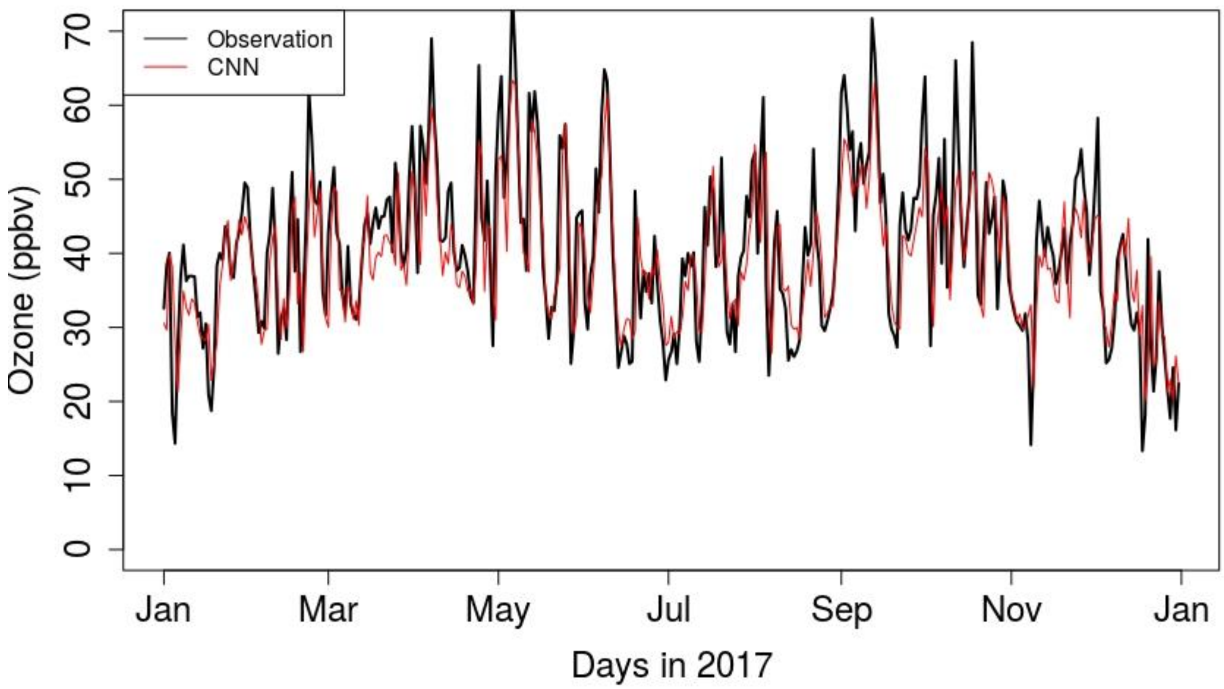


Figure 3.8. *Model-measurement comparisons for daily maximum ozone concentrations for 2017. Black lines represent observations and red model predictions.*

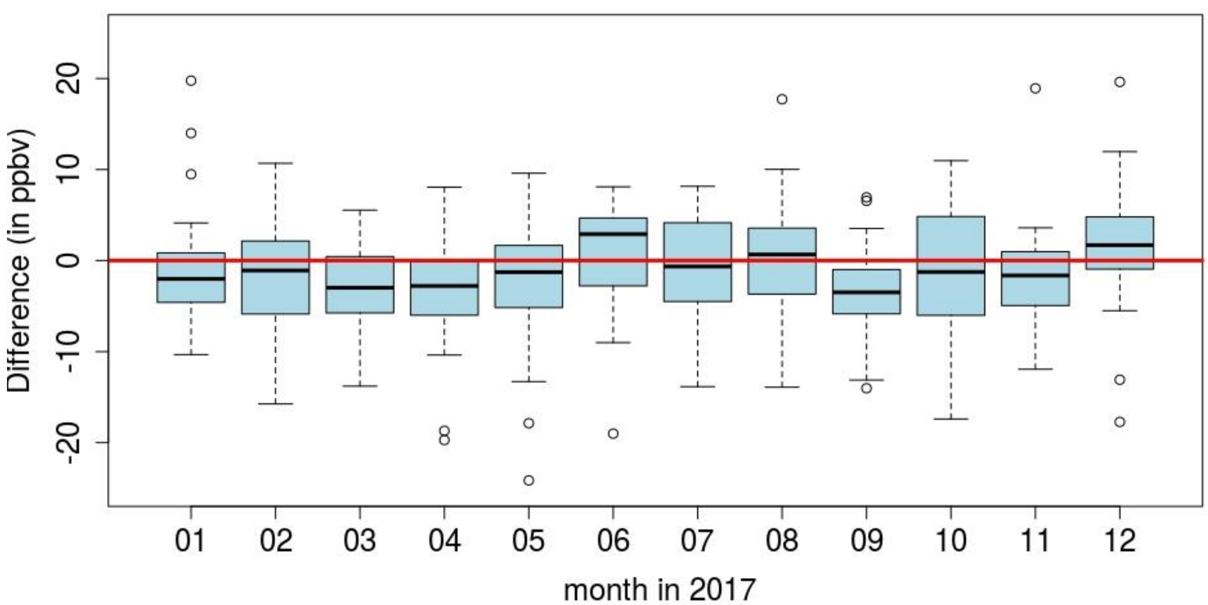


Figure 3.9. *Box and whisker difference plot for a daily maximum of model-predicted and measured ozone concentrations.*

model response to the decreasing trend led to over predictions. In September and August, observed ozone concentrations again increased, and the model again lagged, thus under-predicting. This behavior could be attributed to a lack of examples that could explain these variations for an

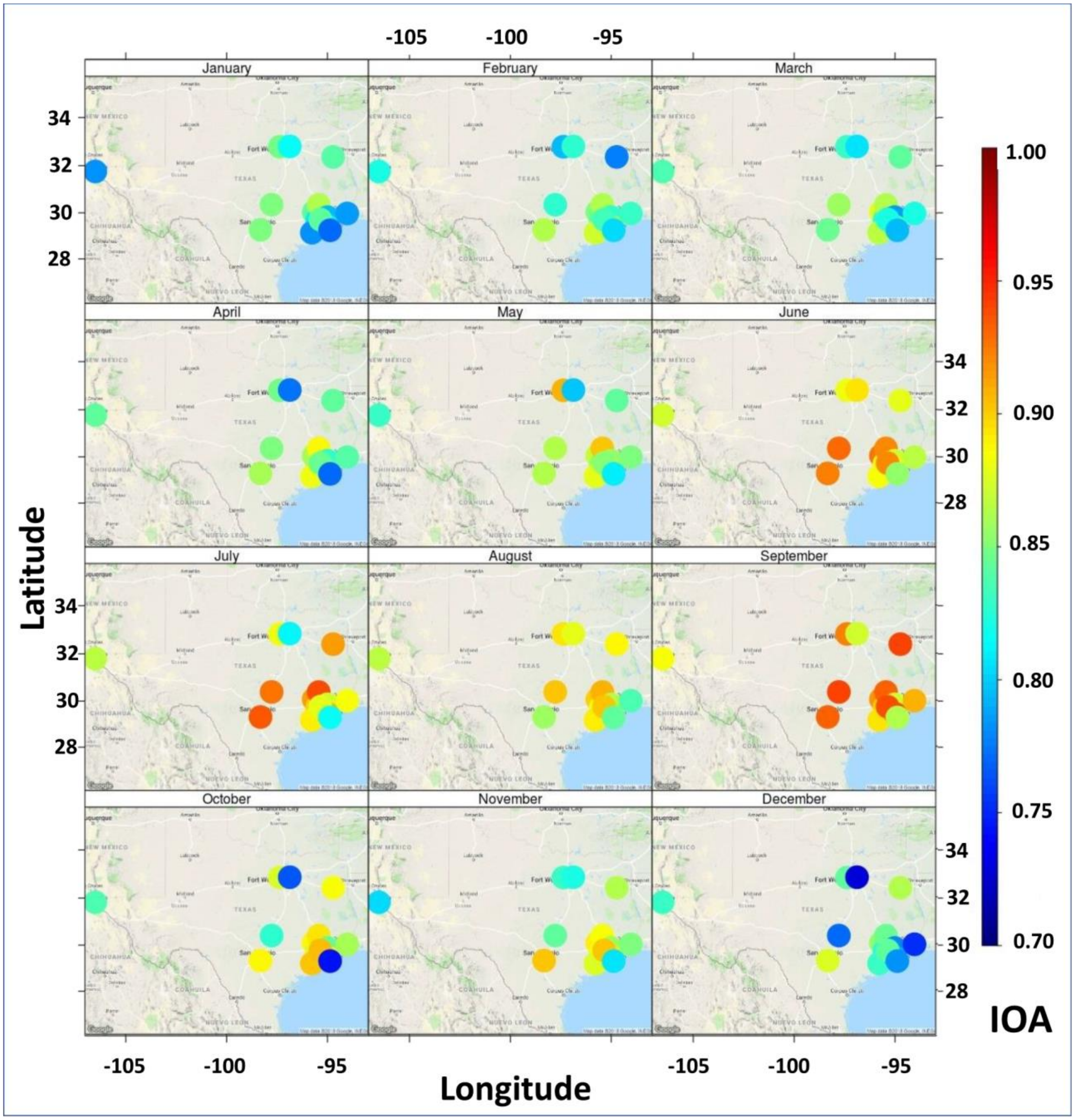


Figure 3.10. *Monthly Index of Agreement for all stations in Texas for the year 2017. Each circle represents the location and corresponding value.*

efficient training process. In addition, the input variables for ozone prediction were inadequate; that is, the measurements for several important ozone predictors such as the cloud fraction and PBL height were not available for the study area, causing the unexplained bias in ozone prediction on some days, which weakened the overall performance of the model. Figure 3.7 shows that the bias in the Houston area varied more than it did in other regions in Texas. In Houston, the formation of ozone, a secondary pollutant, was triggered by reactions between primary pollutants emitted by the oil and gas industry, automobiles, and biogenic sources at various locations in and around the city (Pan et al., 2017). In addition, the influence of the Gulf of Mexico on the meteorological condition of Houston results in large hourly and monthly variations of these input variables of the model. Figures 3.8 and 3.9 plot variations in daily maxima of predictions and observations. The trends were similar to daily mean bias. The maximum range of the bias for daily maximum ozone was ~ ±10 ppbv (barring a few outliers), while the IOA and correlation for a daily maximum ozone concentration (averaged over all stations) were 0.87 and 0.81, respectively.

Model accuracy, represented by the IOA, was shown in Figure 3.10 (refer to Table T2 in the appendix for values). The IOAs of 19 stations were in the range of 0.85-0.90. Station CAMS-078 had the highest IOA of 0.90. Only two stations (CAMS-045 and CAMS-620) had an IOA below 0.85. Both stations are in the Galveston Bay area and positioned downwind of petrochemical refineries in Houston. One explanation for the low IOA at these stations can be attributed to frequent changes in hourly observed ozone concentrations resulting from the combined effect of the land-sea breeze and the presence of petrochemical refineries (NO_x sensitive region) (Pan et al., 2017). The month-wise analysis suggests summer months (June, July, August, and September) had maximum IOAs, and winter (DJF) months had the lowest at almost every station. This trend was similar to the mean monthly trend, suggesting less overprediction and more underprediction of ozone concentrations.

3.3.2. Based on MDA-8 (Categorical Statistics)

Categorical statistics, which are based on a threshold, determine the likelihood of occurrence of an event (Zhang et al., 2012). Therefore, they are more applicable to the evaluation of a forecasting model. To evaluate the model, the following categorical statistics were used: HIT (the Hit Rate), CSI (the Critical Success Index), FAR (False Alarm Rate), ETS (Equitable Threat Score), and POC (the Proportion of Correct) (Chai et al., 2013; Eder et al., 2006). Based on the threshold value, four different cases arise:

a) N_a = Number of times a prediction was above, and observation was below the threshold.

b) N_b = Number of times a prediction and an observation are above the threshold.

c) N_c = Number of times a prediction and an observation are below the threshold.

d) N_d = Number of times a prediction was below, and observation was above the threshold.

From these cases, the following quantities were defined to evaluate the model:

HIT represents the fraction of instances in which the model predicts an extreme event (events above the threshold) from all actual occurrences of extreme events.

$$HIT = \frac{N_b}{N_b+N_d} \tag{3.2}$$

CSI was the fraction of instances in which correct prediction of an extreme event out of all events after removing correctly predicted below the threshold instances.

$$CSI = \frac{N_b}{N_a+N_b+N_d} \tag{3.3}$$

FAR was the fraction of instances in which the wrong prediction of an extreme event of all predictions of an extreme event.

$$FAR = \frac{N_a}{N_a+N_b} \tag{3.4}$$

ETS was similar to CSI but was more accurate for measuring the performance skill of a model. It ranges from -1 (poor model) to 1 (perfect model) (Schaefer, 1990). A positive value of ETS means a skillful model.

$$ETS = \frac{N_b - N_r}{N_a+N_b+N_d-N_r} \tag{3.5}$$

$$\text{where, } N_r = \frac{(N_a+N_b)\times(N_b+N_d)}{N_a+N_b+N_c+N_d} \tag{3.6}$$

POC was the fraction of instances in which accurate prediction an exceedance or a non-exceedance.

$$POC = \frac{N_b+N_c}{N_a+N_b+N_c+N_d} \tag{3.7}$$

To set the threshold, two parameters were selected, i) effects on human health; ii) the fraction of observations above the threshold that make sense of the categorical statistics. Considering the severe health repercussions of ozone, the Environmental Protection Agency (EPA), on October 1, 2015, strengthened the ground-level maximum 8-hour averaged ozone standard from 75 ppbv to 70 ppbv (National Ambient Air Quality Standards for Ozone-2015). US-EPA guideline for ambient air quality states that an ozone concentration of 70 ppbv was unhealthy for individuals who are sensitive to ozone (https://cfpub.epa.gov/airnow/index.cfm?action=pubs.aqiguideozone). For all 21 stations in this study, the number of occurrences above the 70 ppbv thresholds was 95/7,665 (i.e., only 1.23%). Thus, using the 70 ppbv thresholds will not produce meaningful value for the parameters of categorical statistics. The next level, 55 ppbv, was unhealthy for individuals who are exceptionally sensitive to ozone. The number of occurrences above this threshold was 734 (i.e., 9.57%). Therefore, the threshold was set at 55 ppbv for the evaluation of the categorical statistics.

Table 3.2 lists the categorical statistics obtained from a comparison between MDA-8 observations and predictions (Table T4 in the appendix represents categorical statistics from Chai et al. (2013) for comparison). The model has an overall ETS of 0.26, a HIT of 0.34, and a FAR of only 0.32. The accuracy, or POC, of the model was also adequate, with a score of 0.92. Except for CAMS-019, all stations have an ETS greater than 0.10. Since the model mostly

Table 3.2. Daily MDA8 ozone categorical statistics for 2017 with 55 ppbv threshold. See text for details.

	N_a	N_b	N_c	N_d	N_r	HIT	CSI	FAR	POC	ETS
CAMS003	6	18	325	16	2.24	0.53	0.45	0.25	0.94	0.42
CAMS008	10	21	314	20	3.48	0.51	0.41	0.32	0.92	0.37
CAMS012	19	33	268	45	11.11	0.42	0.34	0.37	0.82	0.25
CAMS013	10	19	305	31	3.97	0.38	0.32	0.34	0.89	0.27
CAMS015	2	7	342	14	0.52	0.33	0.30	0.22	0.96	0.29
CAMS019	2	1	338	24	0.21	0.04	0.04	0.67	0.93	0.03
CAMS026	4	11	321	29	1.64	0.28	0.25	0.27	0.91	0.22
CAMS035	1	7	336	21	0.61	0.25	0.24	0.13	0.94	0.22
CAMS045	4	4	334	23	0.59	0.15	0.13	0.50	0.93	0.11
CAMS053	3	14	320	28	1.96	0.33	0.31	0.18	0.92	0.28
CAMS059	4	5	331	25	0.74	0.17	0.15	0.44	0.92	0.13
CAMS078	9	16	316	24	2.74	0.40	0.33	0.36	0.91	0.29
CAMS401	6	10	322	27	1.62	0.27	0.23	0.38	0.91	0.20
CAMS403	2	9	335	19	0.84	0.32	0.30	0.18	0.94	0.28
CAMS416	10	4	332	19	0.88	0.17	0.12	0.71	0.92	0.10
CAMS617	5	8	331	21	1.03	0.28	0.24	0.38	0.93	0.21
CAMS618	4	3	345	13	0.31	0.19	0.15	0.57	0.95	0.14
CAMS620	4	16	329	16	1.75	0.50	0.44	0.20	0.95	0.42
CAMS695	7	19	319	20	2.78	0.49	0.41	0.27	0.93	0.38
CAMS1034	4	19	309	33	3.28	0.37	0.34	0.17	0.90	0.30
CAMS1035	2	5	341	17	0.42	0.23	0.21	0.29	0.95	0.19
Overall	**118**	**249**	**6813**	**485**	**35.14**	**0.34**	**0.29**	**0.32**	**0.92**	**0.26**

under predicts, the HIT ranges from 0.04 to 0.53, with one station (CAMS-019) at 0.04 and the other stations above 0.15. More than half of the stations exhibit a HIT greater than 0.30. The FAR was between 0.13 and 0.71, with only four stations reporting a FAR greater than 0.50 and ten stations reporting a FAR as low as 0.20 or below.

3.3.3. Case Study

3.3.3.1. Stations with fewer meteorological input features

To evaluate the model's effectiveness, two stations (CAMS-003 and CAMS-059) were selected that monitor three meteorological input features and six stations (CAMS-045, 053, 078, 617, 618, and 620) that monitor four meteorological input features. Although these stations monitor few meteorological variables (i.e., wind speed and direction, temperature, and solar radiation), the performance of these stations was on par with other stations. The IOAs for CAMS-003 and 059 were greater than 0.89. CAMS-003 had the highest ETS of 0.42, while CAMS-059 had an ETS of 0.13. While stations CAMS-053, 078, 617, and 618 performed well with IOAs greater than 0.88 and correlations greater than 0.79, stations CAMS-045 and 620 had IOAs of 0.84 (an explanation of their weak performance and solution for these stations are discussed in the next section). An analysis of the importance of input features to predicted ozone concentrations suggested that the results stemmed from the dependence of the model on previous day ozone concentrations. The model assigns higher weights to the previous day ozone than to the previous day meteorology, indicating that the model considers ozone concentration (previous day) the most important input variable. Thus, even with only three or four meteorological input variables, the model was relatively accurate at forecasting ozone concentrations.

3.3.3.2. Station CAMS-045

In this study, the stations with the weakest performance were CAMS-045, CAMS-620, and CAMS-1034. These stations are uniquely located near large bodies of water, which can initiate a cooling effect that results in the lowering of the height of the PBL (planetary boundary layer). As a result, observed ozone concentrations varied considerably hour by hour (Chai et al., 2013). During the training phase of the model, these variations created noise that may have led to lower performance metrics. To mitigate the problem of noise, more input data were used in the model for training purposes. Station CMAS-045 was trained with ten years (2007-2016) of data instead of three years (2014-2016) and observed improvement in the results: The IOA increased by 2.5% and the correlation by 3.6%. Table 3.3 shows a month-by-month comparison of the effect of additional training examples.

Table 3.3. *Comparison of IOA and correlation for station CAMS045 based on 10 and 3-year training*

Months	IOA		Correlation	
	10 Years	3 years	10 Years	3 years
Jan	77.74	74.87	65.46	61.09
Feb	82.90	80.57	73.36	68.84
Mar	83.26	79.28	74.84	72.18
Apr	80.72	77.16	68.13	64.10
May	85.41	83.71	76.35	73.17
Jun	86.77	86.98	77.35	78.16
Jul	85.62	84.30	75.52	73.84
Aug	86.98	84.41	77.60	74.01
Sep	87.02	84.59	78.22	76.51
Oct	83.72	82.44	73.22	71.29
Nov	80.80	79.87	69.20	67.61
Dec	79.14	75.82	65.71	59.73
Overall	86.49	84.38	77.43	74.76

Table 3.4. *Specifications of neural network models compared in this study.*

Model	Hidden/convolutional layer(s) structure†	Number of epochs†	Optimizer‡	Computational Time (in secs)
CNN	5 layer of CONV1D/264*	100	Adam	16.67
RNN (GRU)	128/64/32	400	Adam	1179.23
DNN	128/64/32	100	Adam	4.90
MLP Repressor	-	100	SGD	1.8-4.3
Ridge Regression	-	-		<1
Lasso Regression	-	-		<1

† Optimized using trial and error tests.
‡ Both Adam and stochastic gradient descent (SGD) was explained in Kingma et al. (2014).
* Convolutional layers with filter size 32 and kernel size 2, following with a fully connected hidden layer with size 264.

3.3.4. Model Comparison

To test the robustness of the model used for this study (i.e., CNN model), various machine learning models were compared that are commonly used in predicting time-series, particularly those with high nonlinearity (e.g., Eslami et al., 2019a, 2019b). They include multilayer perceptron (MLP) structure (Glorot and Bengio, n.d.; Hinton, 1989), deep neural networks (DNN) formed by having densely-connected layers, recurrent neural networks (RNN) with Gated Recurrent Unit (GRU) (Cho et al., 2014), Lasso Regression (Tibshirani, 1996), and Ridge Regression (Tikhonov, 1998). These models, including CNN, were trained on the same dataset as discussed above. Table 3.4 shows the model configuration and performance comparisons based on the computational time required to train each model. Once the model was trained, it was used to predict the whole year of 2017.

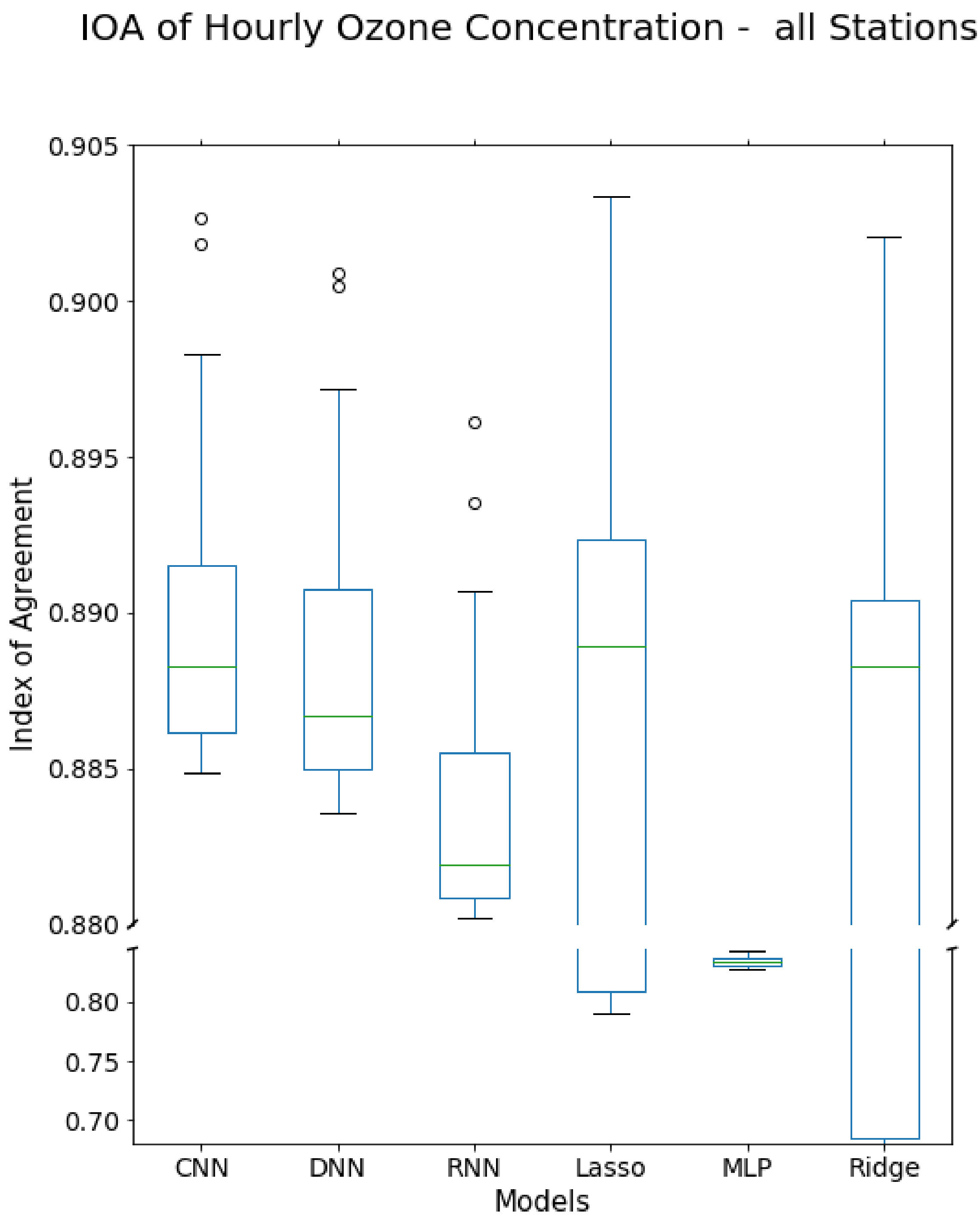


Figure 3.11. *Box and whisker plot of the index of agreement (IOA) based on hourly ozone concentration for the year 2017 of all the stations. Y-axis represents IOA, and X-axis are the models in the study.*

For comparing the performance of different models, the IOA of the 24-hour time series (Figure 3.11) and IOA of the daily maximum ozone concentration (Figure 3.12) for all stations were compared. The CNN model performed better than all other models, while DNN was the close second. CNN model had the highest mean IOA (0.89) for the 24-hourly time series. When comparing the daily maximum ozone concentration (Figure 3.12 and Figure 3.13), the CNN model (IOA = 0.78) performed better than the other models. Furthermore, Figures 3.12 and 3.13 show all deep learning models (CNN, DNN, and RNN) performed notably better than other machine learning model (MLP) and regression model (Lasso and Ridge). It means that these models understood the topology within the daily ozone time series (the relationship between different hours during a day), even though they were trained with the same amount of training samples. This can also be seen in the categorical analysis. CNN model has the highest mean POC (0.92), CSI (0.273), and ETS (0.244) among all the models (Figure F4 in the appendix). CNN was also better than other models in predicting an extreme event with the highest mean HIT rate (0.313). To summarize the abovementioned comparisons, the CNN model performed statistically better than the other machine learning and regression models.

Even though Ridge and Lasso regression shows a better computational efficiency than the other models, they were unable to predict the outliers (high ozone peaks) because of the L2 (Ridge regression) and L1 (Lasso regression) regularizations. The regularization process has superb computational efficiency; however, it negatively affected the accuracy performance of real-time prediction compared with deep learning models. Also, these regression techniques have high variance in IOA for hourly time-series as well as in the IOA of daily maximum concentration across all stations. Although the MLP model was about three times faster than the CNN model used for this study, the accuracy (in terms of IOA) was 2-8% less than the CNN model for all stations. Both DNN and RNN models show comparable accuracy but have performance lower than that of CNN.

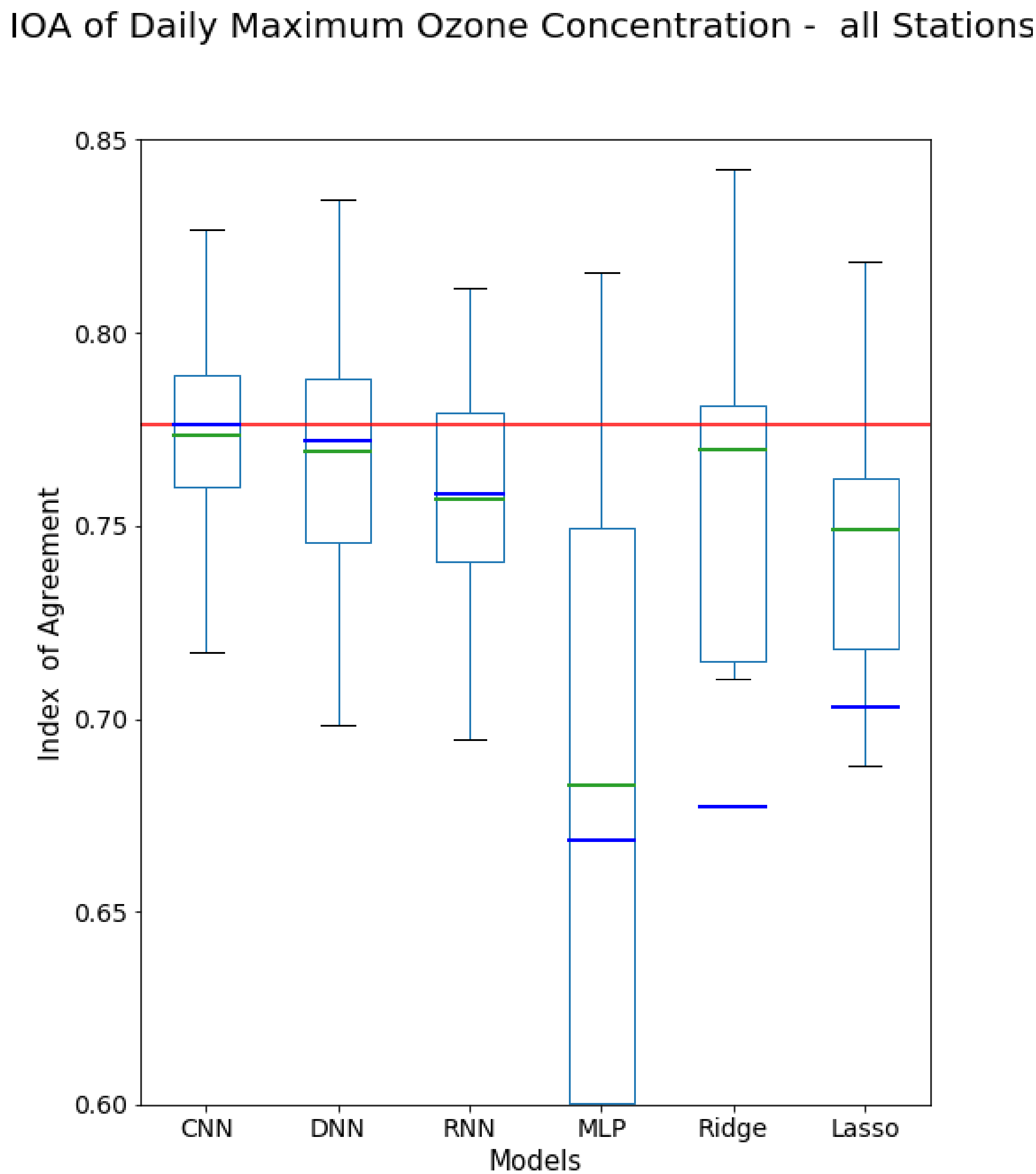


Figure 3.12. *Box and whisker plot for the index of agreement based on daily maximum ozone concentration of all stations. Y-axis represents IOA, and X-axis are the models in the study.*

Additionally, RNN was ~70 times slower in training as compared to CNN for this study. It means that, even though the RNN model can reach a comparable level of prediction accuracy as the other two deep learning models, it needs notably higher computational time. From the above discussion, it can be concluded that the CNN model was better than the other models tested for this study across various evaluation parameters for the real-time prediction of hourly ozone concentrations.

One commonly used practice in improving machine learning models, especially in predictive regression modeling problems, was feature importance analysis. It was the automatic selection of features in the data that are most relevant to the predictive modeling problem. In feature importance analysis, the method selects the most important features present in the data without changing their values; and uses these selections in the training process of the predictive model (e.g., a CNN model). In addition, using irrelevant features can negatively impact the performance of the machine learning model in predicting hourly ozone time-series by making the model learn based on unimportant features. In this way, the robustness of the CNN model by presenting inputs with or without the knowledge from feature importance analysis can be tested. Here, the feature selection was performed using a random forest (RF) model (Breiman, 2001). RF was chosen due to its generally good predictive performance, low overfitting, and easy interpretability compared with other machine learning models. First, an RF model was trained to determine the importance of input features in predicting hourly ozone concentrations. Once a list of the features with their importance was obtained, CNN models were trained with the best 24-set (24, 48, 72, etc.) of features. Each model was then compared against the standard CNN model used in this study (Table T5 in the appendix). Results indicated that the averaged improvement in IOA was around 0.5% higher after applying the RF feature selection compared to the standard CNN model. This shows that the CNN model was able to extract enough information needed to make a proper prediction and understanding the importance of each input feature without applying a prior feature selection method. One of the major caveats in using RF for feature selection was that every station has a different number of features on which their performance improved. It would become cumbersome for the user to run multiple models for an optimum result. Thus, to have a generalized model for all stations, the standard CNN was used.

3.4. Conclusions

In this paper, a real-time 24-hour ozone prediction model based on a deep convolutional neural network was developed, discussed, and evaluated. After designing the model architecture, it was trained on examples from 2014 to 2016. The hourly concentrations of ozone for each day of 2017 were predicted by training the model with examples until the last day and evaluated the prediction data using discreet and categorical statistics.

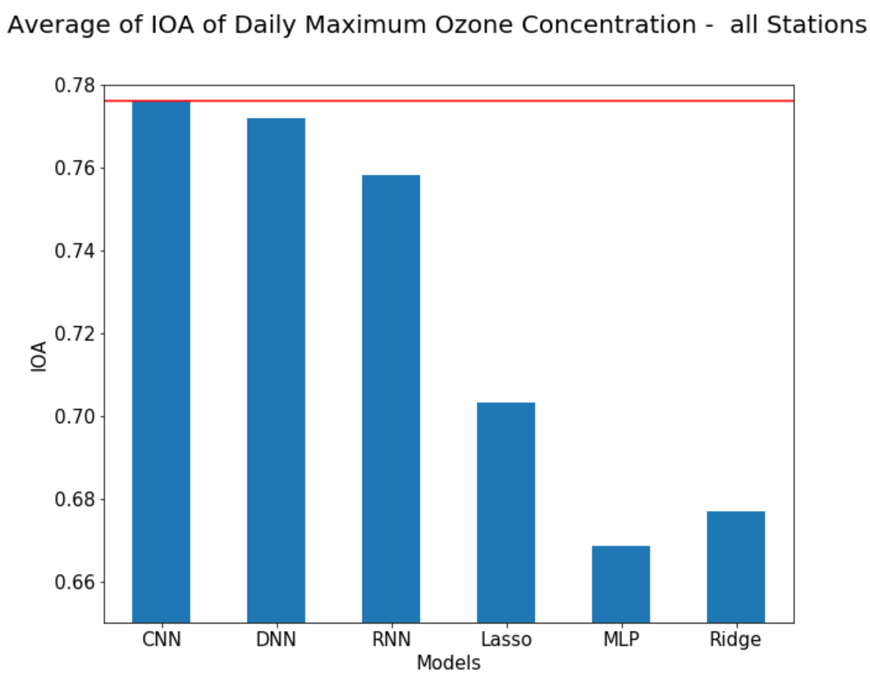


Figure 3.13. *Bar plot average of IOA for all stations. IOA was based on hourly ozone concentration for the year 2017. Y-axis represents IOA, and X-axis are the models in the study.*

In general, the observed ozone concentrations at all stations increased from winter to spring and then decreased during the summer months. During the fall, concentrations again steadily rose until September and then declined to their lowest levels in December. Although the model was able to capture these patterns, the model's response to these changes was slow. For example, when the observed ozone concentration increased, the model generally underpredicted (for both daily and monthly variations), and when observed ozone concentrations declined, it was generally overpredicted. The change in the trends of prediction was largely the result of the delay of the model in responding to changes in the concentrations of observed ozone.

Because of the location of CAMS-012 (El-Paso) 1140m above sea level, it exhibited a meteorological condition that differed markedly from those of other stations in this study. As a result, this station did not follow the general trend mentioned above. Here, the monthly mean ozone concentration steadily increased until May-June and then decreased until December before increasing again. The probable cause for this increase could be the presence of comparatively high temperatures and sunlight during the daytime, leading to the high production of ozone. Another explanation could be the accumulation of ozone from the reversal of the wind direction from westerly to easterly (Figure F2 in the appendix) during the summer and the presence of the Rocky Mountains in the West. Even though this station did not exhibit the general trend, the model captured the trend in the observed ozone concentration with an IOA of 0.89 and a correlation of 0.81. The model's performance at this station suggests that it was able to comprehend the station-specific trends and produce satisfactory results.

At a few stations (e.g., CAMS-045), the model did not perform satisfactorily because of more frequent hourly variations in observed ozone concentrations, either due to NOx emissions or the lowering of the PBL triggered by the cooling effect from the sea breeze. However, with the addition of more years of training data, the performance metrics of these stations improved. The model was trained with seven more years of examples for station CAMS-045, and its IOA improved from 0.84 to 0.86, and its correlation improved from 0.74 to 0.77. In addition, even though the model was capable of predicting daily trends with reasonable accuracy, it mostly under-predicts daily maxima (also evident in the scatter plot Figure F3 Appendix). This drawback can be overcome with the addition of more days of training data as well as more inputs (i.e., more meteorological inputs like PBL, cloud fraction or air pollutants), which can be obtained from a numerical model such as WRF or CMAQ.

The deep CNN model developed forecasted hourly ozone concentrations 24 hours in advance. It showed significant improvement over both numerical and statistical models. Furthermore, when compared with other neural networks (MLP, RNN, and DNN) and regression models (Lasso and Ridge), the CNN model was better in predicting both daily time-series and daily maximum of ozone concentrations. Additionally, it was able to generate the same level of accuracy compared with a model with feature importance analysis.

In addition, the model successfully predicted ozone concentrations with IOAs greater than 0.85 for 19 of the 21 stations. The performance metrics were considerably improved for the stations by adding more years of training examples, as demonstrated in the second case study. The model also proved its effectiveness on stations with fewer meteorological inputs (i.e., CAMS-003, CAMS-059). In addition, it took only a few minutes to predict a single day so that it could be employed at any station. The benefit of this model was that it could be a useful, efficient tool for forecasting air quality and issuing health advisories in advance, thus reducing the serious effects of ozone on human health.

CHAPTER 4

A NOVEL CMAQ-CNN HYBRID MODEL TO FORECAST HOURLY SURFACE-OZONE CONCENTRATIONS FOURTEEN DAYS IN ADVANCE[2]

4.1. Introduction

Surface ozone can pose a significant health risk to both humans and animals alike, affecting crop yields (USEPA - 2006). According to the US Clean Air Act, it was one of the six most common air pollutants, and considering its impact on health; the Environmental Protection Agency (EPA) of the United States has limited the maximum daily eight-hour average (MDA8) concentration of ozone to 70 ppb. Similarly, the Ministry of Environment in South Korea has declared a standard for hourly ozone of 100 ppb and 60 ppb for MDA8. To achieve these attainment goals and to understand future projections (forecasts), researchers have turned to various numerical modeling and statistical analysis tools. One such numerical model was the Community Multi-scale Air Quality Model (CMAQ), a chemical transport model (CTM) developed by the USEPA (Byun and Schere, 2006). Widely used to forecast the air quality of a region with considerable accuracy, CMAQ was an open-source multi-dimensional model that provides estimated concentrations of air pollutants (e.g., ozone, particulates, NO_x) at fine temporal and spatial resolutions. It has been used as a primary dynamical model in regional air pollution studies; CMAQ modeling, however, has several limitations (e.g., parameterization of physics and chemistry) and raises uncertainties that lead to significant biases (overestimations or

[2] Published: Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A.K., Lee, J.-B., Park, H.-J., Choi, M.-H., 2021a. A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. Sci Rep 11, 1–8.

underestimations) overestimations of ozone concentrations (Chatani et al., 2011; Kitayama et al., 2019; Liu et al., 2010; Morino et al., 2010; Trieu et al., 2017). Also, the stochastic nature of the atmosphere results in inherent uncertainty in even comprehensive models that might limit their accuracy (Rao et al., 2020).

CTMs require substantial computational time since they entail multiple physical and chemical processes for each grid. The testing time for various scenarios of the CMAQv5.2beta configuration test (with US EPA Calnex 12km domain July 2, 2011 testing dataset) was in the range of 34-54 minutes (Further detail about the test can be found at CMAQ version 5.2beta (February 2017 release) Technical Documentation - CMASWIKI (airqualitymodeling.org)). ("CMAQ version 5.2beta (February 2017 release) Technical Documentation - CMASWIKI,") Unlike CTMs, machine learning (ML) can be trained to forecast multi-hour output using a certain set of inputs more accurately within a faster processing time (Lops et al., 2019; Sayeed et al., 2020a). In addition, it requires only one training process, further reducing the computational time. Although all ML models are more accurate with faster processing speeds, they are very localized (station-specific) and generate large underpredictions of daily maximum ozone concentrations (Eslami et al., 2019a, 2019b; Sayeed et al., 2020a).

The objective of using this ML technique was to enhance the CMAQ modeling results by taking advantage of i) the deep neural network (DNN), a computationally efficient, artificially intelligent system that recognizes uncertainties resulting from simplified physics and chemistry (e.g., parameterizations) of the CMAQ model; and ii) CMAQ, which computes unmeasured chemical variables along with fine temporal and spatial resolutions. This approach aimed to use the best of both numerical modeling and ML to design a robust and stable algorithm that more accurately forecasts hourly ozone concentrations 14 days in advance and covers a larger spatial domain. The ML technique used in this study was based on the convolutional neural network (CNN) model.

4.2. Material and Methods

The proposed algorithm uses two sets of inputs: i) parameters predicted by numerical models and ii) the previous day's observed air quality (Figure F7 in the appendix displays a schematic of the deep CNN architecture for predicting hourly ozone concentrations for the next fourteen days.)

Like any neural network, a deep CNN was an optimization problem that attempts to minimize the loss function. The most generally used loss functions are the mean squared error, the mean absolute error, and the mean bias error. In this study, two loss functions were tested: i) the mean square error (method 1) and ii) a customized loss function (method 2) based on the index of agreement (IOA) (Willmott, 1981; Willmott et al., 1985). The mathematical expression of IOA appears in the appendix (Appendix D: General Statistics). In method 1, the model attempts to find a solution iteratively such that the mean square error was a minimum. Similarly, in method 2, the model attempts to fit it in such a way that the IOA was maximum. In both cases, two separate models were obtained for each day of prediction. The reason for choosing the IOA as a loss function was that high peaked concentrations in air quality forecasting prediction are critical. IOA, unlike the mean bias or the mean square error, was a better parameter that more accurately reports the quality of a model. The CNN, like any ML technique, was an optimization problem; the model tries to simulate as close to true observations as possible and relies on minimization or maximization of certain performance parameters. In general, ML algorithms, "mean squared errors" (method 1) are used as the cost (loss) function and the model tries to minimize this loss. The issue with this method in the hourly forecast was that it generalized the model and was unable to predict the high peak values because of the sampling bias (only 3-4 high values in 24-hour as compared to 20-21 low or average values). To mitigate this issue, the Index of agreement (IOA) was used as the cost (loss) function and found out that the model was able to predict better high peaks compared to method 1.

4.2.1. Data Preparation and Model Training

The observed air quality was obtained from the Air Quality Monitoring Stations network, operated by the National Institute of Environmental Research (NIER) for 255 urban stations for the years 2014 to 2017 across the Republic of Korea. The network measures and provides real-time air pollutant concentrations such as sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and nitrogen dioxide (NO_2). Since the CNN model requires continuously measured data for training/testing, the missing values of observational datasets were imputed. For these missing values, SOFT-IMPUTE by Mazumder et al. (2010) was used. Then the concentrations of air pollutants were extracted from CMAQ and meteorological parameters from the WRF (processed by Meteorology-Chemistry Interface Processor (MCIP) modules of the CMAQ model). For this purpose, the temporally and spatially matched CMAQ grid points of the NIER station locations were used. Table T6 (appendix) displays all of the parameters extracted from the MCIP and CMAQ.

After acquiring hourly meteorological fields from the WRF model, previous day pollutant concentrations from observations and the forecasted parameters from the CMAQ runs, the inputs were prepared for each station in the form of a two-dimensional matrix in which each column represented a specific parameter (meteorology or gaseous concentration), and each row represented hourly values. Figure F8 (Appendix) represents the schematic diagram of the data preparation used for the 14-day forecasting. For each day (24 hours), separate models were prepared in such a way that inputs remained the same for all the models, but the output (target) was changed from day 1, day 2 until day 14. Thus, a total of 14 models were prepared, one for each forecasting day. Then the model was trained for three years (i.e., 2014 to 2016) and evaluated for the year 2017 (Note: 2017 was never used for training the model; the training of the CNN models was done using data from 2014 to 2016; 1096 days). The input dataset consisted of previous 24-hour observed air pollutant concentrations and the meteorology generated by the WRF; and the following 24-hour forecasted air-pollutants from the CMAQ model (in total 50 input parameters). The output dataset consisted of the next day's 24-hour observed ozone concentration for day 1, 24 to 48 hours for day 2, 48 to 72 hours for day 3, and so on. After the inputs and outputs were defined, the datasets were combined from all stations to construct a matrix for training/testing a generalized deep CNN model across the spatial domain. Since there were 255 stations and three years of hourly data for training, the model was trained with 279,480 (1096×255) examples (days), which were further split randomly into equal parts so that the model was trained on one half and validated on the other. Since each parameter had a unique range of values, each was normalized between zero and 1 to remove the model bias toward any specific high or low valued parameter. It has been observed that having a different maximum and minimum for a training and prediction set destabilizes the model and produces varied results over different runs. Therefore, for the normalization process, a "global" maximum and minimum values were chosen for each parameter. These global maximum and minimum values guaranteed that none of the hourly values exceeded a certain level; thus, the normalization process remained independent of the temporal and spatial variations. After normalization, the deep CNN architecture (defined in the previous section) was used to train the model and generated two models, each with a unique loss function. Once the model was generated, it was used to predict the entire year of 2017.

For long-term training and prediction, the datasets were prepared so that they had the same inputs, but the outputs were changed from the first day to the second, third, and fourth days and so on until the fourteenth day (Figure F8 in the appendix presents a schematic diagram of the data setup used in this study.) Hence, with two loss functions and 14 days of predictions, there were 28 models to evaluate.

4.3. Results and Discussion

The models were trained based on two loss functions (methods 1 and 2) and fourteen days (28 different models), from January 1, 2014, 0000UTC to December 31, 2016, 2300UTC. After training the models, they were evaluated based on various performance parameters. The index of agreement (IOA) based on hourly values of the year 2017 was calculated for each station and then averaged. (The IOA was selected over correlation as the performance metric for reporting the results because i) a correlation of 1 doesn't mean that model captures the high and lows; ii) an IOA considers the bias within the performance metric. Thus, an IOA of 1 will mean that all highs and lows of a time series were captured well. Furthermore, the numerator of IOA addresses the mean bias (Appendix: General Statistical Analysis)). The models based on both methods of the CNN model reported the highest IOA for predicting one day ahead, but the IOA decreased on subsequent days. The average IOAs (method 1 – 0.90, method 2 – 0.91) and correlations (method 1 – 0.82, method 2 – 0.83) for one-day ahead prediction were comparable. The performance of both methods showed improvement over that of the CMAQ model (IOA-0.77, correlation-0.63). The IOA with method 1 increased by 16.86% and that with method 2 by 17.98%. The correlation with method 1 increased by 30% and that with method 2 by 32%.

4.3.1. Performance Comparisons of CMAQ and CNN models

Figure 4.1 shows the yearly IOA (average of all stations). The IOA decreased sharply from day 1 to day 3 but stabilized after the three-day forecasts from both methods. The IOA for day 4 was lowest during the first week of prediction for method 1. After day 4, the IOA increased until day 6 and then decreased until day 10. It increased slightly on day 11 but then decreased further. For method 2, the IOA decreased until day 5, increased until day 7, and then further decreased after day 8. One possible explanation for the weekly trend relates to the weekly cycle of ozone

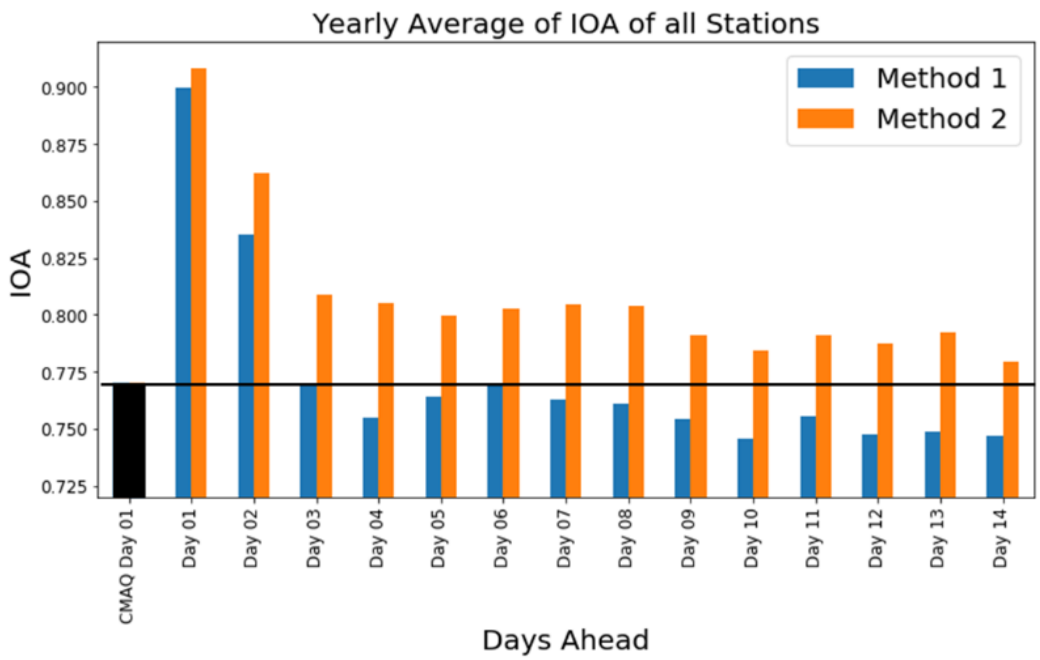


Figure 4.1. *Comparison of Index of Agreement for two-advance prediction using Method 1 and 2. x-axis in the plot shows the days ahead, and the y-axis represents the index of agreement. The blue line represents the IOA of each day's advance prediction using Method 1 (mean squared error as loss function). The orange line represents the IOA of each day's advance prediction using Method 2 (Index of Agreement as loss function).*

concentrations (Choi et al., 2012). Figure F9 (Appendix) shows the autocorrelation (average of all stations) of the current day observed ozone with the subsequent day observations. (The autocorrelation was the correlation of current hourly values with subsequent hours (0 to 336 hours)). The observed ozone followed a weekly cycle, exhibiting a decreasing trend in its correlation until day 3 and then an increasing trend until it peaked on day 7. The same cycle occurred during the second week. Figure 4.1 and figure F9 show that the CNN model also follows this weekly trend. Also, figure 4.1 depicts the superior performance of the CNN model method 2 to that of method 1. The average increase in the IOA of method 2 compared to that of method 1 was 4.77%; a maximum increase of 6.64% occurred on day 4, and a minimum increase of less than 1% occurred on day 1. The greatest increase in the IOA happened on the worst-performing days (days 4, 13, 8, 7, and 12 show an increase of 6.6, 5.8, 5.6, 5.4, and 5.3%, respectively) by method 1.

Figure F10 to F15 (Appendix) shows the hourly time series plots for the month of February and June of stations 131591, 238133, and 823691. (The stations shown here have the highest, median, and least IOA for Day 1 forecasts by the CNN-method 2 model). These figures show that both the CNN models are good up to 7 days forecast. For seventh- and fourteenth-day forecasts, the CNN-method 2 model performed better than the method 1 model. Even though the performance decreases, the CNN models can produce a reliable forecast for up to fourteen days.

In terms of mean bias (Table T7 - Appendix), both the CNN models were under predicting for all days while the CMAQ model for the day 1 forecast was overpredicting. The average of all stations' mean bias for the CMAQ forecast was 1.21. Whereas for CNN-method 1 and CNN-method 2 models, the mean bias was -1.23 and -0.96, respectively. From day 2 onwards, the model, with IOA as loss function, performed significantly better than the model with MSE as loss function. The root means square error (RMSE) for the CMAQ model, the CNN model method 1, and the CNN model method 2 were 18.98, 11.00, and 11.01 for day 1 forecasts, respectively (Table T8 - Appendix). While for day 1, both the methods of the CNN model were equivalent for RMSE, in subsequent days, CNN-method 2 performed slightly better than CNN-method 1. In terms of correlation (Table T9 – Appendix), method 2 of the CNN model outperformed method 1 as well as the CMAQ model. The correlation for all 14 days forecast by the CNN-method 2 model was greater than or equal to the CMAQ's forecast for the first day.

The models were also evaluated based on categorical statistics (Chai et al., 2013; Eder et al., 2006) as defined in Chai et al. (2013). For categorical statistics, the National Ambient Air Quality Standards (NAAQS) for daily maximum 8-hour average (MDA-8) was used. The standard threshold of 70 ppb was too high to properly evaluate model performance with categorical statistics. Thus, the standard (threshold) was further reduced to 55 ppbv as done in Sayeed et al., 2020a). Tables S10 to S14 in the appendix show the hit rate (HIT), false alarm rate (FAR), critical success index (CSI), equitable threat score (ETS), and proportion of correct (POC), respectively. The hit rate for the CNN-method 2 increased to 0.80 from 0.77 for the CMAQ model, while it decreased to 0.67 for the CNN-method 1. The hit rate for day 2 to day 14 forecast ranges between 0.47-0.74. The FAR for the CNN-method 1 was better than the CNN-method 2 for all forecast days. The decrease in FAR for the method 1 and 2 were ~46% and ~35% respectively when compared with the CMAQ model for the first-day forecast. POC, ETS, and CSI were used to evaluate model skills. While the CNN- method 2 performed equivalent to method 1 for POC for all 14 days of forecasts, it had better skill in terms of ETS and CSI when compared with method 1.

From the discussion above, it can be concluded that by changing the cost/loss function, better optimization can be achieved. Introducing IOA as the cost function (see equation 1 in Appendix) not only reduced the bias but also increased the correlation and IOA. This improvement was attributed to the reduction of bias in both high and low values rather than the average/overall bias. Compared to correlation, IOA was a better metric to evaluate a time series since correlation only considers the shape of the time series and not the bias.

4.3.2. Performance Evaluation of Selected Method

It was evident from the above discussion that the performance of the CNN-method 2 overshadowed that of the CNN-method 1; therefore, the performance was further analyzed for method 2 below. Figure 4.2 lists the average yearly IOA of each district in South Korea (Figures were created using R ggplot2 package: https://ggplot2.tidyverse.org/). If a district had more than one station, IOAs were averaged. The inland cities performed slightly better than the coastal ones, and their performance improved the farther they were from the coast (Figures 4.2 & 4.3a and Figure F16 in the appendix). For example, Seoul performed slightly better than Incheon, the former being farther away from the coast. One explanation for the better performance in the central region was that it has ozone chemistry exhibiting a typical diurnal ozone cycle throughout the year than the coastal region, where predominant land-sea breezes may have an impact on ozone chemistry (Figure F17 in the appendix shows 24-hour observed ozone concentrations throughout the year. Figures F17-a, b, and c display the three worst-performing stations, while Figures F17-d, e, and f display the three best (Kotsakis et al., 2019; Pan et al., 2017)). It was evident from the figures that stations with the typical diurnal ozone cycle (Strode et al., 2019) provided more accurate forecasts than those with less variability in hourly concentrations. Ideally, the ozone concentration starts to increase in the afternoon and peaks a few hours before sunset (Eslami et al., 2019a). This forms a distinct diurnal cycle of ozone concentration (addressed like a typical diurnal ozone cycle in this study). The CNN model also follows this typical ozone chemistry and attempts to make predictions based on this information; hence, the station with generalized (typical) ozone chemistry produced more accurate forecasts than the station with less variability in its concentration of ozone throughout the day. (Since the sample size of the typical ozone diurnal cycle was much greater than the diurnal cycle with less variability, the CNN model was biased toward the former.)

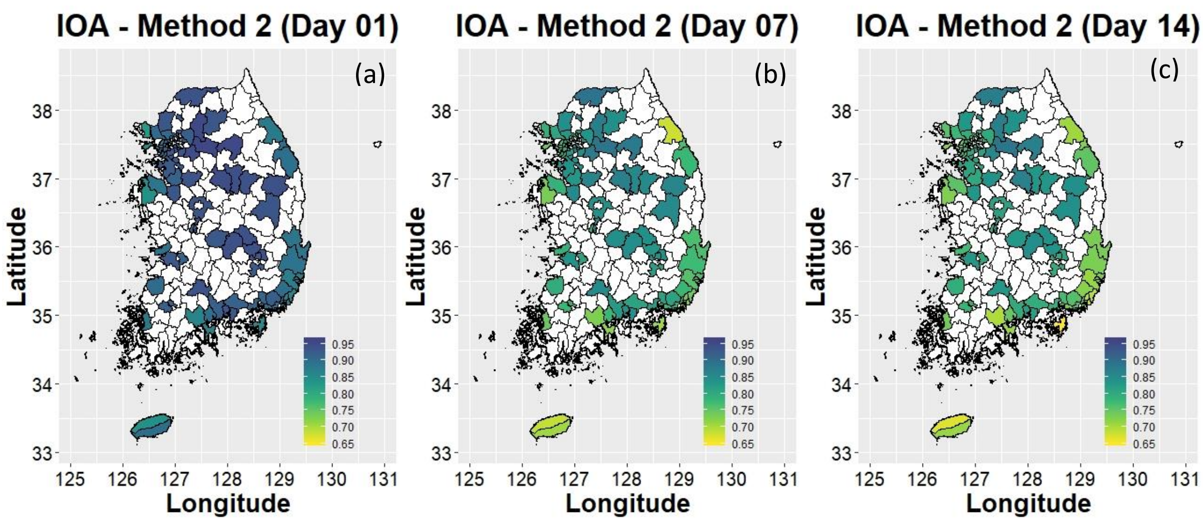


Figure 4.2. *Average IOA all stations (CNN-method 2) in each district of South Korea. a) IOA for Day 1; b) IOA for Day 7; and c) IOA for Day 14. (Figures are created using R ggplot2* ("Create Elegant Data Visualisations Using the Grammar of Graphics,")*: https://ggplot2.tidyverse.org/)*

The accuracy of forecasting was also dependent on the level of urbanization (Figures 4.2 and F18 in the appendix). Out of seven cities with an IOA higher than 0.94, six were among the least urbanized (the 4th and 5th quantiles: urbanization quantiles based on Chan et al. (2015), and only one was an urban region (the 2nd quantile). Ozone precursors are mostly anthropogenic in urban areas that can be highly variable (Choi et al., 2012). This variability leads to a departure from the general (or ideal) diurnal trend of ozone concentrations and thus leads to less accurate forecasting of method 2 in urban areas than in rural areas.

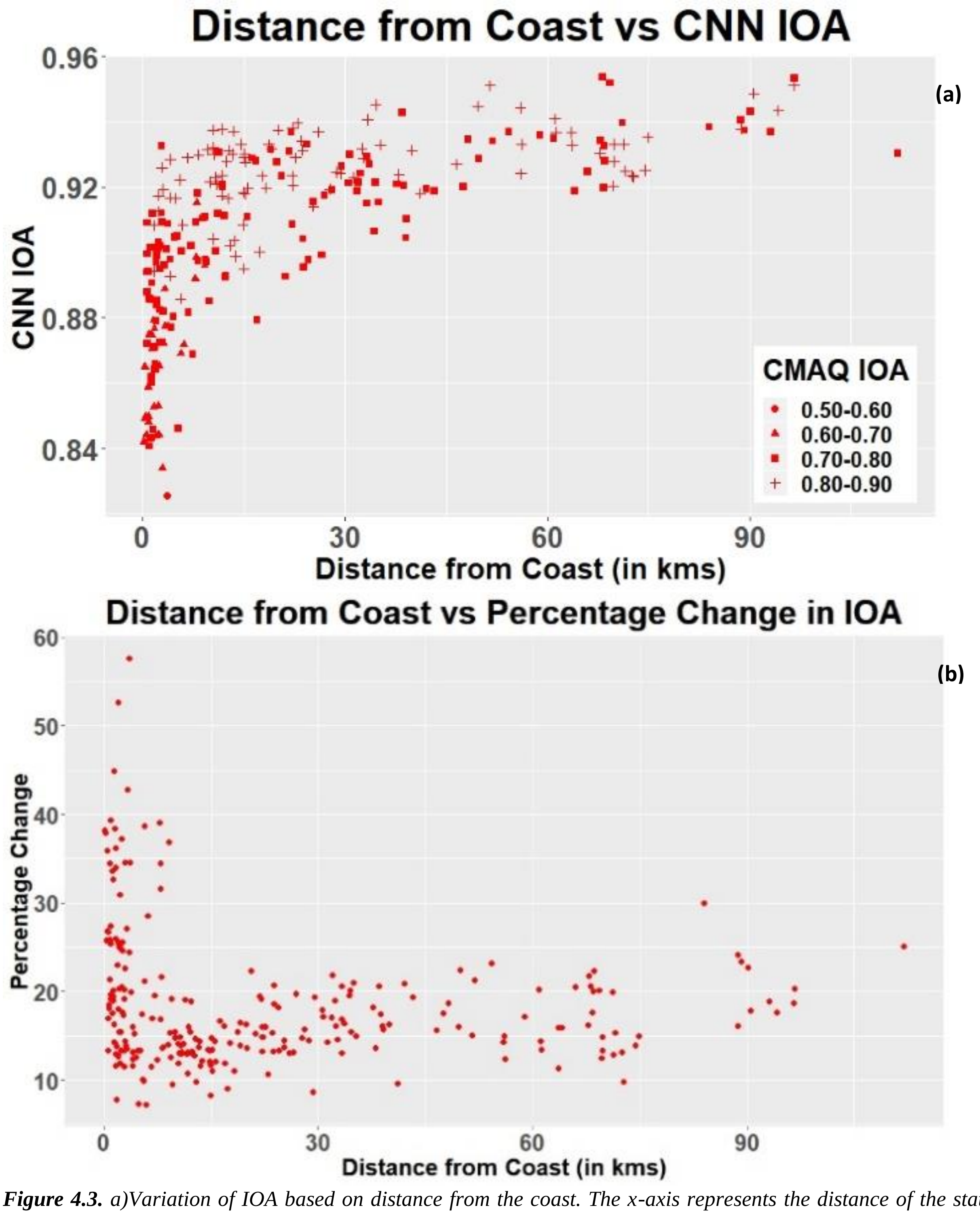


Figure 4.3. *a)Variation of IOA based on distance from the coast. The x-axis represents the distance of the station from the coast, and the y-axis represents the index of agreement. The colored symbols represent the range of CMAQ-IOA for the corresponding station. All IOA are based on one-day ahead prediction only. b) Percentage change in IOA based on distance from the coast. The figure shows the percentage increase in IOA of the CNN model-method 2 when compared with the IOA of the CMAQ model.*

Figure 4.3a shows station-wise IOA based on the distance from the coast. The symbol in the figure represents the bin of the CMAQ model's IOA. It was evident from the figure that the IOA for the CNN model- method 2 increases as the distance from the coast increases. The possible reason for the low IOA was that the CMAQ itself has a lower IOA near the coast. From figure 4.3b, it was evident the stations closest to the coast show more improvement compared to the stations away from the coast, in general. The increase in IOA with the CNN model – method 2 was in the range of 7%-55% when compared with the CMAQ model for day 1 forecast (Figure F19 – Appendix shows the station-wise IOA for the CMAQ and the CNN-method 2 model 1 day forecast). Among the stations on the coastal regions, those on the northwestern coast provided less accurate predictions than those on the northeastern and southeastern coastal cities (Figure F16). A possible explanation for such a trend could be the variability induced by long-range transport from China (Pouyaei et al., 2020). The effects of transport are observable at the three stations on Jeju Island. Because of transport from the Korean Peninsula, two stations (339111 and 339112) on the northern coast have a lower IOA (0.84 for both stations) than the one station (339121) on the southern coast (IOA - 0.90). As a mountain range separates the northern part of the island from the south, the transport was blocked. Note: Location of stations can be found in Figure F20 – Appendix.

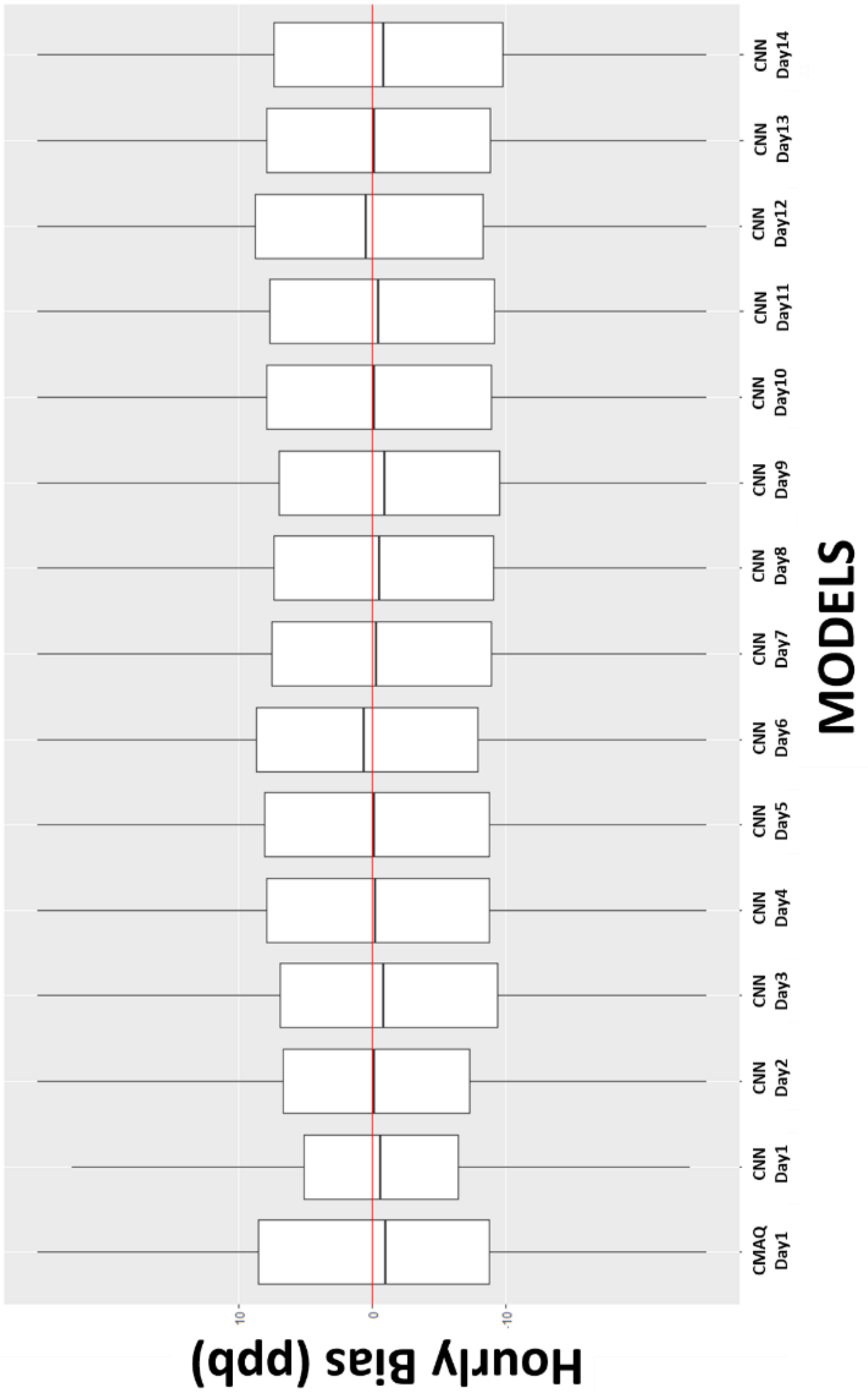


Figure 4.4. *Box plot of hourly bias of all stations combined. The x-axis represents the prediction days, and the y-axis represents the hourly bias in ppb. The Redline represents the zero bias, and the black horizontal line in each box represents the mean bias for that model.*

Figure 4.4 shows the boxplot for the hourly bias of all the stations combined for 14-day advance prediction. The bias for one-day advance prediction using the CNN model -method 2 was the least. As the number of advance prediction days increases, variability in the bias also increases, but the mean bias remains close to 0 for all days. The day 14 forecast has a similar bias as the one-day advance forecast by the CMAQ model. Although the mean bias remains close to zero, it should not be inferred that model performed similarly for all days. The bias resulting from a low observed value will be low, and the frequency of occurrence of a low value in a typical diurnal hourly ozone cycle was more frequent than the high value (Highs occurs only for 4-6 hours in a 24-hour period). Therefore, the evaluation of an air-quality model must be performed in conjunction with other metrics like IOA, correlation, or both. This was demonstrated in Figure F21 (Appendix), which shows a box plot of bias for a daily maximum ozone concentration for all stations (CNN method 2). The interquartile range (IQR) of bias for CNN Day-1 was lower compared to the CMAQ Day-1 bias. The IQR increases with subsequent days and for the 14th day, the mean of bias was -4.89 ppb. From the second day on, the CNN model initiated over predictions, which peaked around days 3and 4 and then began to decrease. Days 7 and 8 showed the fewest over predictions, and the mean of maximum daily ozone was close to the mean of the observations. The second week followed the same trend as that of the first week. Overprediction increased until the 9th and 10th days, and it decreases. The reason for such weekly trends in the IOA of prediction was that ozone concentrations also followed a weekly trend (Choi et al., 2012). Ozone concentrations were strongly auto-correlated with the seventh day, which provided better training of the CNN model for days 7 and 14; hence, the performance of the model on these days improved.

4.4. Conclusions

The predictive accuracy of the CNN model depended on one or a combination of multiple factors: i) the performance of the base model (in this case, CMAQ), ii) distance from the coast, iii) level of urbanization, and iv) transport. These factors, individually or in combination, led to a departure from typical diurnal ozone trends. As a result, an anomaly occurred, and in some cases, the model was not able to successfully understand the anomaly, which led to comparatively less forecasting accuracy. The model generally performs better when the CMAQ performs well. The quantification of variance in performance can provide a future direction for improving the performance further.

The variability caused by the cyclic reversal of land and sea breeze in ozone concentrations led to poor performance by the CNN model in the coastal region. Distance from that coast has an inverse effect on the prediction accuracy of this CNN model. The accuracy improves towards the inland. Similarly, as a less urbanized locale has a more consistent diurnal ozone trend, training of the CNN model becomes easier, enhancing its prediction accuracy.

The highly contrasting performance of the model, when applied to the western and eastern coasts of South Korea, suggests that transport also plays a significant role in determining the accuracy of model predictions. Unlike the east coast, the western coast was subject to long-range transport that adds to the variability of ozone trends. This hypothesis was supported by observations of the effects of transport at the three stations on Jeju Island.

Apart from affecting individually, the combination of these factors can also lead to poor performance. The low correlation over the northwest coast regions near Incheon was possibly from the combination of land-sea breeze and poor emission inventory. The airport located in Incheon was well-known for missing the NOx source (aircraft emission) in the emission inventory and thus affecting the CMAQ's performance. Seosan was a big industrial region located south of Incheon where both land-sea breeze and the NOx chemistry affect the typical ozone diurnal cycle. Also, the KORUS-AQ report mentioned that the emission data highly underestimated VOC emission over the region) which was another reason for the poor performance by the CMAQ model and the CNN model.

The developed CNN model was successful in reducing the uncertainties arising from the systematic biases in the system while, to some extent, unable to account for the uncertainties arising from high variability in the atmospheric dynamics, emission, and chemistry. The current systems for air quality prediction are either a short-term forecasting system or a low-accuracy system that covers a more extended forecasting period. Since this model provides a reasonable forecast two weeks in advance, it can provide an actionable window within which government agencies can deploy effective measures for reducing the occurrence of extreme ozone episodes.

CHAPTER 5

BIAS CORRECTING AND EXTENDING THE PM FORECAST BY CMAQ UP TO 7 DAYS USING DEEP CONVOLUTIONAL NEURAL NETWORKS[3]

5.1. Introduction

Concentrations of particulate matter (PM) arise from primary PM emissions and gaseous precursors, such as sulfur dioxide (SO_2), volatile organic compounds (VOCs), nitrogen oxides (NO_X), and ammonia (NH_3), through secondary formation in the atmosphere (Behera and Sharma, 2010; Hodan and Barnard, 2004). In addition, urban fine PM consists of chemical constituents such as organic carbon (OC), sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), trace metals, elemental carbon (EC), and organic matter (OM) (de Gouw et al., 2008; Zhang et al., 2015). Two components of NO_X, produced from various sources such as the combustion of fossil fuels (Noxon, 1978), biomass burning (van der Werf et al., 2006), soil microbial activity (Yienger and Levy, 1995), and lightning (Choi et al., 2009) are nitrogen dioxide (NO_2) and nitric oxide (NO). These compounds play an important role in the troposphere because they determine levels of ozone (O_3) and the formation of PM formation. A major air pollutant in the world, PM (Koulouri et al., 2008; Li et al., 2014; Mukherjee and Agrawal, 2017), comes in two aerodynamic diameters of fine particles: less than 10µm (PM_{10}) and less than 2.5µm ($PM_{2.5}$) (US EPA, 2016). World Health Organization guidelines of ambient air have defined hourly mean concentration thresholds of 25 $\mu g/m^3$ for $PM_{2.5}$ and 50 $\mu g/m^3$ for PM_{10} (WHO, 2018). Short-term exposure to ozone and PM has been linked to respiratory and cardiovascular diseases and mortalities (Brunekreef and Holgate,

[3]Published: Sayeed, A., Lops, Y., Choi, Y., Jung, J., Salman, A.K., 2021b. Bias correcting and extending the PM forecast by CMAQ up to 7 days using deep convolutional neural networks. Atmospheric Environment 253, 118376.

2002), and estimates show that over two million deaths per year resulting from damage to the respiratory system are associated with PM pollution (Kim et al., 2015; Shah et al., 2013). Thus, to mitigate $PM_{2.5}$ and PM_{10} pollutants, researchers, in an effort to reduce local emissions before thresholds are reached, have focused on developing strategies to forecast PM more accurately.

The development of new technology has enabled atmospheric scientists to receive an overwhelming amount of data on air quality from space, in-situ monitoring sites, and numerical simulations. These data sources provide significant unexploited opportunities to improve the understanding of atmospheric constituents and to model and forecast these processes. Researchers have mainly used historical meteorological data to forecast weather and atmospheric constituents by applying statistical models. Some, however, have applied chemical transport models (CTMs), which are often dependent on the accuracy of numerical weather prediction models and initial input data to accurately estimate distributions of atmospheric constituents.

Although CTMs have shown significant improvement in the last two decades, they are unable to provide fully reliable air quality forecasts. Their reliability was especially low in topographically complex regions because of their shortcomings in horizontal resolution, physical parameterizations, and initial- and boundary conditions. Overcoming the limitations of both numerical and statistical model require comprehensive observational data for model tuning and selection (i.e., estimating the best possible parametrizations) and data assimilation (i.e., estimating the system state for more accurate predictions), both of which can use the same data.

The most widely used air pollution chemical transport model was the Community Multiscale Air Quality (CMAQ) model, developed by the United States Environmental Protection Agency (Byun and Schere, 2006). To improve the reliability of CMAQ for accurately forecasting atmospheric constituents' concentrations, several studies have applied pollution transport and trajectory methods. While they have improved the transport modeling and simulating trajectories in CMAQ, cases of significant bias in higher concentrations of constituents are still present (Hur et al., 2021; Jeon et al., 2016; Pouyaei et al., 2020). To bridge the gap between observations and models, several studies that have applied data assimilation techniques have incorporated ground observations and remote sensing products for atmospheric chemistry and weather forecasting and reported significant improvement in model accuracy (Bocquet et al., 2015; Jung et al., 2019).

Another significant parameter affecting the performance of CMAQ was emissions. For cases regarding NO_2, CMAQ uses bottom-up NO_X emissions to simulate temporal and spatial variations in ozone (Li et al., 2016); several problems involving the simulation of NO_X, however, have arisen (Choi and Souri, 2015). For one, these emissions are susceptible to rapid obsolescence resulting from both rapid changes in anthropogenic emissions and the fast response of NO_2 to these emissions, resulting from the relatively short lifetime (Martin et al., 2003). In addition, the uncertainties of NO_X emissions within a region are even more pronounced when studies use different emissions inventories in their models, which generate significant changes in concentrations (Choi and Souri, 2015). Hence, because of the atmospheric transport of emissions (Sadeghi et al., 2020) and non-linear interactions (Pandis, 2004), the difficulty of estimating them under all possible conditions was apparent (Davidson et al., 2005). In addition, as meteorological factors are significant in the formation of atmospheric constituents, which further emphasize the non-linear process within the atmosphere, they should also be considered in forecasting (Ghahremanloo et al., 2021; Memarianfard et al., 2017). Due to their inherent non-linear algorithm, machine learning (ML) and deep learning (DL) algorithms have shown significant potential in addressing the limitations of CMAQ (e.g., simulating non-linear processes).

Various ML and DL algorithms have been implemented in a variety of fields in the atmospheric sciences to forecast atmospheric constituents (Biancofiore et al., 2017; Díaz-Robles et al., 2008; Eslami et al., 2019a; Hooyberghs et al., 2005; Lops et al., 2019; Sayeed et al., 2021a, 2020a) and improve current models in forecasting or modeling air quality (Pouyaei et al., 2020; Sayeed et al., 2020b). A significant number of these papers have implemented convolutional neural network (CNN) models, which are capable of joint feature and classifier learning. In addition to demonstrating greater accuracy with large-scale datasets (LeCun and Bengio, 1995), CNN models used for feature extraction are more efficient than other neural network methods, particularly when multiple hidden layers are structured. This feature of CNNs has been responsible for significant advancement in classification (Scarpa et al., 2018) and image processing in the atmospheric sciences and other diverse set of applications (Krizhevsky et al., 2017; Lawrence et al., 1997). The research in this study focuses on expanding already developed DL forecasting systems (Eslami et al., 2019a; Lops et al., 2019; Sayeed et al., 2021a, 2020a) by integrating observational and CTM data to improve and extend the forecasting capability of DL forecasting systems. This process addresses several limitations of DL models in forecasting NO_2, $PM_{2.5}$, and PM_{10} with only in-situ observational data. This research also demonstrates the capability of DL models to optimize the output of current CTMs that experience significant biases of atmospheric chemical constituents in short and long-term forecasts.

Unlike conventional machine learning methods, deep CNNs are capable of automatically identifying the most informative required features, which facilitates predictions from the input-output information of the bias-correction and long-term forecasting system used in this paper.

5.2. Material and Methods

This study focuses on two tasks: a) improving the air-quality forecast of the CMAQ model by using a convolutional neural network (CNN) model, and b) obtaining high-quality forecasts for up to 7 days. South Korea was selected as the study region because of its diverse topography, small spatial scale dimensions, active pollution areas, and air pollution transport from surrounding regions. The observational data were acquired from the National Institute of Environmental Research (NIER) in South Korea for 404 air-quality monitoring stations, the locations of which appear in Figure 5.1. The stations monitor and record hourly average concentrations of $PM_{2.5}$, PM_{10}, carbon monoxide (CO), sulfur dioxide (SO_2), ozone (O_3), and nitrogen dioxide (NO_2). The data cover a period from 2016 to 2019. The CNN model proposed in this study requires both observed pollutant concentrations from the 24 hours and air quality and meteorology forecasts from the following 24-hour period from numerical models. The meteorology was generated using Weather Research and Forecasting (WRF) model version 3.8 and processed with Meteorology Chemistry Interface Processor (MCIP). The air quality forecasts were obtained by running the CMAQ v5.2. Detailed configurations of the CMAQ and WRF models can be obtained from Jung et al. (2019). The WRF and CMAQ models were run for the years 2016 to 2019.

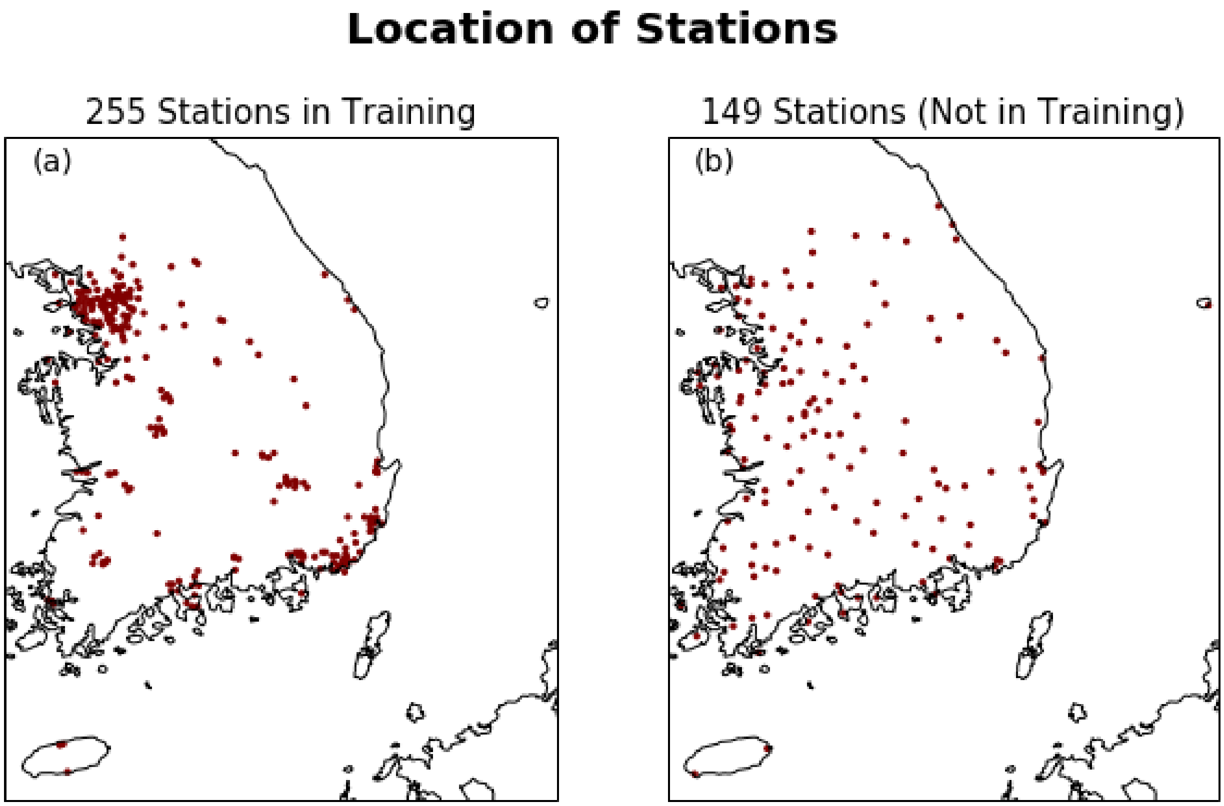


Figure 5.1. Map of South Korea showing: a) the locations of the 255 stations used in training; b) map of the locations of the 149 stations (i.e., out-of-box stations) not used for the training, but instead for validating the model.

In this study, three air-quality parameters were forecasted -$PM_{2.5}$, PM_{10}, and NO_2 -seven days in advance. Figure 5.2 shows the schematic diagram of the arrangement of inputs and output of the CNN model used in this study. For each parameter and each day, a specific model was trained, so 21 different CNN models were trained. The inputs described in Section 2.1 were the same for each model but with a different target (output). Since the Day 1 CNN model uses the CMAQ model's day one forecasts as inputs, it was essentially a bias-correction model, while days 2-7 are forecasts based on the day 1 forecast of the CMAQ model. The models were trained for the years 2016 to 2018 with 255 stations (252 for $PM_{2.5}$). The remaining 149 stations were not used in training but for the spatial validation of the CNN model (Figure 5.1b). In addition, for temporal validation, $PM_{2.5}$, PM_{10}, and NO_2 were forecasted separately for each of the seven days for the year 2019. For each station, the closest grid point was located on the numerical model grid and assigned the grid to the station. To ensure a generalized model, the inputs for each station were prepared and stacked over one another, and randomly shuffled for training. Since there was a large pool of training examples (1096x255), days with any missing observational data were removed from the training.

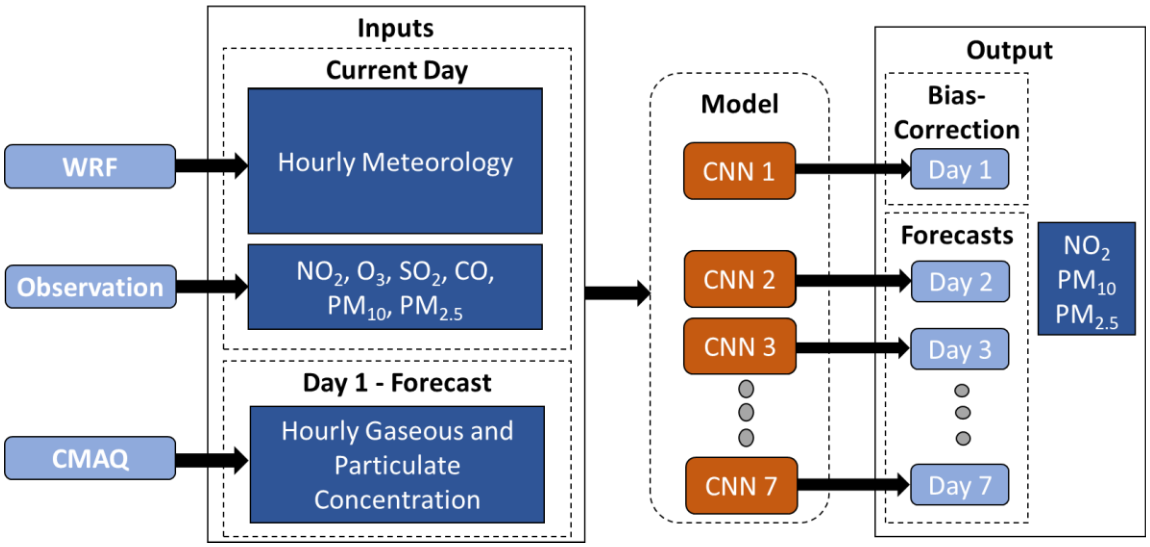


Figure 5.2. *Schematic diagram of the CNN models used in this study. 21 different CNN models were prepared based on each day (7 days) and air quality species (3 species). Day 1 CNN model was a bias correction model for CMAQ day 1 forecast, while Day 2-7 are forecasts of the CNN model. Including CMAQ day 1 forecast extends the forecasting capability of the CNN Models up to 7-days.*

5.2.1. CNN Model

In this study, a state-of-the-art convolutional neural network (CNN) model was used. The architecture used was similar to that used by Sayeed et al. (2021a). The input consists of 24-hourly forecasts of 31 meteorologies (see Table T6 in the appendix), 10 air-quality constituents (see Table T6 in the Appendix) and six current air-quality parameters (CO, SO_2, NO_2, PM_{10}, $PM_{2.5}$, O_3) observed daily. Thus, the number of parameters, which form the input layer to the CNN model, was 1,128 (24×47). The input layer was followed by five convolutional layers, each consisting of 32 filters and convolved through a kernel of size 2×1. The output of the five-layer convolutions was a two-dimensional array, which was then converted to a one-dimensional array using flattening. The flattening layer was followed by a dense layer with 264 neurons that were processed through an output layer with 24 parameters. In the end, a 24-hour forecast was obtained of the desired air-quality parameter. The activation function used in each layer was ReLU. Table T15 (Appendix) provides a detailed description of the architecture of each layer used in the CNN model. Since a deep neural network system was an optimization problem, in which the quality of a model depends on the feedback mechanism (i.e., maximization or minimization of the loss function) used to optimize the model, a loss function specific to air-quality studies (Sayeed et al., 2021a) was developed. The loss function used in this study was based on the index of agreement (IOA) (Willmott et al., 1985).

5.3. Results and Discussion

The performance of the models was evaluated on both temporal and spatial scales. For the spatial evaluation process, the models were trained on only 255 stations out of the available 404 stations and reserved the remaining 149 stations for the spatial evaluation process (see Figure 5.1b). For this evaluation, only stations with more than a month of observations (more than 720 hours) were selected in 2019 (141 stations for NO_2, 138 stations for $PM_{2.5}$, and 140 stations for PM_{10}). For the temporal evaluation process, the models were trained from 2016 to 2018; the model then produced forecasts for all of 2019, which were compared to in-situ measurements. The IOA was used to evaluate the hourly performance of the model and categorical statistics (Eder et al., 2006; Chai et al., 2013) to evaluate its daily performance based on the daily maximum values of $PM_{2.5}$ and PM_{10}.

Figure 5.3a shows the IOA for the hourly forecasts of NO_2 generated by the CMAQ and CNN models. The blue bars in the figure represent the average IOA of all stations during the training phase, and the orange bar represents the average IOA of all the stations that were not included in the training phase (out-of-box stations). Figure 5.3a demonstrates the success of the

CNN model at improving the average IOA for first-day forecasts from the CMAQ model from 0.6 to 0.86 for the stations used for training and from 0.6 to 0.83 for the stations used for validation. Although the forecasts of the IOA from the CNN models saw less and less improvement from day to day, the seventh-day forecast from the CNN model still outperformed the one-day forecast from the CMAQ model. In fact, the improvement in the forecasted IOA from CNN on the seventh day was 10% higher than that from the CMAQ on the first day. The lowest forecasted IOA was for the fifth day. The performance of the sixth- and seventh-day forecasts, however, were equivalent to that of the third-day forecasts (see Table T16 in the appendix).

Figure 5.3b represents the spatial IOA distribution for the first-day forecast by both CMAQ and CNN models. The performance of the CNN model was generally dependent on the performance of the CMAQ model for the station locations (Sayeed et al., 2021a). The stations that are farther away from the coast performed slightly better than those near the coast. The main reason for such variation in performance was due to issues with the performance of CMAQ in the coastal regions (Sayeed et al., 2021a). In addition, some out-of-box stations on the eastern coast performed poorly because of an insufficient number of stations in the region for training. The errors were reduced further, based on geospatial location, by including topographical and demographical details of the stations as input to the CNN model. The regions with more stations for training show comparatively stronger performance than those with fewer stations available for training (a smaller cluster).

Figures 5.4a and b show the average yearly IOA of the CNN model for $PM_{2.5}$ at all stations for the 7-day forecast and the station-wise yearly IOA of the CMAQ and CNN models for the first-day forecast, respectively. For the first day, the CNN model improved the $PM_{2.5}$ forecast by ~13%; for the second day, however, the CNN model generated a similar IOA to that of the first-day

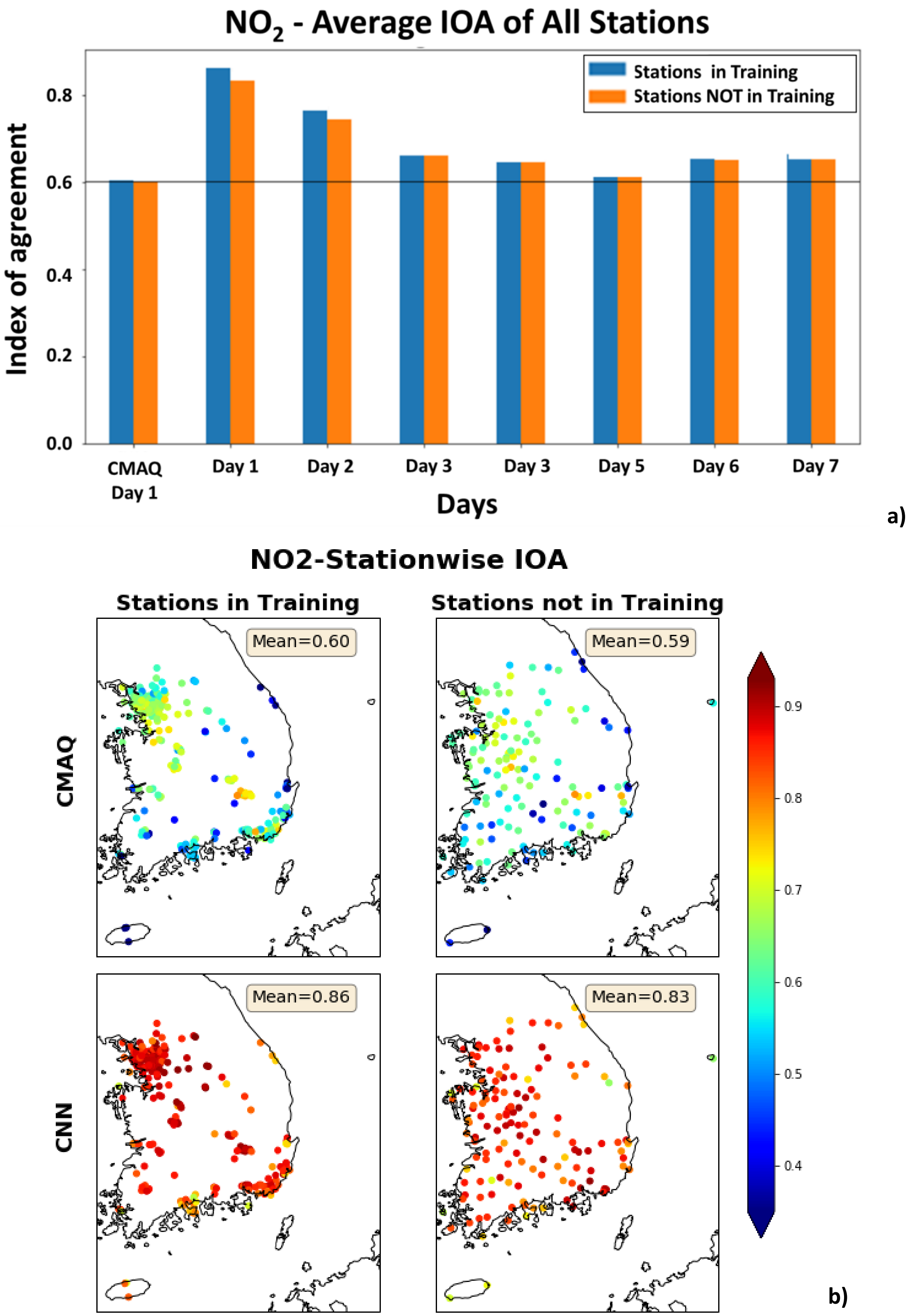
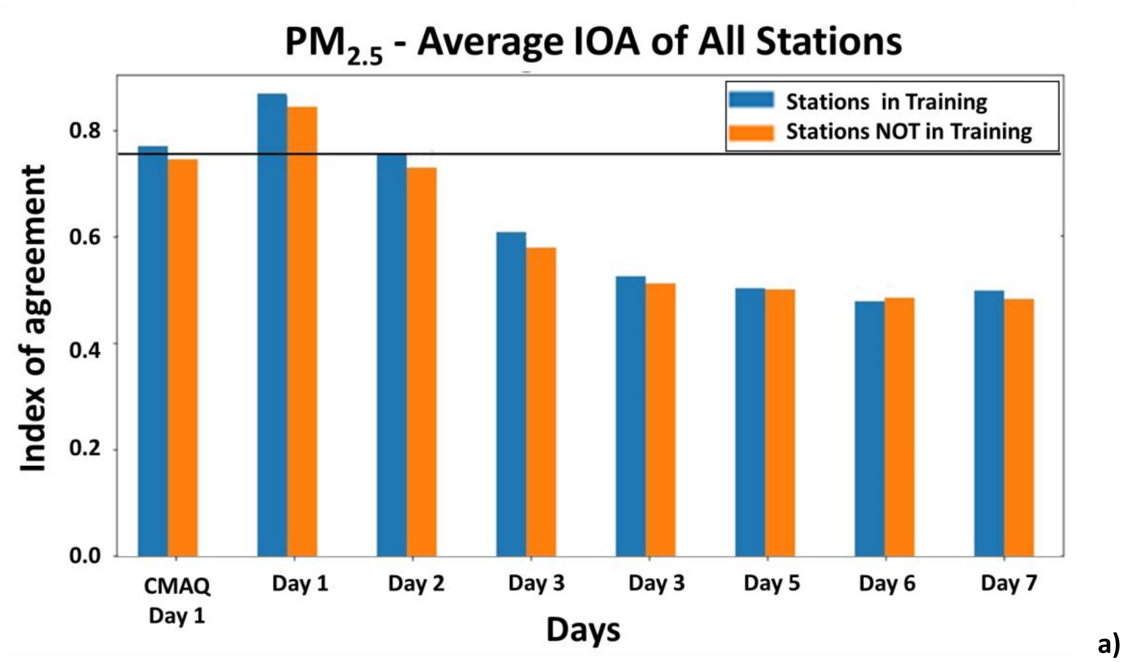


Figure 5.3. *a) Yearly IOA of all stations (averaged) of the CMAQ and the CNN models for forecasting NO_2. b) Station-wise yearly IOA of both models for the one-day forecast of NO_2.*

forecast of the CMAQ model. Nevertheless, for the days following the first day, the accuracy of the CNN forecasts declined. The IOA increased in all stations, except for stations 336471 and 339131, which showed decreases of less than 1%. Stations 525162 (see Figure F25 in the appendix) and 525143 had the lowest IOA. These stations had an unusually high value of 500 μg/m^3 on several days, significantly impacting the performance of the models. These abnormal values may have resulted from measurement errors (Figure F28 in the appendix shows the location of these stations). While the IOA performance for the second-day forecast of the CNN model was equivalent to that of the CMAQ second day forecast (see Table T17 in the appendix), the CNN model was much faster than CMAQ (it takes 2-3 minutes to forecast all stations), and can be used as an alternative to the CMAQ model. (Note: More time series examples can be found in figures F22- F27)

Figures 5.5a and b show a comparison of the performance of the CMAQ and CNN models for PM_{10} 7-day forecasts. The CNN model for PM_{10} increased the average IOA by ~22% for stations used in the model training and 21% for the stations used for the spatial validation. The second-day CNN forecasts for PM_{10} were more accurate than the CMAQ first-day forecasts. From the third day onwards, however, the performance of the CNN model decreased drastically. Figure 5.4b shows a significant improvement in the station-wise IOA for stations in both the training and validation phases. The increase in the IOA was greater than 10% for 387 out of 393 stations. For 138 stations, the IOA increased by more than 25% when comparing the CMAQ model to the CNN model. The smallest increase in performance was 4% for station 238471. The performance of the CNN model for PM_{10} was better than the CMAQ model by up to two days (see Table T18 in the appendix). In addition, the faster computational speed of the CNN model makes it a good candidate for PM_{10} forecasts for up to two days.



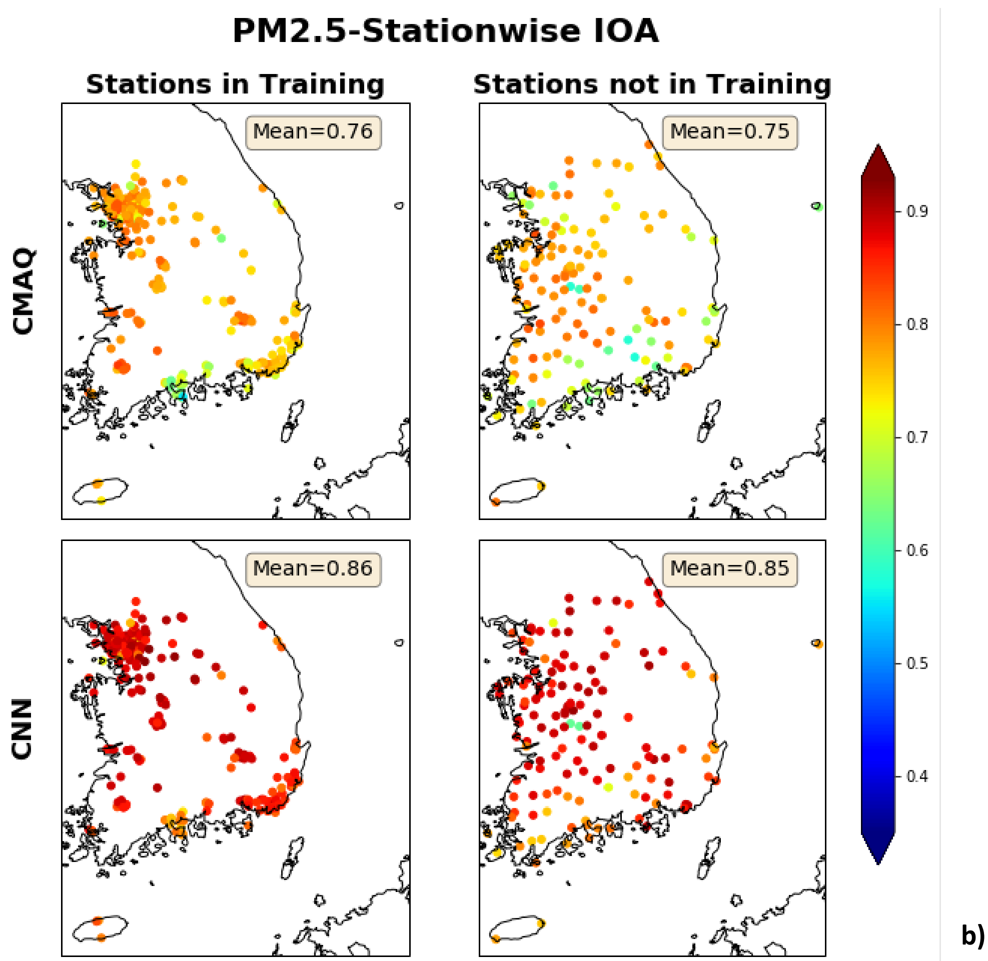


Figure 5.4. *a) Yearly IOA of all stations (averaged) of both CMAQ and CNN models for forecasting $PM_{2.5}$. b) The station-wise yearly IOA of both models for the one-day forecast of $PM_{2.5}$.*

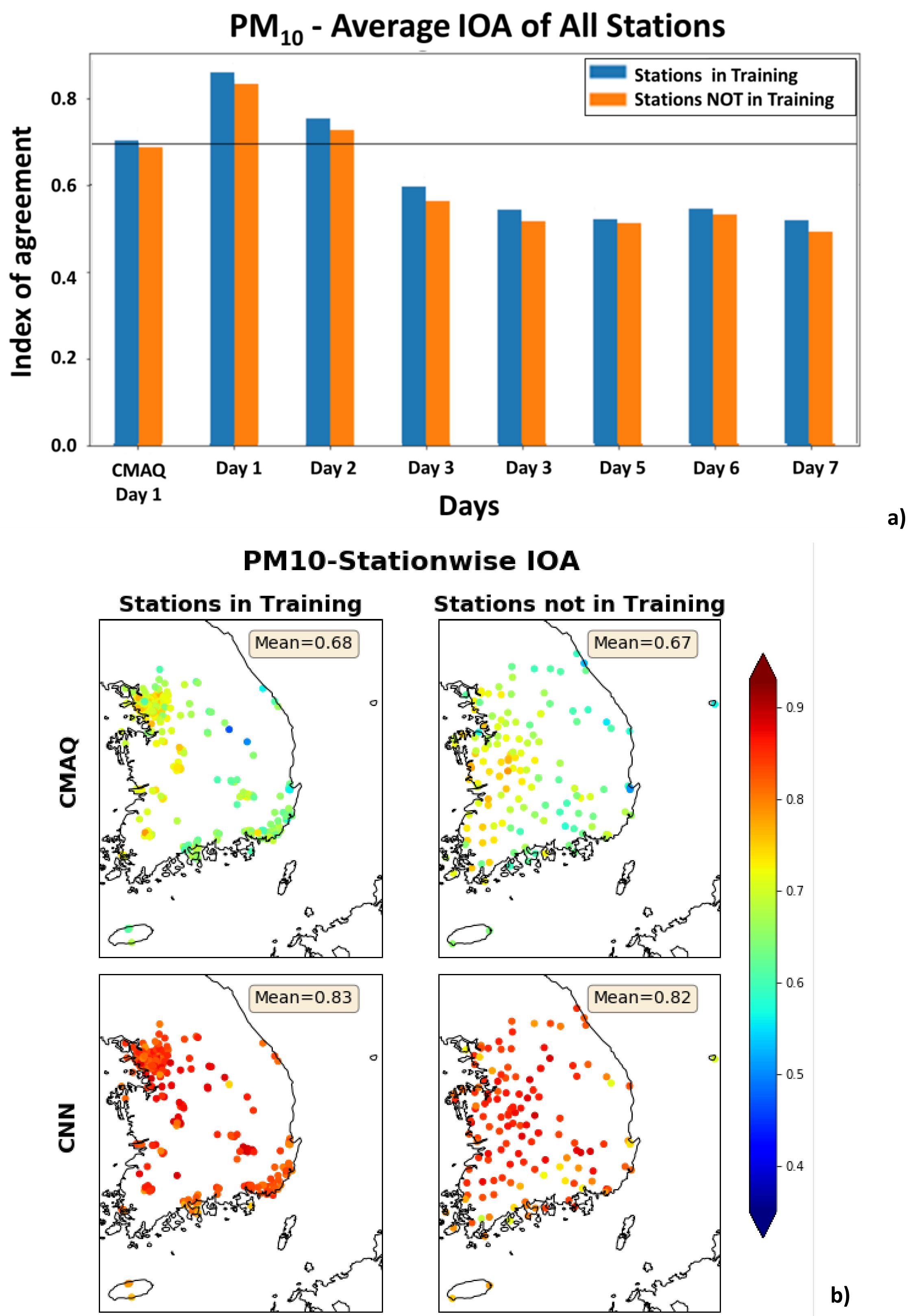


Figure 5.5. *a) Yearly IOA of all stations (averaged) generated by the CMAQ and CNN models for forecasting PM_{10}. b) Station-wise yearly IOA generated by both models for the one-day forecast of PM_{10}.*

While the accuracy of hourly forecasts of NO_2 was reasonable for up to 7 days in advance, those of $PM_{2.5}$ and PM_{10} were only reasonable for up to 2 days in advance. Possible explanations for this finding follow: i) more uniform uncertainty in the CMAQ forecasts of NO_2 than for those of $PM_{2.5}$ and PM_{10}; ii) the diurnal cycle of NO_2, which was mostly high during heavy traffic time (mornings and evenings); and/or iii) the effect of the transport of PM from adjoining regions. All of these factors contribute to the ease with which an AI model can be trained. Because of the high variability of PM throughout the day, it was difficult for the model to achieve a high level of accuracy for a longer period. For particulate matter, the model produced stronger forecasts for the less variable $PM_{2.5}$ than it did for PM_{10}; it also produced more accurate forecasts for the more uniform diurnal ozone and a longer forecast duration (Sayeed et al., 2021b)

5.3.1. Categorical Performances:

The performance of the model was evaluated by daily maximum values. To evaluate the model, categorical statistics (Chai et al., 2013, Eder et al., 2006) was used and pairs of observations and predictions were divided as follows:

a) N_a, number of days when an observation was below the threshold and a prediction was above.

b) N_b, number of days when both observations and predictions are above the threshold.

c) N_c, number of days when both observations and predictions are below the threshold.

d) N_a, number of days when an observation was above the threshold and a prediction was below.

After categorizing, the observations and predictions, the following metrics were defined based on the following: the hit rate (HIT), which represented the capability of the model to correctly forecast an extreme event (i.e., an event above the threshold); the false alarm rate (FAR), which represented times when the model falsely forecasted an extreme event; and the equitable threat score (ETS), which defined the skill of a model on a scale of -1 to 1, in which 1 indicated that the model was skillful; and the proportion of correctness (POC), which defined times the model was able to correctly predict the occurrence of an event (both exceedances and non - exceedances).

$$HIT = \frac{N_b}{N_b + N_d} \tag{5.1}$$

$$FAR = \frac{N_a}{N_a + N_b} \tag{5.2}$$

$$ETS = \frac{N_b - N_r}{N_a + N_b + N_d - N_r} \tag{5.3}$$

$$\text{where } N_r = \frac{(N_a + N_b) \times (N_b + N_d)}{N_a + N_b + N_c + N_d} \tag{5.4}$$

$$POC = \frac{N_b + N_c}{N_a + N_b + N_c + N_d} \tag{5.5}$$

The thresholds were set based on the WHO guidelines for ambient air. The standard for NO_2, PM_{10}, and $PM_{2.5}$ was a 100 ppb one-hour mean, 50 µg/m^3 for the daily maximum, and 25 µg/m^3 for the daily maximum, respectively. The categorical statistics were inconclusive for NO_2 because the NO_2 in-situ measurements did not exceed the 100-ppb threshold. Without threshold exceedances, the categorical analysis could not be performed. Table 5.1 represents the categorical statistics for PM_{10} based on the 50 µg/m^3 threshold, averaged over all stations. The average hit rate of the CNN model at all stations increased by 95%, but the false alarm rate remained almost equivalent to that of the CMAQ model. The equitable threat score increased by 123%, from 0.22 for the CMAQ model to 0.42 for the CNN model. The POC, which signifies the times a model accurately predicts an event, also increased by ~9% for the CNN model.

Table 5.1. *Categorical statistics; average of all of the stations for PM_{10}.*

PM_{10}	HIT	FAR	ETS	POC
CMAQ	0.42	0.32	0.22	0.80
CNN	0.82	0.31	0.49	0.87

Table 5.2 represents the categorical statistics for $PM_{2.5}$ based on the 25 µg/m^3 threshold, averaged over all stations. The CMAQ model performed comparatively better for $PM_{2.5}$ than it did for PM_{10}. Since the CNN model was based on the CMAQ model, the performance of CNN for $PM_{2.5}$ was also better than it was for PM_{10}. While the hit rate increased by 12%, the false alarm rate decreased by ~40%. The ETS skill of the CNN model was ~68% higher than that of the CMAQ model. The POC also increased by ~14%, indicating that the forecasts of the CNN model were 14% more accurate overall than those of the CMAQ model.

Table 5.2. *Categorical statistics; average of all of the stations for $PM_{2.5}$.*

$PM_{2.5}$	HIT	FAR	ETS	POC
CMAQ	0.75	0.38	0.32	0.77
CNN	0.84	0.23	0.54	0.88

5.4. Conclusions

This research demonstrates the capability of AI models to address the non-linear relationships of in-situ measurements and model output. Integrating the AI system and the CMAQ model significantly reduced the biases of CMAQ model output and extended the forecasting capability up to 7 days in advance. The NO_2 bias correction and forecasting of the AI model showed the most significant improvement. The model showed same-day bias correction, improving IOA, on average, by 38% for all stations; in addition, its seventh-day forecast of NO_2 still exceeded the accuracy of the one-day forecast by the CMAQ model. Improving the $PM_{2.5}$ and PM_{10} bias correction of the CMAQ model also improved the IOA by 13% and 21%, respectively. Comparing the forecasting time of the CMAQ and AI models, the AI model's two-day forecast was able to maintain the one-day forecast performance of the CMAQ model. The AI model was not able to maintain the one-day forecast performance of the CMAQ model on the third day, which significantly declined, and finally tapered off beyond the fourth day.

A significant benefit of the AI bias correction system was its near real-time processing capability. While the training time of the AI model may require time to process, deployment of the AI system for bias correction performs in near real-time. Thus, the integration of the AI system for model output optimization will not have a significant or noticeable impact on the overall model run time. In addition, the performance of the AI model via a spatial validation process (see Figures 5.3-5.5) further emphasizing its scalability for locations and adaptability to changes in regional meteorology and atmospheric constituents.

The AI bias correction model was still subject to limitations, particularly with regard to the extended forecasting of $PM_{2.5}$ and PM_{10}. One reason was the dependency of AI model performance on CMAQ model performance. However, once the CMAQ model improves its forecasting and simulation of short-term atmospheric constituents, the AI model will adapt to such improvements and further optimize and extend the forecasting days with sufficient accuracy. The system was also limited to effective implementation in the locations of in-situ measurements. Thus, the integration of remote sensing data as input to the AI model has the potential to expand the capability of the AI system to bias correct each grid cell of the CMAQ model.

CHAPTER 6

A DEEP CONVOLUTIONAL NEURAL NETWORK MODEL FOR IMPROVING WRF SIMULATIONS[4]

6.1. Introduction

The atmosphere sciences, particularly weather forecasting, have at their disposal a deluge of data from space, in-situ monitoring, and numerical simulations. These diverse data sources offer new opportunities, still largely underexploited, to improve the understanding, modeling, and reconstruction of geophysical dynamics. Several academic studies devoted to the problem of forecasting difficult-to-retrieve weather events and their associated uncertainties typically employ weather forecasting techniques that fall into three main categories: numerical weather predictions (NWP), statistical forecasting, and artificial intelligence (AI - forecasting). Dynamical (physical) models such as the Weather Research and Forecasting (WRF) model use meteorological and topological information to determine the mesoscale weather parameters of a specific region (Su et al., 2014), and statistical methods mainly use historical meteorological data to simulate the state of the weather (Eslami et al., 2019a, 2019b; Lops et al., 2019; Nie et al., 2020; Sayeed et al., 2020b).

To obtain the various meteorological parameters, NWP models generally entail the parameterization of physical phenomena using initial and boundary conditions and a series of partial differential equations (Wilgan et al., 2015). Unfortunately, despite advancements in these models, the shortcoming in resolving horizontal grid resolutions through discretization and

[4]Revised Submission: Sayeed, A., Choi, Y., Jung, J., Lops, Y., Eslami, E., Salman, A.K., 2020c. A Deep Convolutional Neural Network Model for improving WRF Simulations. (2nd revised manuscript submitted to IEEE-Transactions on Neural Networks and Learning Systems)

interpolation has led to unreliable weather simulations (Cassola and Burlando, 2012). NWPs are also computationally expensive, particularly with regard to fine-resolution simulations (Sayeed et al., 2020b). In addition, because of the misrepresentation of unresolved small-scale features or neglected physical processes, parts of numerical models are represented by empirical sub-models or parameterizations (Crétat et al., 2012; Lu et al., 2013; Stensrud, 2007), which tend to simplify involved physics that may lead to uncertainties in simulation.

Unlike NWPs, statistical models require a large amount of historical data and completely neglect the physics of the atmosphere; thus, they do not consider meteorology (Erdem and Shi, 2011; Nie et al., 2020). Since statistical methods are easily implemented and less computationally intensive than NWPs, they are popular among researchers. Nevertheless, owing to the scarcity of representing complex meteorological phenomena and non-linear patterns in the training data, statistical models are unreliable and inaccurate for simulating extreme weather episodes, which exacerbate long-range forecasting.

Because of the chaotic nature of the weather system, errors in weather models are unavoidable but quite often significant regardless of the implemented modeling approach. The parametrization of physical process and discretization of differential equations lead to biases, which increases at every step of space and time in a numerical model. Overcoming these limitations still remains a challenging task. In the past several decades, the volume and quality of observations have increased dramatically, particularly thanks to remote sensing. At the same time, new developments in machine learning (ML), particularly deep learning (DL) (LeCun et al., 2015), have demonstrated impressive capabilities at reproducing complex spatiotemporal processes (Tran et al., 2015) by efficiently using an enormous amount of data, thus creating a path for their use in the atmospheric sciences.

Researchers have applied various ML algorithms in a variety of fields in the earth and atmospheric sciences, including air quality forecasting (Eslami et al., 2019a, 2019b; Lops et al., 2019; Sayeed et al., 2020b) and hurricane tracking (Eslami, 2020). ML has also been applied to nowcasting based on real observations such as the sea surface temperature (Bézenac et al., 2019) and precipitation (Shi et al., 2017). Most studies for the bias-correction using statistical methods or machine learning methods focus on only one meteorological parameter, or the temporal resolution was very coarse (3-hourly to daily mean values) (Costoya et al., 2020; Holman et al., 2017). Moreover, most studies use the CNN model for either image processing or classifications of images. The CNN models were used as a non-linear regressor for air-quality forecasts (ozone and pollen) (Eslami et al., 2019a, 2019b; Lops et al., 2019; Sayeed et al., 2020b). In this study, an alternative approach was applied: a fully data-driven framework that combines a deep neural network (CNN) and physical models (WRF) that simulate the dynamics of a complex weather system. A weather-AI as real-time weather simulating model was developed that reduces the model-measurement error of the WRF model. The model developed can be used to bias-correct any observed meteorology parameter for an hourly temporal resolution. The system, using a convolutional neural network algorithm (Krizhevsky et al., 2017), post-processes and bias-corrects the WRF output (observation network of the 24-hour simulations) in real-time at each grid linked to a station location.

6.2. Material and Methods

The algorithm was divided into two sections: i) hourly simulation by a WRF model and ii) a deep CNN model that reduces uncertainty and improves simulation accuracy. Figure 6.1 shows the process flow diagram for the Weather-AI model.

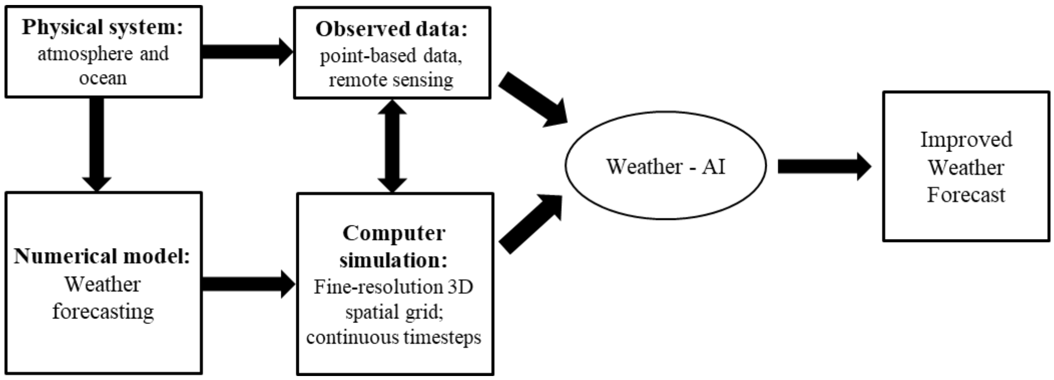


Figure 6.1. *Process flow for the Weather - AI model in bias correcting WRF forecasts. A Weather-AI model uses historical simulation by a numerical model (WRF) and uses the actual observation to understand the biases. The process was called training an AI model. Once a model was trained, it was used to forecast unseen scenarios.*

6.2.1. Deep Convolutional Neural Network

The deep architecture of the convolution neural network (CNN) used in this study was similar to the model (Sayeed et al., 2020b). In general, CNN models are used for image processing and image classifications. The CNN models have proven to be an effective tool in developing regression models for air-quality forecasts (Eslami et al., 2020, 2019a; Eslami, 2020; Sayeed et al., 2021b, 2020b, 2021; Yeo et al., 2021). Since all meteorological parameters are inter-dependent, the convolution feature of a CNN layer provides an excellent tool for convolving different paraments. Several recent studies have tried to leverage this feature by either convolving a single parameter in time and variables using 1-D CNN (Wang et al., 2020) or established spatial inter-correlation using 2-D CNN (Hong and Satriani, 2020). Although these studies have shown some promising results, the forecast was either a shifted time-series (Wang et al., 2020) by the same amount as the prediction window, or they don't provide time-series to evaluate the shifting (Hong and Satriani, 2020). For this study, a generalized architecture of the CNN model was developed, capable of bias-correcting various weather parameters modeled (by WRF) weather parameters. The use of the numerical model (WRF) enabled us to remove the shifts in time series by removing the previous day's observations from the inputs for forecasting. The model entails five one-dimensional convolutional layers (Figure 2.1, which shows the model architecture), a fully connected layer, and an output layer. Each convolutional layer, with 8, 16, 32, 32, and 32 filters, respectively, was activated by the rectified linear unit. To find the best architecture, several architectures were tested for wind speed. Figure F30 shows the comparison of the CNN model with a different number of layers for windspeed, u-wind, and v-wind, respectively. However, the model with 3 layers performs equivalent to the model with 5 layers for wind speed. The results from the figure show the model with 5 layers performs better for u-wind and v-wind components over the CNN models with fewer layers. To have generalized model architecture, a 5-layer model was used for this study. The input for the first layer consists of various hourly meteorological parameters extracted from the WRF model (Table T19 in the Appendix lists all the WRF meteorological parameters used as input). The convolutions are applied to the input features with the elements of a randomly initialized kernel (with a kernel window size of 2 × 1). The feature maps are obtained through the output of the first layer, then used as input for the second layer. The same process was applied in the succeeding layers. The output of the fifth convolutional layer was then passed to the fully connected layer, which contains 264 nodes (neurons) (selected by using grid-search CV (sklearn.model_selection.GridSearchCV — scikit-learn 0.24.1 documentation). Furthermore, several learning rates, optimizers and batch-sizes were tested for the model and the best configuration (based on highest IOA and correlation on out of box test set) was selected (i.e., 0.001 learning rate, adam optimizer (Kingma and Ba, 2014), and 72 batch size). The hourly output was obtained at the last layer (output layer). A deep CNN, like any neural network, was an optimization problem that attempts to minimize the loss function. In general, deep learning models use mean squared error or mean absolute error as loss function to optimize the model. In this study, the loss function developed by Sayeed et al. (2021a), based on the index of agreement (IOA) (Willmott et al., 1985), was used.

6.2.2. Data Preparation and Model Training

The observed meteorology was obtained from the 93 Automated Synoptic Observing System (ASOS) stations operated by the Korea Meteorological Administration (KMA) for the years 2014 to 2018 across South Korea. Figure 6.2 displays the location of all the meteorology monitoring stations in the country. The meteorological parameters obtained from these stations were wind speed, wind direction, precipitation, relative humidity, temperature, dewpoint temperature, and surface pressure.

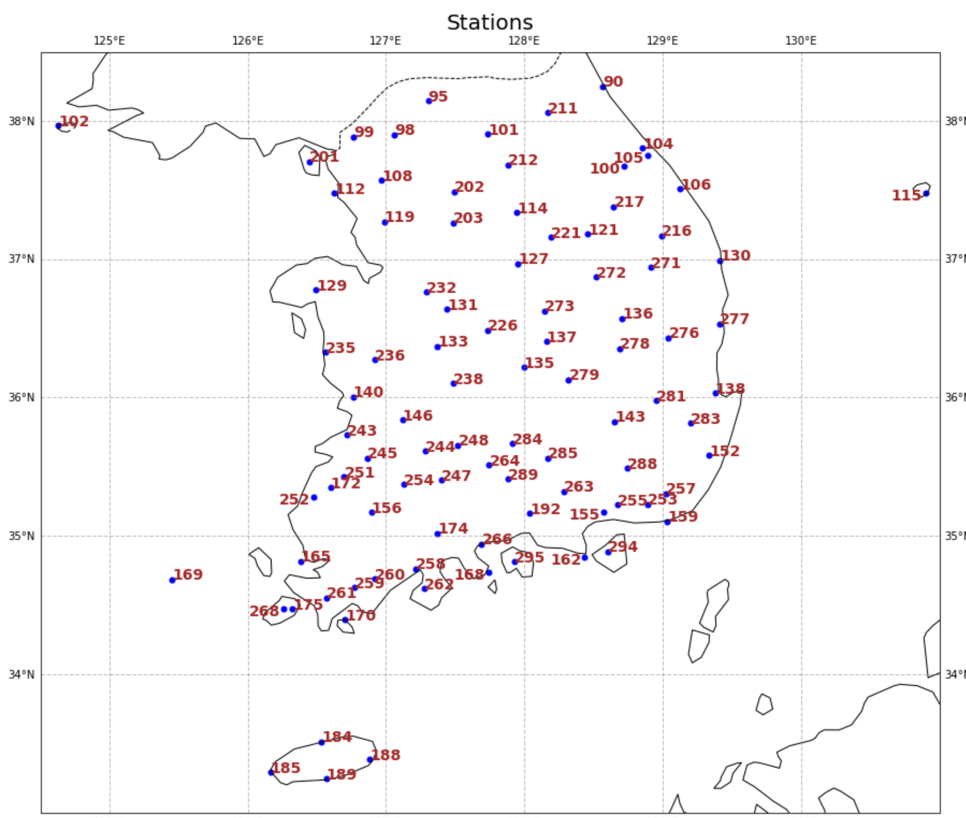


Figure 6.2. *Map of South Korea with the location of the meteorology stations used in the Weather-AI model for bias-correcting WRF weather simulations.*

Upon completion of the WRF run, the closest WRF grid to each station was identified, to which the station was assigned (Table T20 in the Appendix), and then extracted hourly meteorology at each grid point (Table T19 in the Appendix). After acquiring hourly meteorological fields from the output of the WRF model, the inputs were prepared for each station in the form of a two-dimensional matrix in which each column represented a specific meteorology parameter from the WRF model and each row represented hourly values. As each column represented a specific meteorological parameter (Figure F29a, appendix shows the arrangement of inputs and outputs), it displayed a range of values. To establish uniformity over all inputs, each column was normalized between 0 and 1 with a global minimum and maximum (Sayeed et al., 2020b). The output dataset consisted of the hourly observed meteorology. To construct a matrix for training/testing a generalized deep CNN model across the spatial domain, all station data were combined row-wise and further split the training dataset into a 50-50 ratio (randomly) for training and validation. Then, the model was trained for four years (i.e., 2014 to 2017) and evaluated for the year 2018 (Note: The data from 2018 was not used in the model training). A separate model was trained for each of the observed meteorological parameters. One of the major challenges with any ML algorithm was overfitting. To minimize the overfitting, several tests run by varying the number of iterations(epoch) were performed for the model and evaluated the model on the evaluation set in terms of IOA and correlation. In all these sets, the model performed best just before when the validation loss becomes larger than the training loss. So, for the final model, it was trained until this point (i.e., the point before the intersection of train and test curve).

A) Special Case: Precipitation model

Simulating the amount of hourly rainfall for a specific region requires complex physics and chemistry pertaining to atmospheric conditions. Thus, the simulated rainfall was divided into two sections: a classification model (Rain-CM: Rain Classification model) that identified rain hours;

and an hourly quantity prediction model (Rain-RM: Rain Regression Model). The two models are combined to simulate the hourly and daily accumulated total rainfall (in mm).

The Rain-CM model was similar to the model discussed in previous sections but differs in its output, consisting of 0's for no rain and 1's for rain hours. The data setup of the Rain-RM model differed slightly from that of the models discussed in this study (Figure F29b, appendix, shows the data arrangement of the Rain-RM model). The output consisted of observed 24 hourly rain amounts (in mm) arranged in rows, and the inputs consisted of the daily simulated meteorology and simulated 24-hourly rain amounts (in mm) by the WRF (This model has 87 inputs instead of 64 inputs and 24 outputs instead of 1). Therefore, each row in the setup consisted of daily values instead of hourly values.

6.3. Results and Discussion

For the Weather-AI model, the following meteorological parameters were obtained: wind speed, wind direction, temperature, pressure, dewpoint temperature, relative humidity, vapor pressure, and precipitation at the surface. The CNN model was then used with the WRF model to obtain predictions for all of 2018. The CNN models, developed for each meteorological variable, were evaluated against the WRF model performance in the following sections. In addition to using WRF as a benchmark, linear regression and lasso regression models were also used for the evaluation of the performance of wind speed. The windspeed was evaluated as a benchmark as it was a more difficult meteorological parameter to simulate. Both regression models were fitted as a generalized model and a station-specific model (the generalized model used all stations as input). The average performance of the CNN model (generalized) exceeds the performance of the best regression model (Table T21, appendix). However, the station-specific CNN model was better than the generalized model in terms of IOA and correlation. The problem with the station-specific model was that it takes greater computational time to train and it would need 93 different models for each variable for each station (93 stations). Furthermore, a generalized model can be used at any station apart from the 93 used in this study. Thus, to have a generalized and easy-to-use model, the performance of the CNN generalized model for each variable was discussed in further sections.

6.3.1. Wind Speed and Direction:

Figure 6.3 shows the performance of the WRF model (Figure 6.3a) and the Weather-AI model (Figure 6.3b) for each station in terms of IOA. The Weather-AI models show an average increase of 27% in IOA for all stations; IOA increased from 0.67 (correlation = 0.66) for the WRF model to 0.85 (correlation = 0.75) for the Weather-AI model. Overall, the Weather-AI model improved the performance of WRF simulations for all stations, with more than two-thirds (64 out of 93) of the stations showing an IOA increase greater than 20% (Figure F31 shows the percentage change in the IOA at all stations).

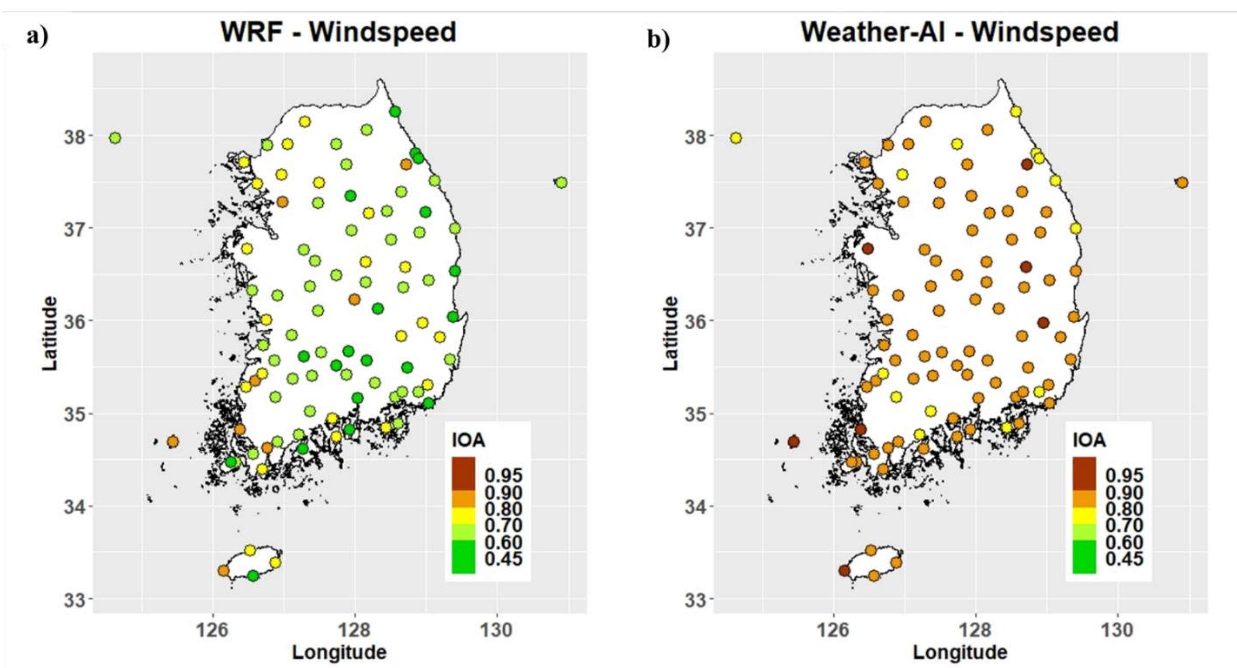


Figure 6.3. *Station-wise IOA comparison of wind speed for a) WRF model and b) Weather-AI models.*

Figure 6.4 shows Taylor diagrams (separated by month) comparing the performance of the two models for all stations combined. The figure shows that the model closest to the observed point on the diagram performs the best (Taylor, 2001), demonstrating the superior performance of the Weather-AI model in all months. Although the root mean squared error (RMSE) for the WRF varied each month and was larger in the cold months, the RMSE for the Weather-AI remained constant at 1 m/s. Similarly, while the standard deviation (SD) and correlation of the WRF varied each month, those of the Weather-AI remained stable throughout the year. From Figure 6.4, one can conclude that seasonality does not affect the performance of the Weather-AI model for wind speed.

Predicting the wind direction was challenging because of its circular nature. To do so, first, predict u and v components of winds were predicted and the direction was calculated. To evaluate the performance of the wind direction, all the predictions that are in the bin of $\pm 45^{o}$ from observed values are treated as true predictions, and all other values are treated as false predictions. Hence, categorical statistic evaluations, in this case, are as follows:

$$HR_{wd}, \text{ Hit Rate} = \frac{\begin{array}{c} No.\ of\ hours\ when\ both \\ predictions\ and\ observations\ are\ in\ the\ range \\ of\ \pm 45^{0}\ from\ observed\ values \end{array}}{Total\ no.of\ hours\ of\ observations} \tag{6.1}$$

$$FAR_{wd}, \text{ False Alarm Rate} = \frac{\begin{array}{c} No.\ of\ hours\ when\ predictions \\ are\ \ not\ in\ the\ range\ of\ \pm 45^{0} \\ from\ observed\ values \end{array}}{Total\ no.of\ hours\ of\ observations} \tag{6.2}$$

The HR_{wd} for all stations combined for the Weather-AI was 54.83% and HR_{wd} for the WRF was 52.16%.

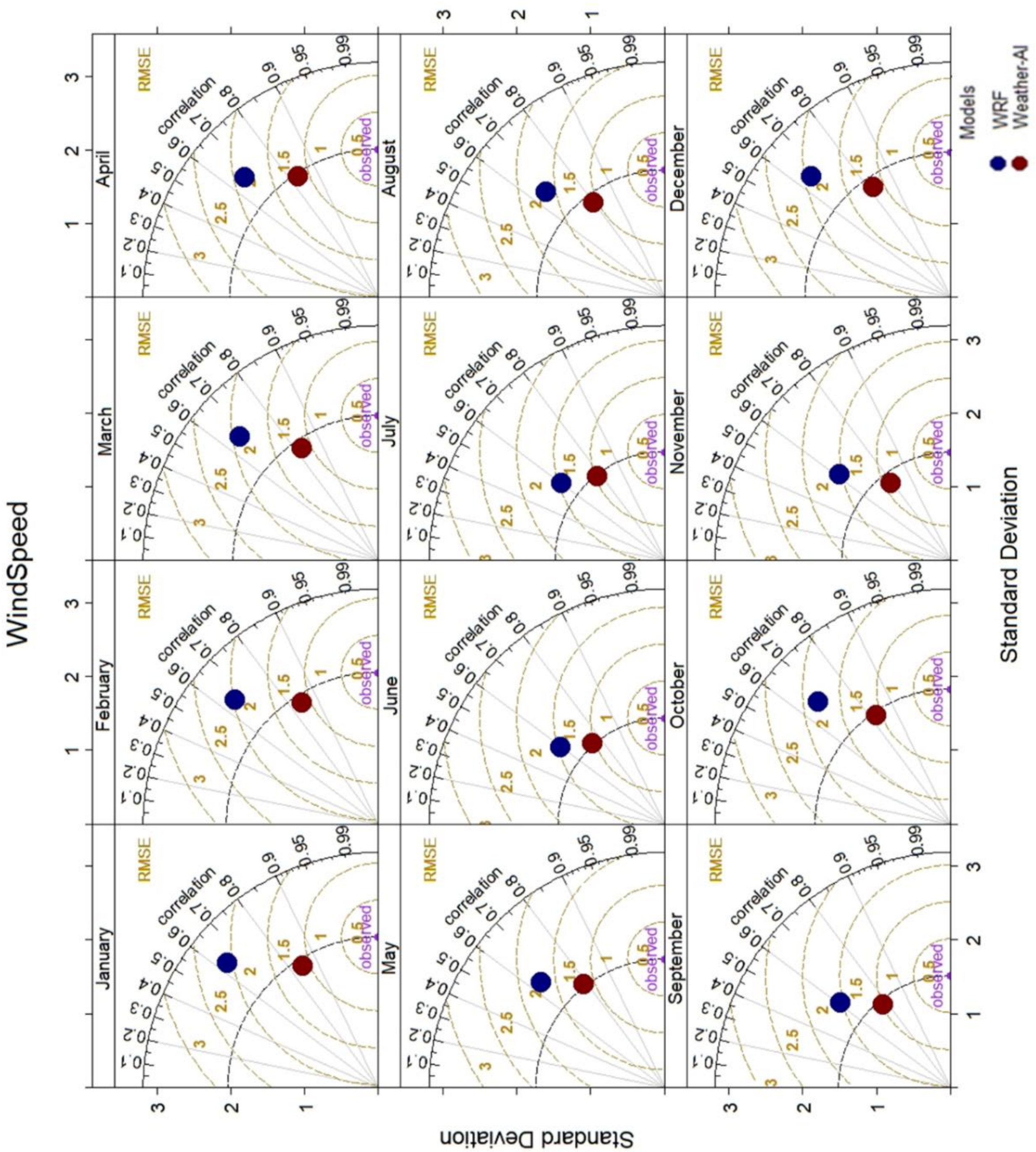


Figure 6.4. *Taylor diagram of each month comparing the performance of the WRF model and Weather-AI model for wind speed.*

Figure F32a shows the yearly time series of wind speed and Figure F32b shows the wind direction at station 115. This station was unique because it was situated near the southeastern coast of a small island, Ulleng-do (120 km east of the Korean Peninsula). The WRF model significantly overpredicted wind speeds during the cold months (Figure F32a). As summer approached, its performance improved (also shown in Figure 6.4), with the most dramatic improvement in the JJA season. The Weather-AI model was able to reduce the seasonal biases of the WRF, out-performing it in all months for predicting wind speed and more accurately predicting the wind direction (Figure F32b). Furthermore, the model significantly improved the wind direction predictions by successfully predicting dominant southwestern and northeastern wind directions.

6.3.2. Precipitation

The bias correction of precipitation consisted of two models. Therefore, different techniques were used to evaluate them. Rain-CM was evaluated based on categorical statistics, that is, the hit rate (HR) and the false alarm rate (FAR), defined as follows:

$$HR_{rain}\text{, HR Rain Condition} = \frac{\substack{No.\ of\ hours\ when \\ both\ prediction\ and \\ observation\ are\ a\ rain\ hour}}{\substack{Total\ no.of\ hours\ when \\ observation\ is\ \ a\ rain\ hour}} \tag{6.3}$$

$$FAR_{rain}\text{, FAR Rain Condition} = \frac{\substack{No.\ of\ hours\ when\ \ prediction \\ is\ a\ no\ rain\ \ and\ o \\ bservation\ is\ a\ rain\ hour}}{\substack{Total\ no.of\ hours\ when \\ observation\ is\ a\ rain\ hour}} \tag{6.4}$$

$$HR_{no\text{-}rain}\text{, HR No-Rain Condition} = \frac{\substack{No.\ of\ hours\ when\ both \\ prediction\ and\ observation \\ are\ no\ rain\ hour}}{\substack{Total\ no.of\ hours\ when \\ observation\ is\ a\ no\ rain\ hour}} \tag{6.5}$$

$$FAR_{no\text{-}rain}\text{, FAR No-Rain Cond.} = \frac{\substack{No.\ of\ hours\ when \\ prediction\ is\ rain \\ observation\ is\ no\ rain\ hour}}{\substack{Total\ no.of\ hours\ where \\ observation\ is\ a\ no\ rain\ hour}} \tag{6.6}$$

Tables 6.1a and 6.1b show the HR and FAR of the WRF and Weather-AI models, respectively, for the year 2019 for all stations combined (observations with “NaN” values were removed). The Weather-AI Rain-CM model showed 7% and 1% improvement over the WRF model in the HR for rain and no-rain hours, respectively, and 37.5% and 6.25% decrease in the FAR for rain and no-rain, respectively.

Table 6.1a. *Categorical evaluation of the rain classification for the WRF model.*

	Observed Rain Hours		**Observed No Rain Hours**	
Predicted Rain	44058	HR_{rain} = 0.84	122670	$FAR_{no\text{-}ain}$ = 0.16
Predicted No Rain	8256	FAR_{rain} = 0.16	638654	$HR_{no\text{-}rain}$ = 0.84
Total Hours	52314		761324	

Table 6.1b. *Categorical evaluation of the rain classification for the Weather-AI (Rain-CM) model.*

	Observed Rain Hours		**Observed No Rain Hours**	
Predicted Rain	47214	HR_{rain} = 0.90	117572	$FAR_{no\text{-}ain}$ = 0.15
Predicted No Rain	5100	FAR_{rain} = 0.10	643752	$HR_{no\text{-}rain}$ = 0.85
Total Hours	52314		761324	

After obtaining the predictions from the classification model, the regression model (Rain-RM) was used to predict the hourly amount of precipitation. To merge both models and predict rain more accurately, all the non-rain hours from the Rain-CM were converted to zero. The average IOA for all stations for hourly rain was 0.62 (WRF = 0.56) and the correlation was 0.51 (WRF = 0.43). According to Figure 6.5, which presents a station-wise IOA comparison for hourly rain, 90% of the stations show an improved IOA, and 95% show an improved correlation for hourly rain.

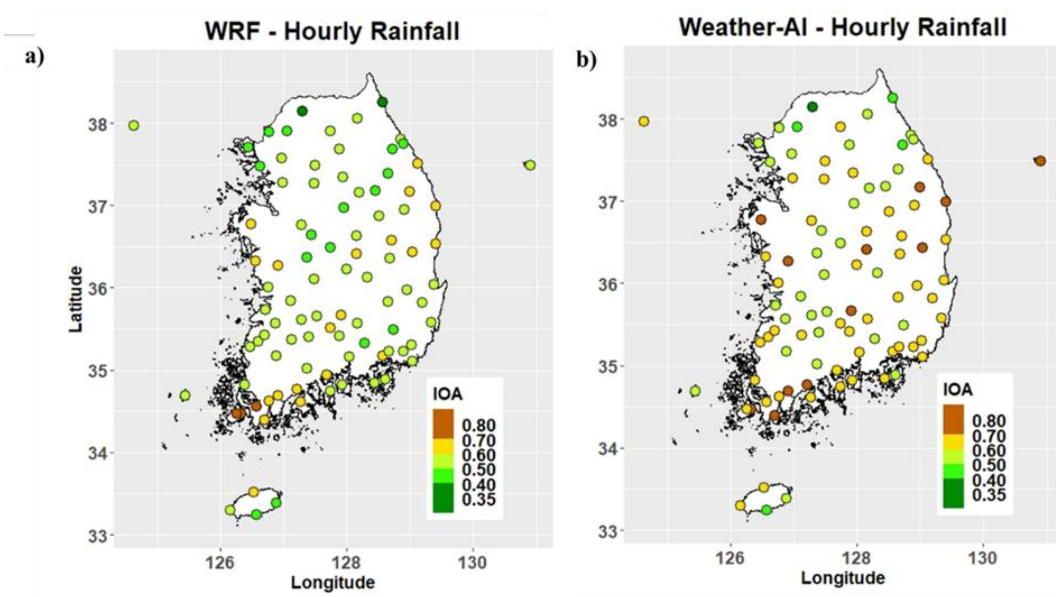


Figure 6.5. *Station-wise IOA comparison of hourly rainfall for a) the WRF model and b) the Weather-AI model.*

The next step in rainfall prediction was daily accumulated rainfall, calculated from the hourly rain predicted by the Rain-RM model. Figure 6.6 represents a station-wise IOA comparison of the WRF and Weather-AI models. The average IOA and correlation of the Weather-AI model were 0.87 (WRF-0.86) and 0.79 (WRF-0.77), respectively.

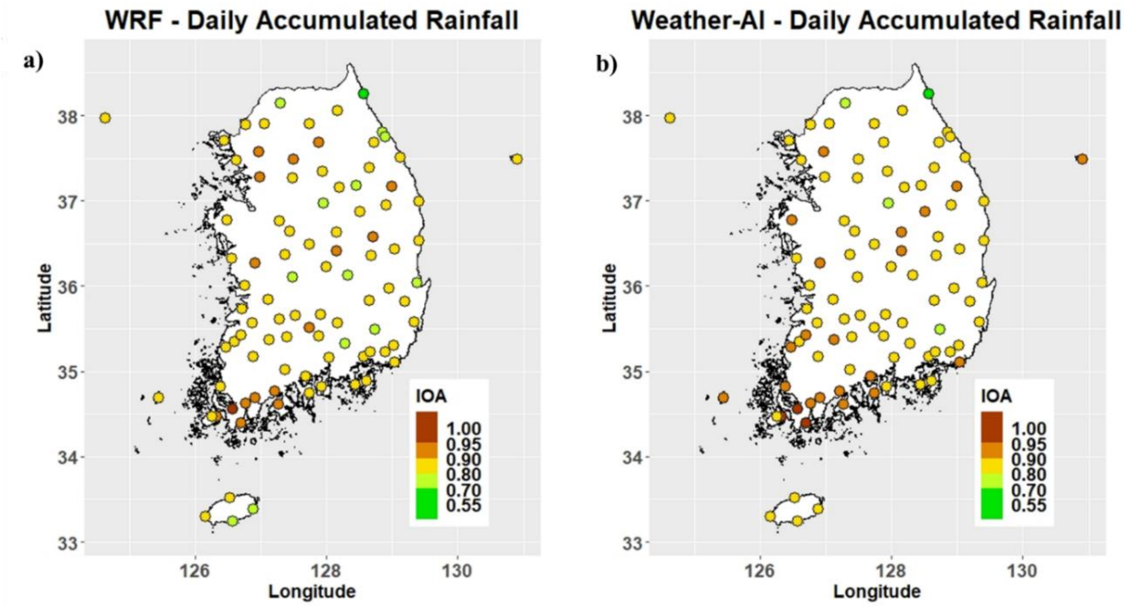
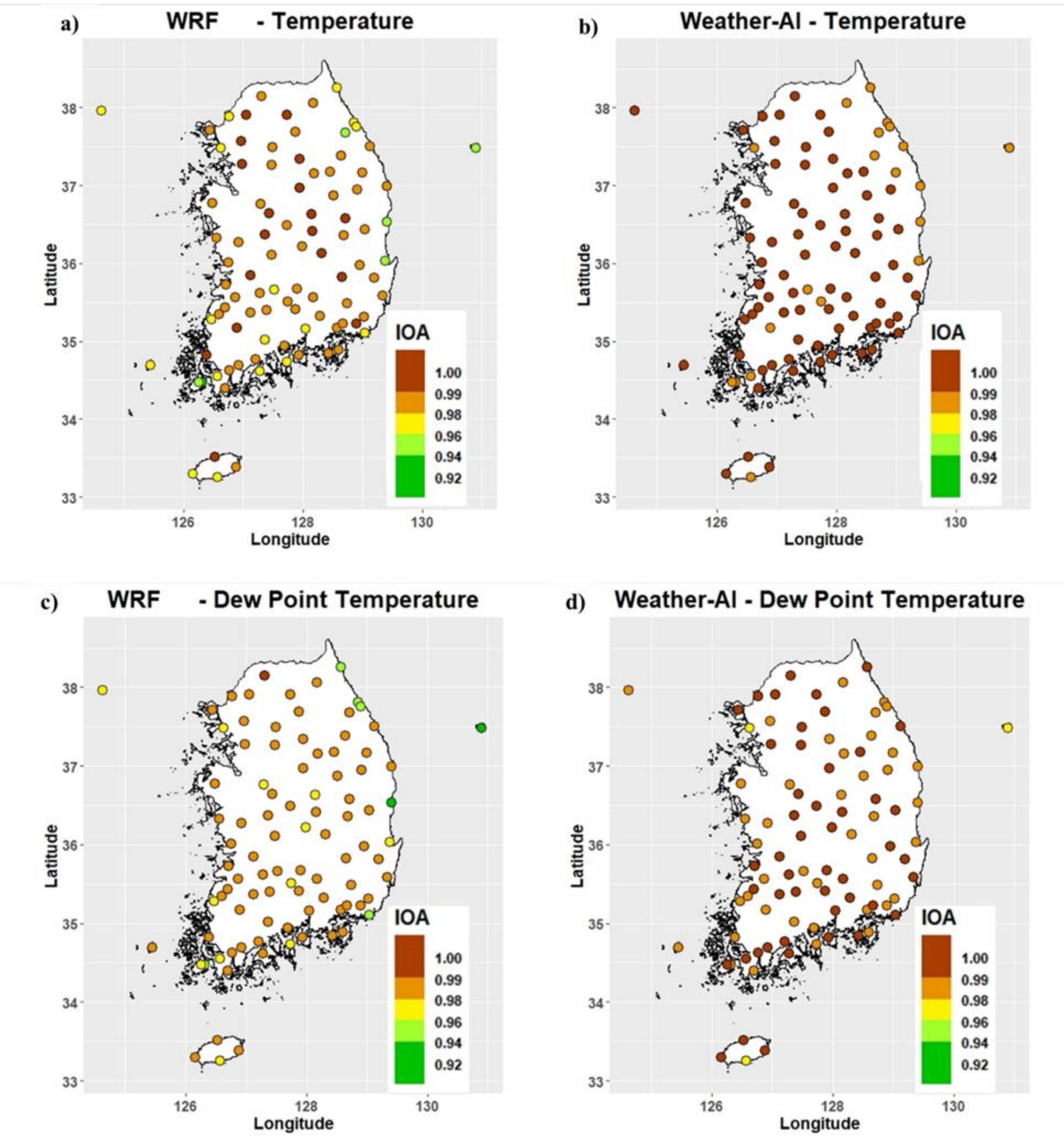


Figure 6.6. *Station-wise IOA comparison of daily accumulated rainfall by a) the WRF model and b) Weather-AI models.*

6.3.3. Other Weather Variables:

Figure 6.7a and 6.7b present the station-wise IOA of hourly temperature for 24 hours predictions by the WRF and Weather-AI models, respectively. Both models performed well in predicting temperature, with an average IOA for all stations combined of 0.98 from the WRF and 0.99 from the Weather-AI models. The range of the IOA for the WRF was 0.92-0.99 and for the Weather-AI 0.98-0.99. Even though the temperature predictions of the WRF were exceptionally accurate, those of the Weather-AI still showed improvements in all stations. A similar improvement occurred for the dewpoint temperature (Figure 6.7c and 6.7d). A monthly Taylor diagram comparison of both models for temperature and dewpoint temperature are shown in Figure F33a and F33b. Results have shown the RMSE and the SD from WRF were slightly larger during the DJF (December, January, and February) season with a weaker correlation. Whereas during the warmer months, WRF had smaller RMSE and SD with a higher correlation. In contrast, the Weather-AI generated more accurate predictions than the WRF for all months. The RMSE and SD did not vary or exceed 2°C for each month throughout the season.



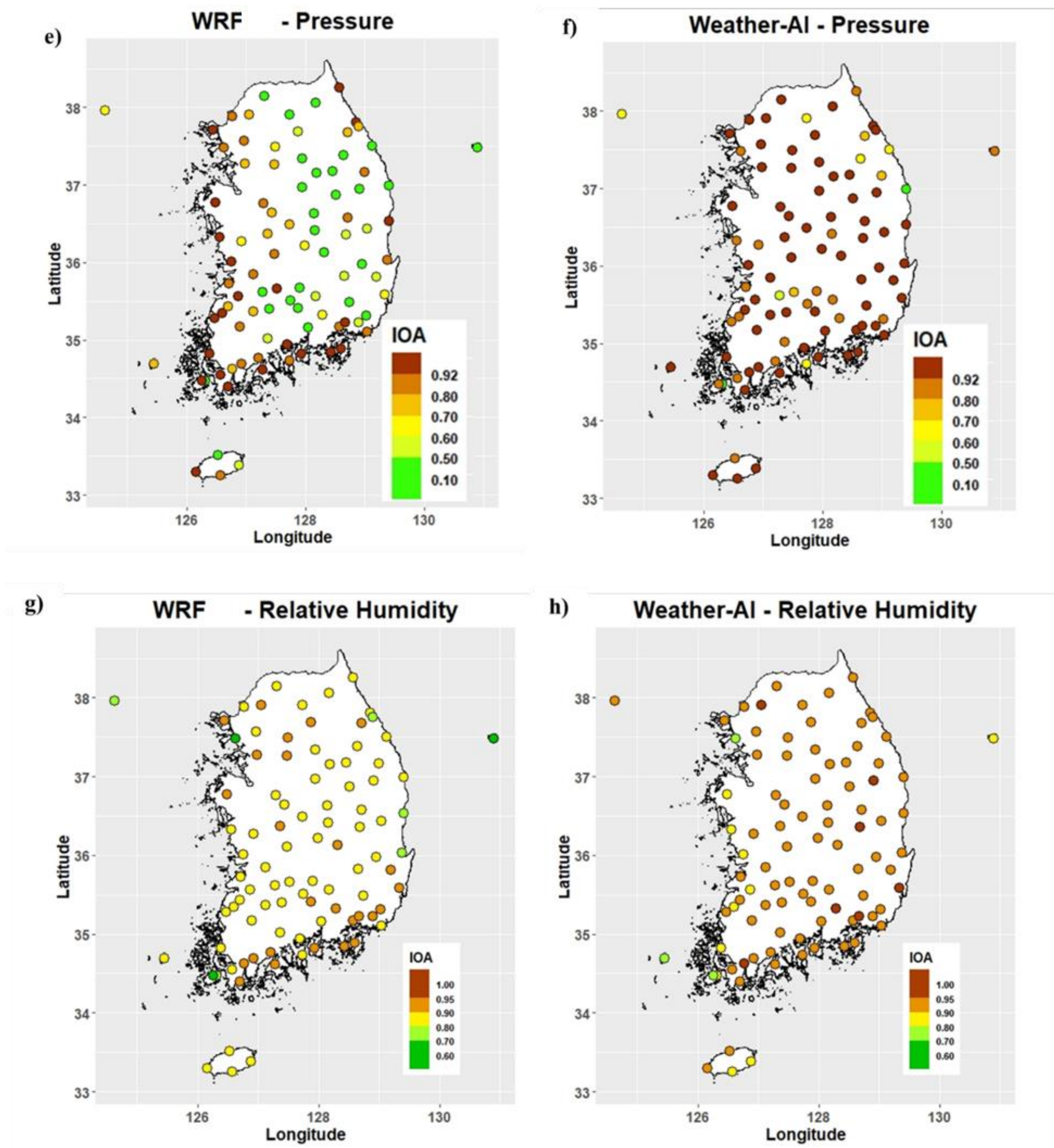


Figure 6.7. *Station-wise IOA comparison of the forecasts of the WRF and Weather-AI models for temperature, dew-point temperature, surface pressure, and relative humidity. a), c), e), and g) represents IOA for temperature, dewpoint temperature, pressure, and relative humidity, respectively, for the WRF model. b), d), f), and h) represents IOA for temperature, dewpoint temperature, pressure, and relative humidity, respectively, for the Weather-AI model.*

The IOA for the hourly surface pressure predictions for 24 hours increased significantly, as shown in Figures 6.7e and 6.7f. The average IOA of the WRF and Weather-AI models were 0.69 and 0.91, respectively. For several stations, the WRF produced uniform bias in simulating surface pressure, which was adjusted by the Weather-AI (Figure F34 in Appendix). Since the bias from the WRF was uniform, the correlation was stronger for these stations, but the IOA was weaker. However, as the bias from the Weather-AI decreased, the IOA increased.

Figure 6.7g and 6.7h show the yearly IOA of the hourly simulations of relative humidity from the WRF (IOA-0.87) and Weather-AI (IOA-0.92) models, respectively. All, except for five (Station 169, 165, 129, 140, and 170), stations show improvement in the IOA. According to Figure F33c, the Weather-AI model performed better than the WRF model for relative humidity. Also, the bias-corrections by the Weather–AI model were slightly more accurate than the simulations of the WRF model in all months.

6.4. Conclusions

In this study, a deep CNN model was developed and discussed that reduced bias in an NWP model and significantly improved predictions. Although the same model configuration was used, several meteorology-specific models based on the target/output were developed. The models showed improved predictions over the WRF model and significantly reduced bias.

The IOA for wind speeds from the Weather-AI model improved for all 93 stations in South Korea. Improvement fell within the range of 2.3 – 39.3%, with a mean of 17.83% in absolute terms. For wind direction, the predictions of the Weather-AI model improved in 52 out of 93 stations. Moreover, the performance remained consistent throughout the year. Since the Weather-AI model uses the WRF meteorology as an input, it does not indicate any time-series shift as the input doesn't have true observations (Figure F35, Appendix, shows the best, the median, and the least performing station based on IOA). The Rain-CM improved the hit rate by 6% over the WRF model for the prediction of rain hours, but it remained the same as the WRF for the prediction of no-rain hours. The bias correction of the hourly rainfall amount by the Rain-RM model improved in most of the stations; nevertheless, simulating the absolute amount of hourly rainfall remains a challenge.

Predictions of the daily accumulated rainfall amount showed a slight improvement in the IOA and a 2% improvement in the correlation. The performance statistics of other meteorological parameters— temperature, dewpoint temperature, and relative humidity— also improved.

The Weather-AI model significantly improved and bias-corrected the simulated wind speed, relative humidity, and hourly precipitation by the WRF model. As WRF predictions were already relatively accurate, they did not show significant improvement in the bias-correction of temperature and dewpoint temperature. The correlation and IOA for temperature were in the range of 0.92-0.99 for the WRF model. However, for the CNN, the range of correlation and IOA was 0.97-0.99 and 0.98-0.99, respectively. Similarly, for dew point temperature, the IOA and correlation for the WRF model were in the range of 0.93-0.99 and 0.95-0.98, respectively. For the CNN model, the range of IOA and correlation were 0.96-0.99 and 0.97-0.99, respectively. The simulation of surface pressure from WRF contained a uniform bias in several stations that were corrected by the Weather-AI model. Even though the Weather-AI model was trained for South Korea WRF simulation, a similar model can be trained and reproduced for any numerical model (simulations and forecasts). In addition, the system can be utilized for any location to bias-correct any number of meteorological parameters while being computationally fast. Although the AI model showed significant improvement over the WRF model, it does not cover WRF domains over the sea/ocean (because of the lack of observations). In addition, unlike the WRF and more advanced architectures of CNN, the Weather-AI model has no spatial gridded structure. Therefore, developed AI models are capable of spatial and temporal simulation and forecasting, specifically long-range forecasting, based on Weather-AI.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1. Conclusion

In this study, several deep learning algorithms, particularly CNN, were used to develop computational fast and accurate techniques to bias-correct numerical models and forecast air quality and meteorology with better accuracy. The models developed were also able to forecasts the air-quality parameters like surface ozone, PM_{10}, $PM_{2.5,}$ and NO_2 for a longer duration (up to two weeks in advance). The meteorology model, Weather-AI, was able to achieve better accuracy than the conventional numerical model (WRF).

In the first task, a deep CNN model was developed to forecast hourly ozone concentration for 24-hours in advance at various monitoring stations in Texas, USA. For this model, previous day (24- hour, hourly) values of observed meteorology and atmospheric constituents like ozone and NO_2 were used. The model was trained on examples from 2014 to 2016 and evaluated for the year 2017. The robustness of the model was evaluated on discrete as well categorical parameters. For 19 of the 21 stations in the study, results show that the yearly index of agreement (IOA) was above 0.85, confirming the acceptable accuracy of the CNN model. The results also showed that the model performed well, even for stations with varying monthly trends of ozone concentrations (specifically CAMS-012, located in El-Paso, and CAMS-013, located in Fort Worth, both with IOA=0.89). In addition, to ensure that the model was robust, we tested it on stations where fewer meteorological variables were monitored. Although these stations have fewer input features, their performance was similar to that of other stations. However, despite its success at capturing daily trends, the model mostly underpredicts the daily maximum ozone, which provides a direction for future study and improvement. As this model predicts ozone concentrations 24-h in advance with greater accuracy and computationally fewer resources, it can serve as an early warning system for individuals susceptible to ozone and those engaging in outdoor activities.

Learning from the previous task, a deep CNN architecture was designed and developed to forecast hourly ozone concentration two weeks in advance for the second task. In this task, we developed a modeling system based on a convolutional neural network (CNN) model that is not only fast but covers a temporal period of two weeks with a resolution as small as a single hour for 255 stations. The CNN model uses meteorology from the Weather Research and Forecasting model (processed by the Meteorology-Chemistry Interface Processor), forecasted air quality from the Community Multi-scale Air Quality Model (CMAQ), and previous 24-hour concentrations of various measurable air quality parameters as inputs and predicted the following 14-day hourly surface ozone concentrations. The model achieves an average accuracy of 0.91 in terms of the index of agreement for the first day and 0.78 for the fourteenth day, while the average index of agreement for one day ahead prediction from the CMAQ is 0.77. Through this task, the best features of numerical modeling (i.e., fine spatial resolution) and a deep neural network (i.e., computation speed and accuracy) were amalgamated to achieve more accurate spatio-temporal predictions of hourly ozone concentrations. Although the primary purpose of this study is the prediction of hourly ozone concentrations, the system was extended to various other pollutants in the third task.

In the third task, the algorithm developed in the second task was extended to forecast other atmospheric constituents like $PM_{2.5}$, PM_{10}, and NO_2 for up to two weeks, but it was found that the forecast was only reliable for up to one week only. The model was also used to evaluate the stations that were not in training for robustness. The CNN model developed in this task, bias-corrects hourly concentrations of air pollutants from the CMAQ model on the first day and forecasts the remaining six days. Our results show improved performance of the average yearly index of agreement (IOA) from the CMAQ to the CNN model by 13% for $PM_{2.5}$, 22% for PM_{10}, and 43% for NO_2 for the first-day bias correction; and the seventh-day forecast of NO2 by the CNN model was more accurate than the first-day forecast of the CMAQ model. The forecasts for $PM_{2.5}$ and PM_{10}, however, were reliable only up to two days in advance. The trained model was also capable of forecasting pollutants at stations not included in the training and showed similar performance metrics as that of the stations included in the training. The increase in the average yearly IOA at such stations was 13% for $PM_{2.5}$, 22% for PM_{10}, and 40% for NO_2. Although the CNN model enhances the performance of the CMAQ model, it can be further improved by adding remote sensing data.

In the fourth task, various deep CNN models specific to specific meteorology (like temperature, pressure, relative humidity, etc.) were developed. The use of a computationally efficient deep learning method, the Convolutional Neural Network (CNN), as a post-processing technique that improves mesoscale Weather and Research Forecasting (WRF) one-day simulation (with a one-hour temporal resolution) outputs were investigated. Using the CNN architecture, several meteorological parameters calculated by the WRF model for 2018 were bias-corrected. The CNN model was trained with a four-year history (2014-2017) to investigate the patterns in WRF biases and then reduce these biases in simulations for surface wind speed and direction, precipitation, relative humidity, surface pressure, dewpoint temperature, and surface temperature. The WRF data, with a spatial resolution of 27 km, covers South Korea. The surface measurement was obtained from the Korean Meteorological Administration station network for 93 weather station locations. The results indicate a noticeable improvement in WRF simulations in all station locations. The average annual index of agreement for surface wind, precipitation, surface pressure, temperature, dewpoint temperature, and relative humidity of all stations were 0.85 (WRF:0.67), 0.62 (WRF:0.56), 0.91 (WRF:0.69), 0.99 (WRF:0.98), 0.98 (WRF:0.98), and 0.92 (WRF:0.87), respectively. While this study focuses on South Korea, the proposed approach can be applied for any measured weather parameters at any location.

The models developed in this study were evaluated both spatially and temporally for a very complex region. These evaluations have shown that the model is robust enough to estimate various parameters of meteorology and air-quality. The dissertation successfully developed technology that integrated both numerical models and the deep neural networks techniques. With this integration a better and longer forecast (up to two weeks) were made possible for difficult to estimate quantities like ozone, NO_2, $PM_{2.5}$ and PM_{10}. The model was also able to improve the performance of estimation of various meteorology simulated by WRF model. This methodology can also be used to estimated other measures/monitored quantities. Although the region for this study was South Korea, this study can be used for any region. This was also shown by the spatial evaluations in chapter 4.

7.2. Future Work

Although the deep learning models developed performed exceedingly well for various scenarios, they can be further improved with the more advanced ML algorithms like 2D CNN, reinforcement learning, unsupervised learning, etc. Also, the algorithms developed in this study focus mainly on post-processing of simulation results and extending the forecast through ML techniques; the future researcher can develop a pre-processing technique, wherein ML can take care of the parametrization of partial differential equations. The other approaches can also be using the better forecast of meteorology as developed in the fourth task and using it in the CMAQ model for better physics.

Another approach to having a better weather forecast can be the use of ML for solving adjoints for data assimilations. The traditional way of reconstructing the space-time variations of weather events from observations relies on data assimilation (DA) methods involving a known dynamical model, also referred to as a numerical weather prediction (NWP) model. These data assimilation techniques require estimation of a cost function and its minimization, typically through the use of an adjoint-based variational approach, an ensemble-based approach, or a combination of both. However, adjoint versions of the forecast model are non-trivial to develop and can be computationally expensive to run, especially when they involve a linearization (gradient) of highly nonlinear 'physics' components of the NWP model, used for representing the effects clouds, radiation, and turbulence. An alternative approach can be explored for estimating the adjoint operators used in data assimilation, based on a fully data-driven framework that combines machine learning and NWP to develop a computationally fast deep neural network (DNN).

APPENDIX

A: Tables

Table T1. *Sites in Texas: Location of all stations with available (monitored) meteorological and pollutants parameter (WS=Wind Speed, WD =Wind Direction, TEMP=temperature, DPT=Dew point Temperature, RH= Relative Humidity, SOL-RAD = Solar Radiation, PPT=precipitation, NET-RAD = net radiation, PR= Pressure)*

Station ID	Area	Lat.	Lon.	Meteorology	Pollutant
CAMS003	ARR	30.354	-97.76	WS/WD/TEMP	NOx/ O_3
CAMS008	HGB	29.901	-95.326	WS/WD/TEMP/DPT/RH/SOL-RAD/PR	NOx/ O_3
CAMS012	ELP	31.768	-106.5	WS/WD/TEMP/RH/SOL-RAD/PPT	NOx/ O_3
CAMS013	DFW	32.806	-97.356	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3
CAMS015	HGB	29.802	-95.126	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3
CAMS019	TLM	32.379	-94.712	WS/WD/TEMP/SOL-RAD/PPT	NOx/ O_3
CAMS026	HGB	30.039	-95.674	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3
CAMS035	HGB	29.67	-95.128	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3
CAMS045	HGB	29.583	-95.015	WS/WD/TEMP/SOL-RAD	NOx/ O_3
CAMS053	HGB	29.696	-95.499	WS/WD/TEMP/SOL-RAD	NOx/ O_3
CAMS059	SAN	29.275	-98.311	WS/WD/TEMP	NOx/ O_3
CAMS078	HGB	30.35	-95.425	WS/WD/TEMP/SOL-RAD	NOx/ O_3
CAMS401	DFW	32.82	-96.86	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3
CAMS403	HGB	29.734	-95.257	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3
CAMS416	HGB	29.686	-95.295	WS/WD/TEMP/RH/SOL-RAD/ PPT	NOx/ O_3
CAMS617	HGB	29.821	-94.99	WS/WD/TEMP/NET-RAD	NOx/ O_3
CAMS618	HGB	29.149	-95.765	WS/WD/TEMP/NET-RAD	NOx/ O_3
CAMS620	HGB	29.402	-94.946	WS/WD/TEMP/NET-RAD	NOx/ O_3
CAMS695	HGB	29.718	-95.341	WS/WD/TEMP/RH/PR/PPT	O_3
CAMS1034	HGB	29.254	-94.861	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3
CAMS1035	BPA	29.979	-94.011	WS/WD/TEMP/DPT/RH/SOL-RAD	NOx/ O_3

Austin-Round Rock (ARR), Houston-Galveston-Brazoria (HGB), El Paso-Juarez (ELP). Dallas-Fort Worth (DFW), Tyler-Longview-Marshal (TLM), San Antonio (SAN), Beaumont-Port Arthur (BPA)

Table T2. *Month-wise Index of Agreement for all stations.*

Station ID	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Overall
CAMS-003	0.85	0.82	0.86	0.85	0.86	0.93	0.93	0.90	0.94	0.82	0.84	0.77	0.89
CAMS-008	0.80	0.85	0.87	0.91	0.91	0.92	0.92	0.89	0.92	0.89	0.89	0.84	0.90
CAMS-012	0.78	0.82	0.83	0.84	0.83	0.87	0.87	0.87	0.88	0.83	0.80	0.83	0.89
CAMS-013	0.85	0.79	0.83	0.84	0.91	0.88	0.88	0.89	0.92	0.88	0.83	0.84	0.89
CAMS-015	0.75	0.72	0.79	0.78	0.85	0.88	0.89	0.88	0.91	0.84	0.83	0.78	0.85
CAMS-019	0.84	0.77	0.84	0.84	0.84	0.88	0.91	0.89	0.94	0.88	0.86	0.86	0.88
CAMS-026	0.84	0.85	0.88	0.86	0.87	0.93	0.92	0.90	0.92	0.89	0.89	0.85	0.90
CAMS-035	0.79	0.83	0.83	0.84	0.88	0.90	0.86	0.90	0.91	0.89	0.86	0.80	0.88
CAMS-045	0.75	0.81	0.79	0.77	0.84	0.87	0.84	0.84	0.85	0.82	0.80	0.76	0.84
CAMS-053	0.81	0.84	0.83	0.86	0.86	0.89	0.90	0.88	0.90	0.87	0.88	0.82	0.88
CAMS-059	0.85	0.86	0.84	0.86	0.87	0.92	0.94	0.86	0.93	0.89	0.90	0.87	0.90
CAMS-078	0.86	0.86	0.86	0.89	0.90	0.92	0.94	0.91	0.93	0.89	0.88	0.84	0.91
CAMS-401	0.78	0.80	0.83	0.81	0.89	0.88	0.88	0.87	0.90	0.87	0.85	0.81	0.88
CAMS-403	0.79	0.82	0.78	0.82	0.86	0.87	0.87	0.86	0.87	0.84	0.86	0.78	0.86
CAMS-416	0.77	0.84	0.82	0.87	0.86	0.89	0.87	0.76	0.90	0.89	0.88	0.80	0.87
CAMS-617	0.81	0.87	0.84	0.84	0.87	0.89	0.91	0.86	0.94	0.89	0.89	0.79	0.89
CAMS-618	0.84	0.83	0.82	0.84	0.85	0.92	0.88	0.90	0.94	0.91	0.90	0.84	0.90
CAMS-620	0.77	0.80	0.79	0.77	0.81	0.85	0.81	0.84	0.86	0.74	0.80	0.78	0.85
CAMS-695	0.78	0.83	0.82	0.84	0.85	0.87	0.89	0.83	0.91	0.86	0.85	0.75	0.87
CAMS-1034	0.81	0.82	0.80	0.77	0.79	0.89	0.81	0.88	0.87	0.76	0.82	0.73	0.85
CAMS-1035	0.80	0.85	0.78	0.83	0.86	0.89	0.90	0.93	0.91	0.88	0.87	0.80	0.88

Table T3. *Month Wise Correlation for all stations.*

Station ID	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Overall
CAMS-003	0.76	0.72	0.77	0.76	0.77	0.87	0.87	0.82	0.89	0.68	0.73	0.62	0.81
CAMS-008	0.66	0.75	0.79	0.84	0.84	0.85	0.86	0.80	0.88	0.81	0.83	0.74	0.83
CAMS-012	0.64	0.71	0.75	0.73	0.70	0.81	0.78	0.79	0.81	0.72	0.68	0.73	0.82
CAMS-013	0.73	0.64	0.71	0.73	0.86	0.79	0.78	0.81	0.87	0.79	0.72	0.74	0.81
CAMS-015	0.62	0.55	0.66	0.64	0.76	0.79	0.81	0.78	0.87	0.75	0.72	0.65	0.76
CAMS-019	0.74	0.64	0.74	0.74	0.75	0.79	0.85	0.80	0.90	0.80	0.76	0.77	0.80
CAMS-026	0.72	0.74	0.80	0.78	0.78	0.87	0.85	0.82	0.86	0.82	0.81	0.75	0.83
CAMS-035	0.66	0.72	0.75	0.74	0.81	0.84	0.78	0.82	0.87	0.81	0.77	0.67	0.81
CAMS-045	0.61	0.70	0.73	0.64	0.74	0.79	0.74	0.73	0.76	0.72	0.68	0.60	0.75
CAMS-053	0.69	0.73	0.75	0.78	0.77	0.81	0.83	0.79	0.85	0.80	0.82	0.72	0.81
CAMS-059	0.75	0.77	0.75	0.79	0.79	0.86	0.89	0.76	0.88	0.80	0.84	0.79	0.83
CAMS-078	0.76	0.76	0.78	0.81	0.83	0.85	0.89	0.84	0.88	0.81	0.81	0.75	0.83
CAMS-401	0.63	0.68	0.72	0.71	0.82	0.79	0.80	0.78	0.82	0.76	0.76	0.68	0.79
CAMS-403	0.66	0.72	0.69	0.70	0.76	0.78	0.79	0.77	0.82	0.76	0.78	0.63	0.78
CAMS-416	0.68	0.78	0.77	0.78	0.78	0.80	0.81	0.81	0.84	0.84	0.80	0.70	0.81
CAMS-617	0.67	0.77	0.73	0.74	0.77	0.81	0.84	0.76	0.89	0.80	0.81	0.66	0.80
CAMS-618	0.72	0.72	0.75	0.76	0.74	0.85	0.79	0.82	0.89	0.83	0.83	0.73	0.82
CAMS-620	0.63	0.69	0.72	0.64	0.68	0.77	0.69	0.72	0.78	0.59	0.68	0.63	0.75
CAMS-695	0.63	0.71	0.73	0.73	0.74	0.78	0.81	0.71	0.85	0.77	0.74	0.59	0.77
CAMS-1034	0.69	0.72	0.75	0.63	0.68	0.84	0.69	0.79	0.79	0.61	0.69	0.55	0.76
CAMS-1035	0.65	0.76	0.66	0.72	0.77	0.82	0.83	0.87	0.87	0.79	0.79	0.67	0.81

Table T4. *Categorical Statistics results from Chai et al. (2013) compared with the CNN model's categorical statistics.*

	a	b	c	d	N_r	HIT	CSI	FAR	POC	ETS
Pacific Coast	2047	977	47796	681	97.35	0.59	0.26	0.68	0.95	0.24
Lower Middle	1156	168	44875	217	10.98	0.44	0.11	0.87	0.97	0.10
Southeast	2774	345	54107	132	25.94	0.72	0.11	0.89	0.95	0.10
Rocky Mountain	939	70	42667	88	3.64	0.44	0.06	0.93	0.98	0.06
Upper Middle	1963	281	52056	171	18.62	0.62	0.12	0.87	0.96	0.11
North East	2232	770	42637	157	60.77	0.83	0.24	0.74	0.95	0.23
CONUS	11119	2616	284706	1449	186.18	0.64	0.17	0.81	0.96	0.16
Overall	22230	5227	568844	2895	372.17	0.64	0.17	0.81	0.96	0.16

Table T5. *Model comparisons of RF+CNN and CNN.*

			Percent change for RF+CNN from CNN only model		
	Inputs*	Total Inputs#	Time^	IOA¥	Correlationα
CAMS003	72	120	19.96	0.58	0.72
CAMS008	120	216	15.17	0.29	0.22
CAMS012	144	192	19.65	0.60	0.44
CAMS013	48	192	-22.91	0.11	0.63
CAMS015	168	192	35.13	0.39	0.16
CAMS019	96	168	-1.41	1.76	1.28
CAMS026	48	192	-24.22	0.06	-0.08
CAMS035	72	192	-11.91	-0.08	0.41
CAMS045	120	144	29.63	-1.35	-0.69
CAMS053	96	144	14.41	0.34	-0.22
CAMS059	48	120	-6.98	0.19	0.12
CAMS078	48	144	-19.34	0.16	0.84
CAMS1034	48	192	-21.95	-0.29	0.36
CAMS1035	72	192	-9.77	0.51	0.07
CAMS401	48	192	-22.62	1.77	1.48
CAMS403	72	192	-4.46	1.69	1.06
CAMS416	48	192	-28.02	0.26	0.80
CAMS617	72	144	0.48	0.49	0.34
CAMS618	48	144	-15.42	0.98	0.92
CAMS620	72	144	4.03	0.98	1.14
CAMS695	48	168	-21.72	0.17	0.69

*Top features selected by Random Forests
Total Number of available input features
^Percent change in computational time. Negative means improvement.
¥Percent change in IOA. Positive means improvement.
αPercent change in Correlation. Positive means improvement.

Table T6: *List of parameters from WRF/MCIP and CMAQ used to train the CNN model.*

Abb.	Variable Name (WRF/MCIP)	Units
PRSFC	Surface Pressure	Pascal
USTAR	Cell Averaged Friction Velocity	m/s
WSTAR	Convective Velocity Scale	m/s
PBL	Planetary Boundary Level Height	M
MOLI	Inverse Of Monin-Onukhov Length	1/m
HFX	Sensible Heat Flux	watt/m^2
RADYNI	Inverse Of Aerodynamic Resistance	m/s
RSTOMI	Inverse Of Bulk Stomatal Resistance	m/s
TEMPG	Skin Temperature At Ground	Kelvin
TEMP2	Temperature At 2 M	Kelvin
Q2	Mixing Ratio At 2 M	Kg/Kg
WSPD10	Wind Speed At 10 M	m/s
WDIR10	Wind Direction At 10 M	Degrees
GLW	Longwave Radiation At Ground	watt/m^2
GSW	Solar Radiation Absorbed At Ground	watt/m^2
RGRND	Solar Rad Reaching Sfc	watt/m^3
RN	Nonconvec. Pcpn Per Met Tstep	cm
RC	Convective Pcpn Per Met TSTEP	cm
CFRAC	Total Cloud Fraction	fraction
CLDT	Cloud Top Layer Height (M)	meter
CLDB	Cloud Bottom Layer Height (M)	meter
WBAR	Avg. Liquid Water Content Of Cloud	g/m^3
SNOCOV	Snow Cover	fraction
VEG	Vegetation Coverage (Decimal)	Fraction
LAI	Leaf-Area Index	m^2/m^2
SEAICE	Sea Ice	Fraction
WR	Canopy Moisture Content	M
SOIM1	Volumetric Soil Moisture In Top Cm	m^3/m^3
SOIM2	Volumetric Soil Moisture In Top M	m^3/m^4
SOIT1	Soil Temperature In Top Cm	Kelvin
SOIT2	Soil Temperature In Top M	Kelvin
SLTYP	Soil Texture Type By USDA	Category

Abb.	Variable Name (CMAQ)	Units
PM_{10}	Particulate Matter-10	μg/m^3
$PM_{2.5}$	Particulate Matter-2.5	μg/m^3
O_3	Ozone	ppmV
NO_x	Nitrogen Oxides	ppmV
NO	Nitric Oxide	ppmV
NO_2	Nitrogen dioxide	ppmV
ISOPRENE	Isoprene	ppmV
OLES	Olefins	ppmV
AROS	Aromatics	ppmV
ALKS	Allanes	ppmV

Table T7: *Average of all stations mean bias for the CMAQ model and both methods of CNN model*

Mean Bias	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14
CMAQ	1.21	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1 / MSE	-1.23	-2.11	-3.23	-2.59	-1.55	-2.06	-1.62	-1.48	-2.08	-2.78	-2.26	-2.47	-3.24	-2.62
Method2 / IOA	-0.96	-1.01	-1.69	-0.22	-0.25	0.65	-0.84	-0.54	-0.95	0.01	-0.15	0.84	-0.29	-1.89

Table T8: *Average of all stations root mean squared error for the CMAQ model and both methods of CNN model*

RMSE	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14
CMAQ	18.98	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1/ MSE	11.00	13.48	15.82	16.14	16.05	16.00	16.03	15.80	16.43	16.61	16.55	16.39	16.38	16.45
Method2/ IOA	11.01	13.30	15.59	16.27	16.44	15.95	15.60	16.08	16.78	16.77	17.31	16.55	16.64	16.06

Table T9: *Average of all stations correlations for the CMAQ model and both methods of CNN model*

Correlation	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14
CMAQ	0.63	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1/ MSE	0.82	0.72	0.62	0.60	0.60	0.61	0.60	0.61	0.59	0.58	0.59	0.59	0.59	0.59
Method2/ IOA	0.84	0.76	0.67	0.66	0.66	0.66	0.66	0.66	0.64	0.63	0.65	0.64	0.64	0.63

Table T10: *Average of all stations hit rate for the CMAQ model and both methods of CNN model*

Hit Rate	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14
CMAQ	0.77	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1/ MSE	0.67	0.50	0.39	0.36	0.40	0.42	0.41	0.37	0.38	0.34	0.37	0.33	0.31	0.34
Method2/ IOA	0.80	0.74	0.58	0.66	0.63	0.64	0.57	0.62	0.61	0.60	0.69	0.61	0.63	0.47

Table T11: *Average of all stations false alarm rate for the CMAQ model and both methods of CNN model*

False Alarm Rate	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**	**Day 8**	**Day 9**	**Day 10**	**Day 11**	**Day 12**	**Day 13**	**Day 14**
CMAQ	0.43	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1/ MSE	0.23	0.28	0.44	0.47	0.49	0.48	0.46	0.47	0.49	0.51	0.53	0.50	0.48	0.47
Method2/ IOA	0.28	0.35	0.48	0.50	0.50	0.49	0.47	0.48	0.52	0.53	0.54	0.53	0.51	0.49

Table T12: *Average of all stations critical success index for the CMAQ model and both methods of CNN model*

Critical Success Index	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**	**Day 8**	**Day 9**	**Day 10**	**Day 11**	**Day 12**	**Day 13**	**Day 14**
CMAQ	0.47	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1/ MSE	0.56	0.43	0.30	0.28	0.30	0.31	0.31	0.29	0.29	0.25	0.27	0.25	0.24	0.26
Method2/ IOA	0.61	0.53	0.38	0.40	0.39	0.40	0.38	0.39	0.37	0.36	0.38	0.36	0.38	0.32

Table T13: *Average of all stations equitable threat score for the CMAQ model and both methods of CNN model*

Equitable Thread Score	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**	**Day 8**	**Day 9**	**Day 10**	**Day 11**	**Day 12**	**Day 13**	**Day 14**
CMAQ	0.34	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1/ MSE	0.47	0.34	0.20	0.18	0.19	0.20	0.20	0.18	0.18	0.15	0.15	0.15	0.15	0.16
Method2/ IOA	0.51	0.42	0.26	0.26	0.25	0.26	0.25	0.26	0.23	0.21	0.22	0.21	0.24	0.20

Table T14: *Average of all stations proportion of correct for the CMAQ model and both methods of CNN model*

Proportion of Correct	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**	**Day 8**	**Day 9**	**Day 10**	**Day 11**	**Day 12**	**Day 13**	**Day 14**
CMAQ	0.80	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Method1/ MSE	0.88	0.84	0.79	0.78	0.78	0.78	0.78	0.78	0.77	0.76	0.76	0.77	0.77	0.77
Method2/ IOA	0.88	0.85	0.78	0.77	0.76	0.77	0.78	0.78	0.75	0.74	0.73	0.74	0.76	0.77

Table T15: *Detailed architecture of CNN model*

DNN Layers	Dimensions	Additional Comments
Input Layer	(1128, 1)	
Convolutional Layer 1	(1127 x 32)	Filters 32, kernel - (2 x1)
Convolutional Layer 2	(1126 x 32)	Filters 32, kernel - (2 x1)
Convolutional Layer 3	(1125 x 32)	Filters 32, kernel - (2 x1)
Convolutional Layer 4	(1124 x 32)	Filters 32, kernel - (2 x1)
Convolutional Layer 5	(1123 x 32)	Filters 32, kernel - (2 x1)
Flatten	(35936)	
Dense	(264)	
Dense, Output Layer	(24)	

Table T16: *Performance in terms of IOA and correlation for the CMAQ and the CNN models (average of all stations) for* NO_2

	Day - 1	Day - 2	Day - 3	Day - 4	Day - 5	Day - 6	Day - 7
CMAQ - IOA	0.60	0.60	0.57	-	-	-	-
CNN - IOA	0.85	0.76	0.67	0.65	0.62	0.66	0.66
CMAQ - Correlation	0.45	0.45	0.41	-	-	-	-
CNN - Correlation	0.75	0.60	0.46	0.43	0.40	0.45	0.45

Table T17: Performance in terms of IOA and correlation for the CMAQ and the CNN models (average of all stations) for $PM_{2.5}$

	Day - 1	Day - 2	Day - 3	Day - 4	Day - 5	Day - 6	Day - 7
CMAQ - IOA	0.76	0.75	0.69	-	-	-	-
CNN - IOA	0.86	0.75	0.60	0.52	0.50	0.48	0.49
CMAQ - Correlation	0.62	0.60	0.52	-	-	-	-
CNN - Correlation	0.76	0.61	0.37	0.29	0.26	0.25	0.24

Table T18: *Performance in terms of IOA and correlation for the CMAQ and the CNN models (average of all stations) for* PM_{10}

	Day - 1	Day - 2	Day - 3	Day - 4	Day - 5	Day - 6	Day - 7
CMAQ - IOA	0.68	0.65	0.59	-	-	-	-
CNN - IOA	0.83	0.73	0.57	0.52	0.51	0.53	0.50
CMAQ - Correlation	0.53	0.51	0.44	-	-	-	-
CNN - Correlation	0.71	0.55	0.35	0.27	0.26	0.26	0.26

Table T19. *List of WRF meteorological parameters extracted from each grid point and used as input for the CNN model. (*WRF diagnostic variables)*

Symbol	***Description***	***Units***
P_HYD	Hydrostatic Pressure	Pa
Q2	Water Vapor Mixing Ratio at 2m	kg/kg
T2	Temperature at 2m	K
TH2	Potential Temperature at 2m	K
PSFC	Surface Pressure	Pa
U10	U Wind at 10 M	m/s
V10	V Wind at 10m	m/s
QVAPOR	Water Vapor Mixing Ratio	kg/kg
QCLOUD	Cloud Water Mixing Ratio	kg/kg
QRAIN	Rain Water Mixing Ratio	kg/kg
SHDMAX	Annual Max Veg Fraction	
SHDMIN	Annual Min Veg Fraction	
SNOALB	Annual Max Snow Albedo in Fraction	
TSLB	Soil Temperature	K
SMOIS	Soil Moisture	m3/m3
SH2O	Soil Liquid Water	m3/m3
SFROFF	Surface Runoff	mm
UDROFF	Underground Runoff	mm
IVGTYP	Dominant Vegetation Category	
ISLTYP	Dominant Soil Category	
VEGFRA	Vegetation Fraction	
GRDFLX	Ground Heat Flux	W/m2
ACGRDFLX	Accumulated Ground Heat Flux	J/m2
ACSNOM	Accumulated Melted Snow	kg/m2
SNOW	Snow Water Equivalent	kg/m2
SNOWH	Physical Snow Depth	m
CANWAT	Canopy Water	kg/m2
SSTSK	Skin Sea Surface Temperature	K
COSZEN	Cos of Solar Zenith Angle	
LAI	Leaf Area Index	m2/m2
VEGF_PX	Vegetation Fraction for PX LSM	area/area
CANFRA	Satellite Canopy Fraction	
VAR	Orographic Variance	
F	Coriolis Sine Latitude Term	s-1
E	Coriolis Cosine Latitude Term	s-1
HGT	Terrain Height	m
RAINC	Accumulated Total Cumulus Precipitation	mm
RAINSH	Accumulated Shallow Cumulus Precipitation	mm
RAINNC	Accumulated Total Grid-Scale Precipitation	mm

SNOWNC	Accumulated Total Grid-Scale Snow and Ice	mm
GRAUPELNC	Accumulated Total Grid Scale Graupel	mm
HAILNC	Accumulated Total Grid-Scale Hail	mm
CLDFRA	Cloud Fraction	
SWDOWN	Downward Short-Wave Flux at Ground Surface	W/m2
GLW	Downward Long-Wave Flux at Ground Surface	W/m2
SWNORM	Normal Short-Wave Flux at Ground Surface (Slope-Dependent)	W/m2
OLR	TOA Outgoing Long Wave	W/m2
ALBEDO	Albedo	
ALBBCK	Background Albedo	
EMISS	Surface Emissivity	
NOAHRES	Residual of the NOAH Surface Energy Budget	W/m2
TMN	Soil Temperature at Lower Boundary	K
XLAND	Land Mask	
PBLH	PBL Height	m
HFX	Upward Heat Flux at the Surface	W/m2
QFX	Upward Moisture Flux at the Surface	kg m-2 s-
LH	Latent Heat Flux at the Surface	W/m2
SNOWC	Flag Indicating Snow Coverage	
SR	Fraction of Frozen Precipitation	
SST	Sea Surface Temperature	K
Ue10*	U-wind in Earth Coordinate	m/s
Ve10*	V-wind in Earth Coordinate	m/s
WS*	Wind Speed	m/s
WD*	Wind Direction	

Table T20. *Table showing the Latitude, Longitude, and distance between each station and WRF grid points.*

Station ID	***Station Latitude***	***Station Longitude***	***WRF Latitude***	***WRF Longitude***	***Distance (in km)***
100	37.68	128.72	37.59	128.67	10.40
101	37.90	127.74	37.86	127.74	4.68
102	37.97	124.63	37.87	124.58	12.07
104	37.80	128.86	37.83	129.00	12.91
105	37.75	128.89	37.83	129.00	12.94
106	37.51	129.12	37.58	128.99	14.72
108	37.57	126.97	37.62	127.10	13.07
112	37.48	126.62	37.38	126.47	17.69
114	37.34	127.95	37.36	128.04	8.18
115	37.48	130.90	37.51	130.87	4.00
119	37.27	126.99	37.37	127.10	14.75
121	37.18	128.46	37.10	128.34	13.72
127	36.97	127.95	36.86	128.02	13.76
129	36.78	126.49	36.88	126.47	11.58

130	36.99	129.41	37.07	129.28	15.32
131	36.64	127.44	36.62	127.39	4.67
133	36.37	127.37	36.37	127.39	1.43
135	36.22	127.99	36.11	128.00	11.80
136	36.57	128.71	36.60	128.63	7.27
137	36.41	128.16	36.36	128.00	14.60
138	36.03	129.38	36.07	129.53	14.32
140	36.01	126.76	35.88	126.77	13.59
143	35.83	128.65	35.85	128.60	5.32
146	35.84	127.12	35.88	127.07	6.14
152	35.58	129.33	35.59	129.20	12.13
155	35.17	128.57	35.11	128.57	6.80
156	35.17	126.89	35.14	126.76	12.81
159	35.10	129.03	35.09	129.18	13.20
162	34.85	128.44	34.86	128.56	11.67
165	34.82	126.38	34.89	126.45	10.79
168	34.74	127.74	34.63	127.65	14.22
169	34.69	125.45	34.65	125.55	10.04
170	34.40	126.70	34.40	126.75	4.26
172	35.35	126.60	35.39	126.46	13.75
174	35.02	127.37	35.13	127.36	12.58
175	34.47	126.32	34.40	126.45	13.98
184	33.51	126.53	33.41	126.44	13.76
185	33.29	126.16	33.42	126.15	13.56
188	33.39	126.88	33.41	126.74	13.59
189	33.25	126.57	33.17	126.44	14.48
192	35.16	128.04	35.12	127.97	8.05
201	37.71	126.45	37.63	126.47	9.43
202	37.49	127.49	37.37	127.41	15.43
203	37.26	127.48	37.37	127.41	13.24
211	38.06	128.17	38.10	128.06	10.55
212	37.68	127.88	37.61	127.73	15.51
216	37.17	128.99	37.08	128.96	9.81
217	37.38	128.65	37.34	128.66	4.60
221	37.16	128.19	37.10	128.34	14.45
226	36.49	127.73	36.37	127.70	13.72
232	36.76	127.29	36.87	127.40	15.29
235	36.33	126.56	36.38	126.46	10.40
236	36.27	126.92	36.38	126.77	17.93
238	36.11	127.48	36.12	127.38	9.15
243	35.73	126.72	35.64	126.76	11.27

244	35.61	127.29	35.63	127.37	8.03
245	35.56	126.87	35.64	126.76	12.33
247	35.40	127.40	35.38	127.37	3.60
248	35.66	127.52	35.63	127.37	13.75
251	35.43	126.70	35.39	126.76	7.14
252	35.28	126.48	35.39	126.46	11.92
253	35.23	128.89	35.35	128.88	13.14
254	35.37	127.13	35.38	127.06	6.11
255	35.23	128.67	35.11	128.57	15.88
257	35.31	129.02	35.35	128.88	13.08
258	34.76	127.21	34.64	127.35	18.80
259	34.63	126.77	34.65	126.75	2.70
260	34.69	126.92	34.64	127.05	13.10
261	34.55	126.57	34.65	126.45	15.03
262	34.62	127.28	34.64	127.35	7.31
263	35.32	128.29	35.36	128.28	4.70
264	35.51	127.75	35.62	127.68	13.96
266	34.94	127.69	34.88	127.66	7.48
268	34.47	126.26	34.40	126.15	12.79
271	36.94	128.91	36.84	128.95	12.48
272	36.87	128.52	36.84	128.64	11.48
273	36.63	128.15	36.61	128.01	12.24
276	36.43	129.04	36.34	128.93	14.45
277	36.53	129.41	36.57	129.56	13.88
278	36.36	128.69	36.35	128.62	6.06
279	36.13	128.32	36.11	128.30	2.89
281	35.98	128.95	36.09	128.92	13.01
283	35.82	129.20	35.83	129.21	2.10
284	35.67	127.91	35.62	127.98	8.55
285	35.57	128.17	35.61	128.29	11.75
288	35.49	128.74	35.60	128.90	17.95
289	35.41	127.88	35.37	127.97	9.83
294	34.89	128.60	34.86	128.56	4.86
295	34.82	127.93	34.88	127.96	7.24
90	38.25	128.56	38.34	128.70	15.59
95	38.15	127.30	38.11	127.43	11.31
98	37.90	127.06	37.87	127.11	5.32
99	37.89	126.77	37.87	126.79	2.52

Table T21. *The comparison of the WRF, CNN, and various linear models based on IOA and correlation.*

Model	IOA	Correlation
WRF	0.67	0.66
Linear Regression (Station-wise)	0.43	0.42
Lasso Regression (Station-wise)	0.81	0.74
Linear Regression (General)	0.77	0.69
Lasso Regression (General)	0.77	0.69
CNN (STATION-WISE)	0.86	0.77
CNN (General)	0.85	0.75

B: Figures

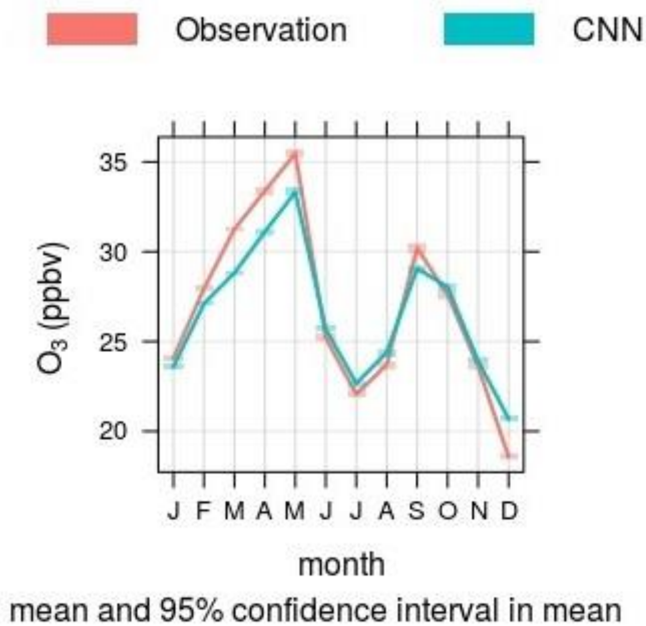


Figure F1. *Time-variation (Observation & Prediction) plot – and average monthly variations.*

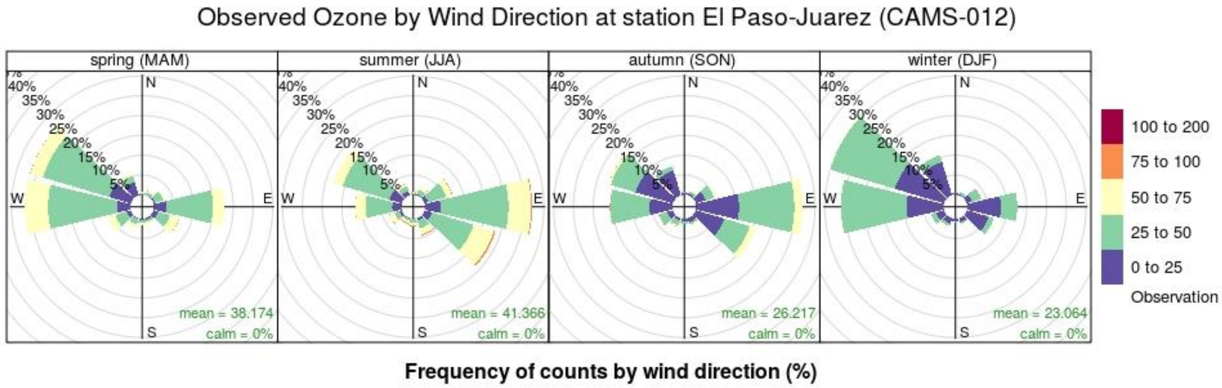


Figure F2. *Season-wise Wind Direction by O_3 concentration for station CAMS-012.*

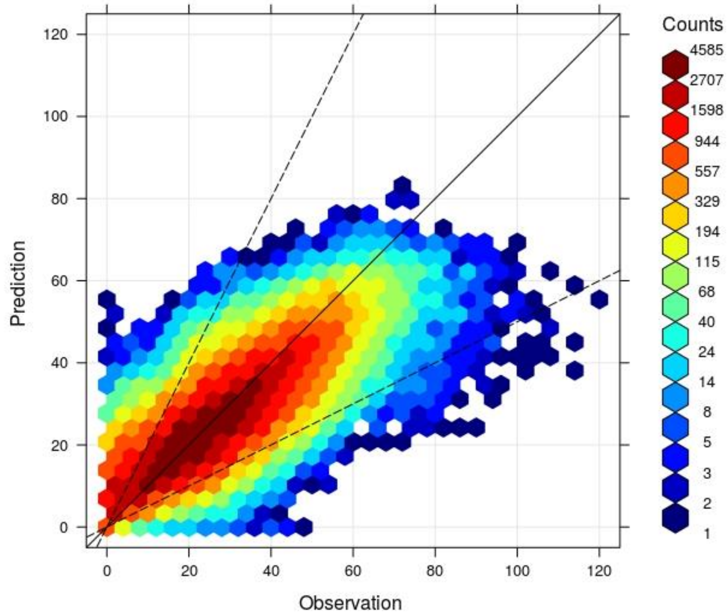


Figure F3*. Scatterplot of Observation vs. Prediction for all stations combined. The color intensity indicates the frequency of occurrence.*

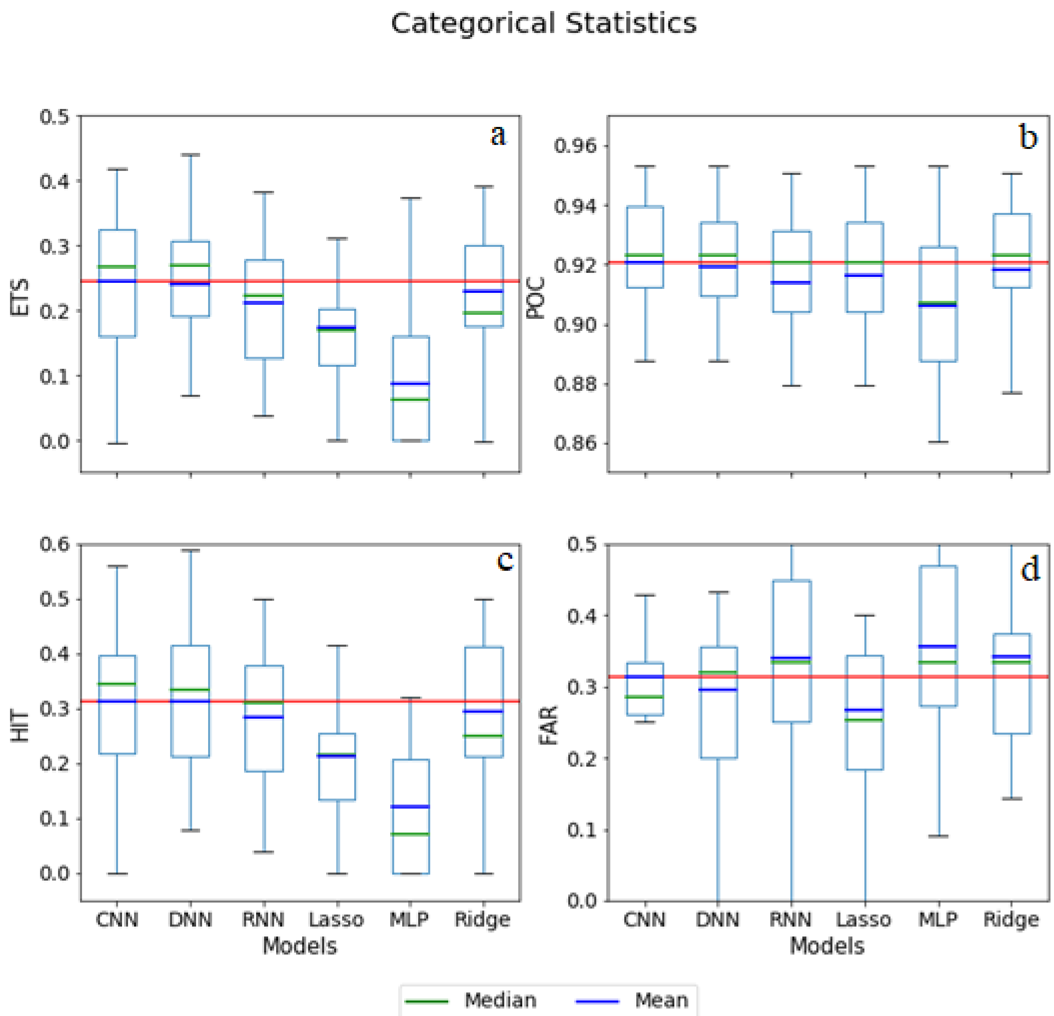


Figure F4. *Box and whisker plot of categorical statistics of the CNN model for all stations. a) represents Equitable Threat Score (ETS), b) represents Proportion of Correct (POC), c) represents Hit rate (HIT), and d) represents False Alarm Rate (FAR).*

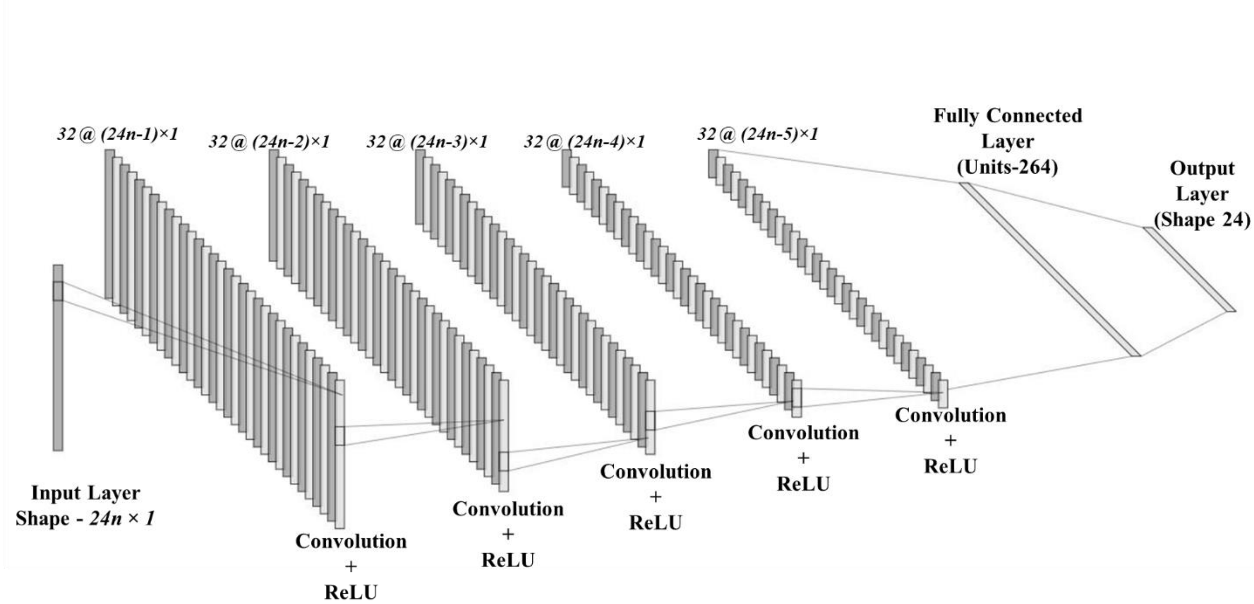


Figure F5. *Block Diagram of the CNN model. "n" was the number of input variables used. Kernel size was 2×1 and the number of filters was 32 for each layer.*

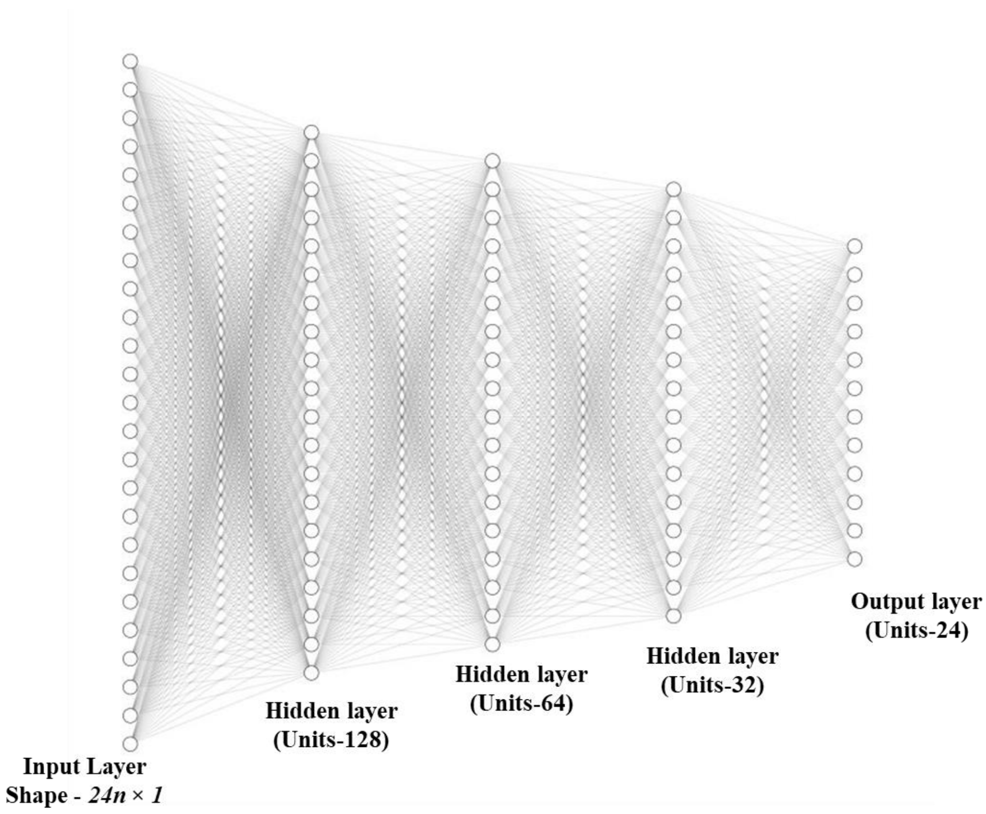


Figure F6. *Block Diagram of DNN and RNN model. For DNN the hidden layers are Dense and for RNN hidden layers are GRU.*

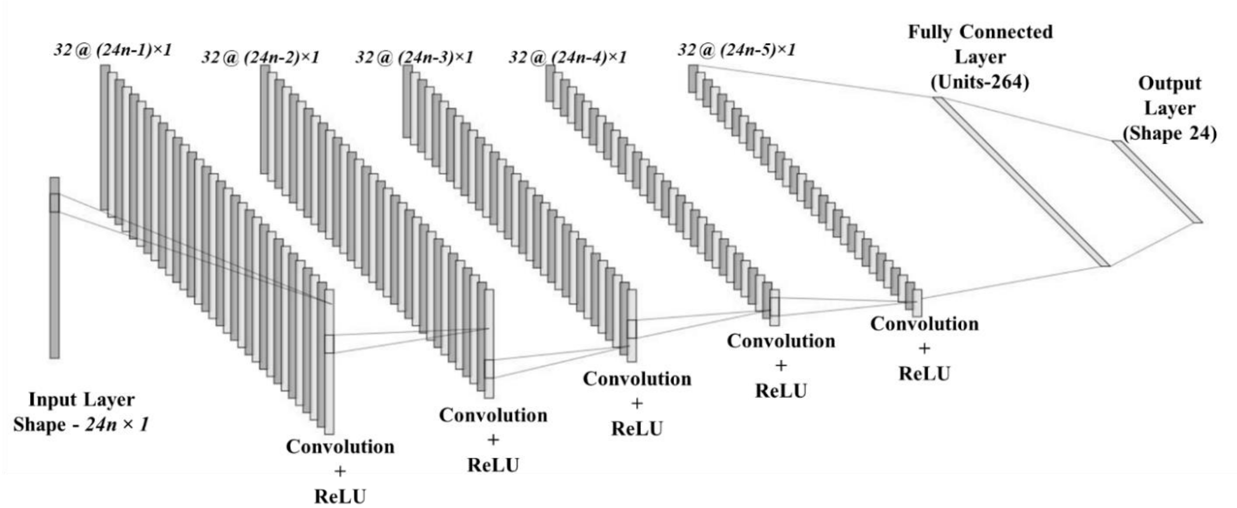


Figure F7: *Schematic diagram of the Convolution Neural Network. 'n' was the number of input parameters (meteorology, air quality, and observations) used. Here, n=50 (32 meteorology, 14 air-quality parameters, and 4 previous day observations), so there are 1200 (=52 × 24) inputs and 24 outputs. Schematics are prepared by the NNSVG tool* (LeNail, 2019).

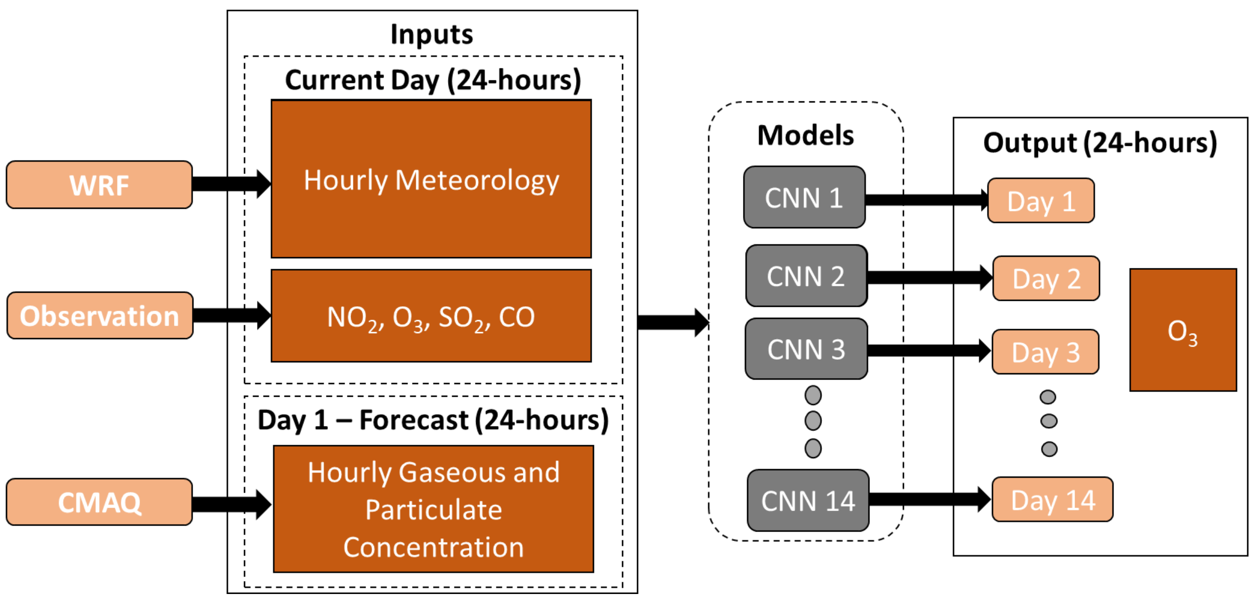


Figure F8: *Schematic diagram of the process flow of the CNN model. Note: there are 1200 input parameters (32 meteorology, 14 parameters from CMAQ and 4 parameters from previous day observation; 50 × 24 = 1200) for 24 output parameters (1 day; 24-hour ozone concentration) for each day model.*

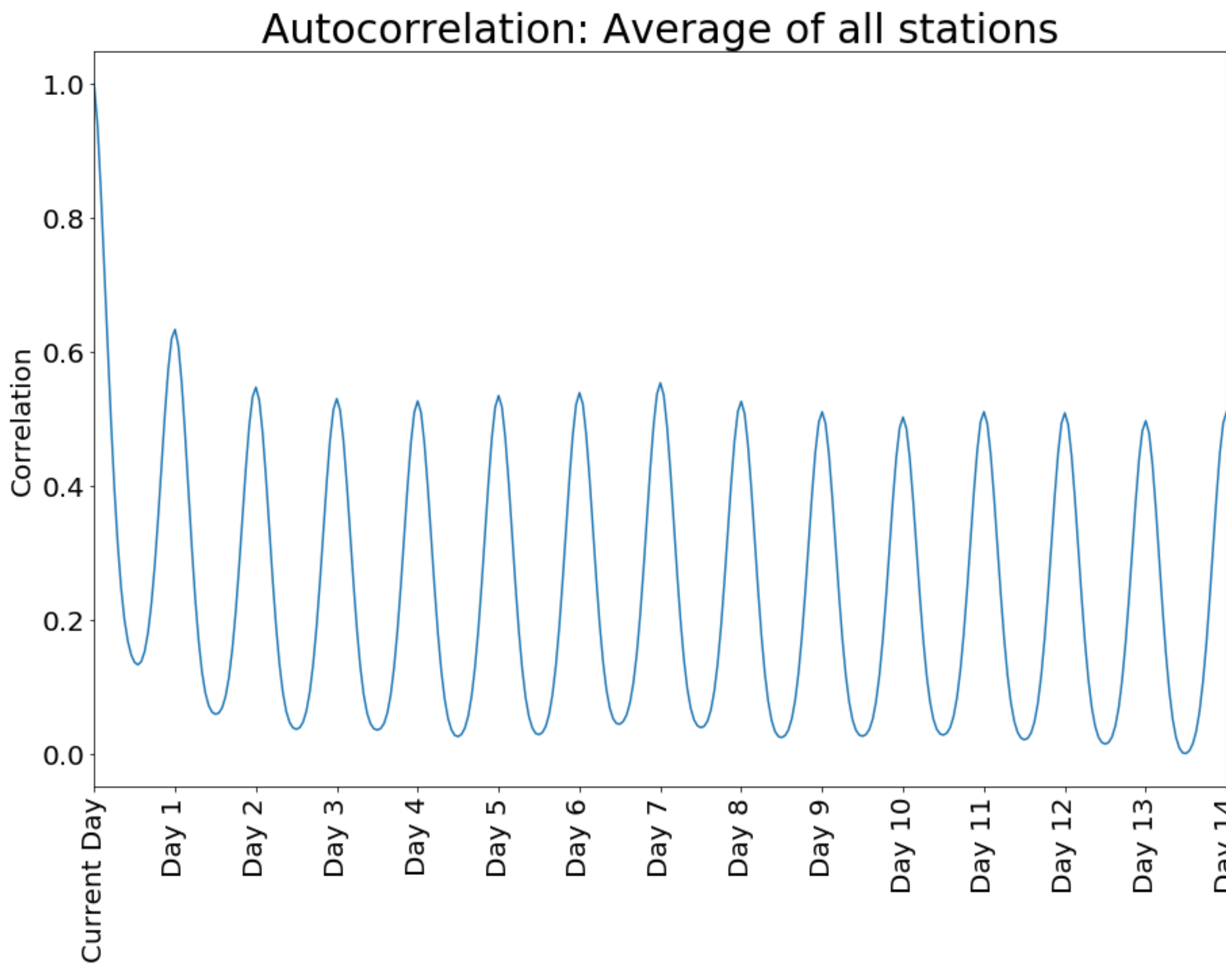


Figure F97: *Auto-correlations (average of all stations) of the observed current hour ozone concentration with the subsequent hour observed ozone concentration. The X-axis represents the hours, and the y-axis represents the correlation. (Correlation of 0th hour with 0th hour, 1st hour, 2nd hour, and so on. It was analogous to delayed response in electrical signals)*

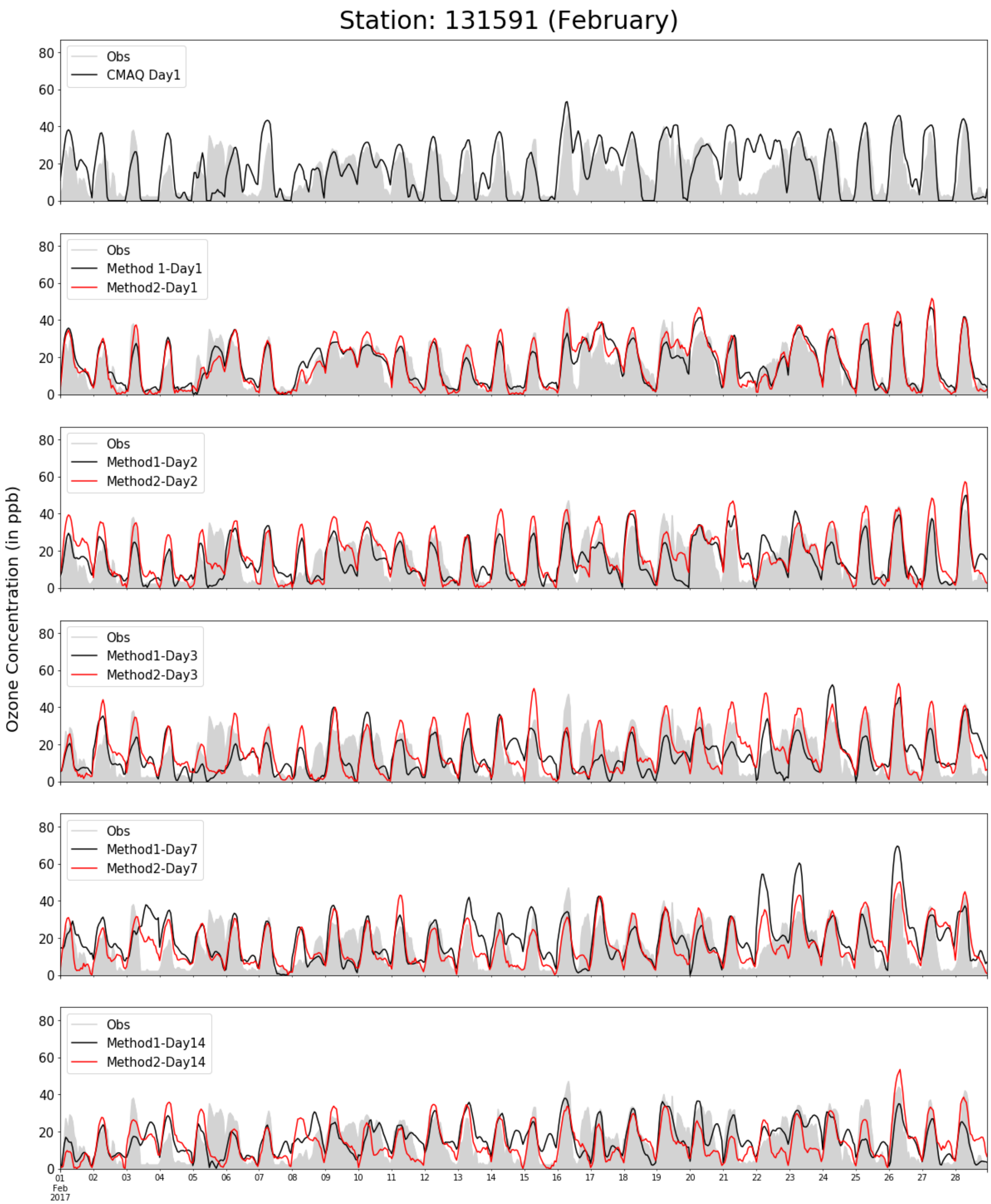


Figure F10: *Hourly time-series plot for station 131591 for February 2017. The panels from top to bottom show a time-series comparison of the observed ozone concentration with the CMAQ day 1 forecast, the CNN-method 2 day1, day 2, day 3, day 7, and day 14 forecasts, respectively. The X-axis represents the days of the month and the y-axis represents the ozone concentration (in ppb).*

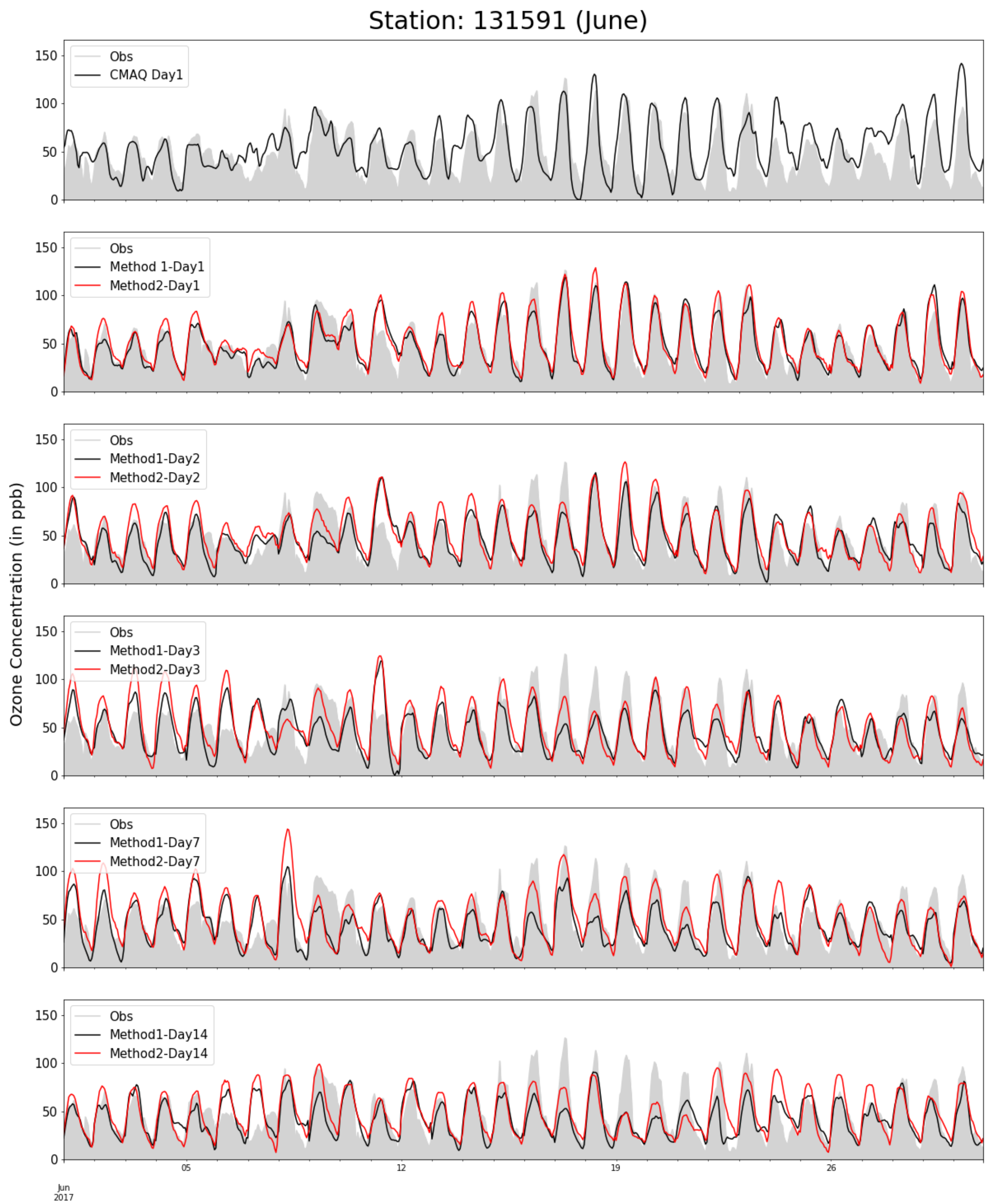


Figure F11: *Hourly time-series plot for station 131591 for the month of June 2017. The panels from top to bottom show a time-series comparison of the observed ozone concentration with the CMAQ day 1 forecast, the CNN-method 2 day1, day 2, day 3, day 7, and day 14 forecasts, respectively. The X-axis represents the days of the month and the y-axis represents the ozone concentration (in ppb).*

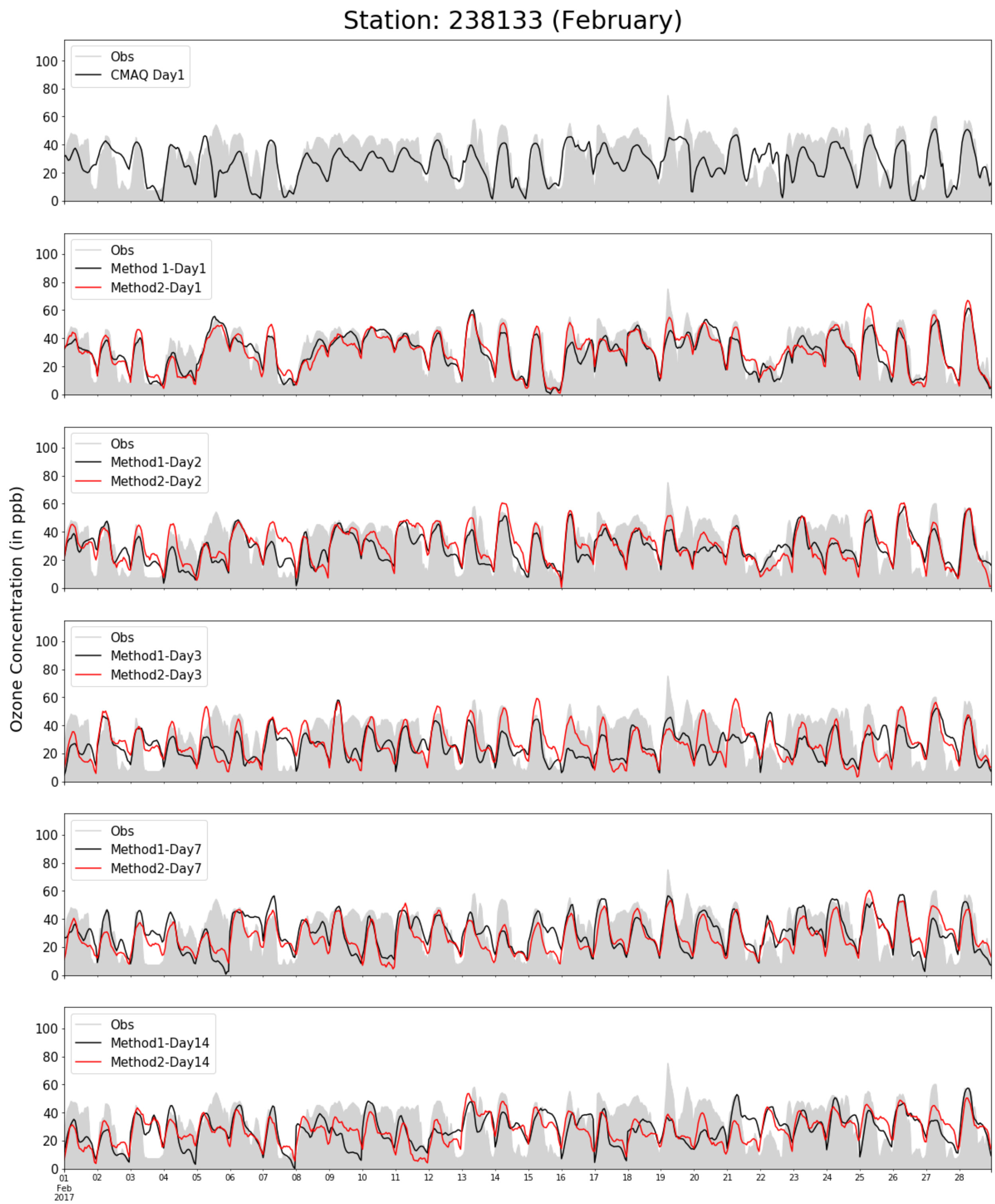


Figure F12: *Hourly time-series plot for station 238133 for the month of February 2017. The panels from top to bottom show a time-series comparison of the observed ozone concentration with the CMAQ day 1 forecast, the CNN-method 2 day1, day 2, day 3, day 7, and day 14 forecasts, respectively. The X-axis represents the days of the month and the y-axis represents the ozone concentration (in ppb).*

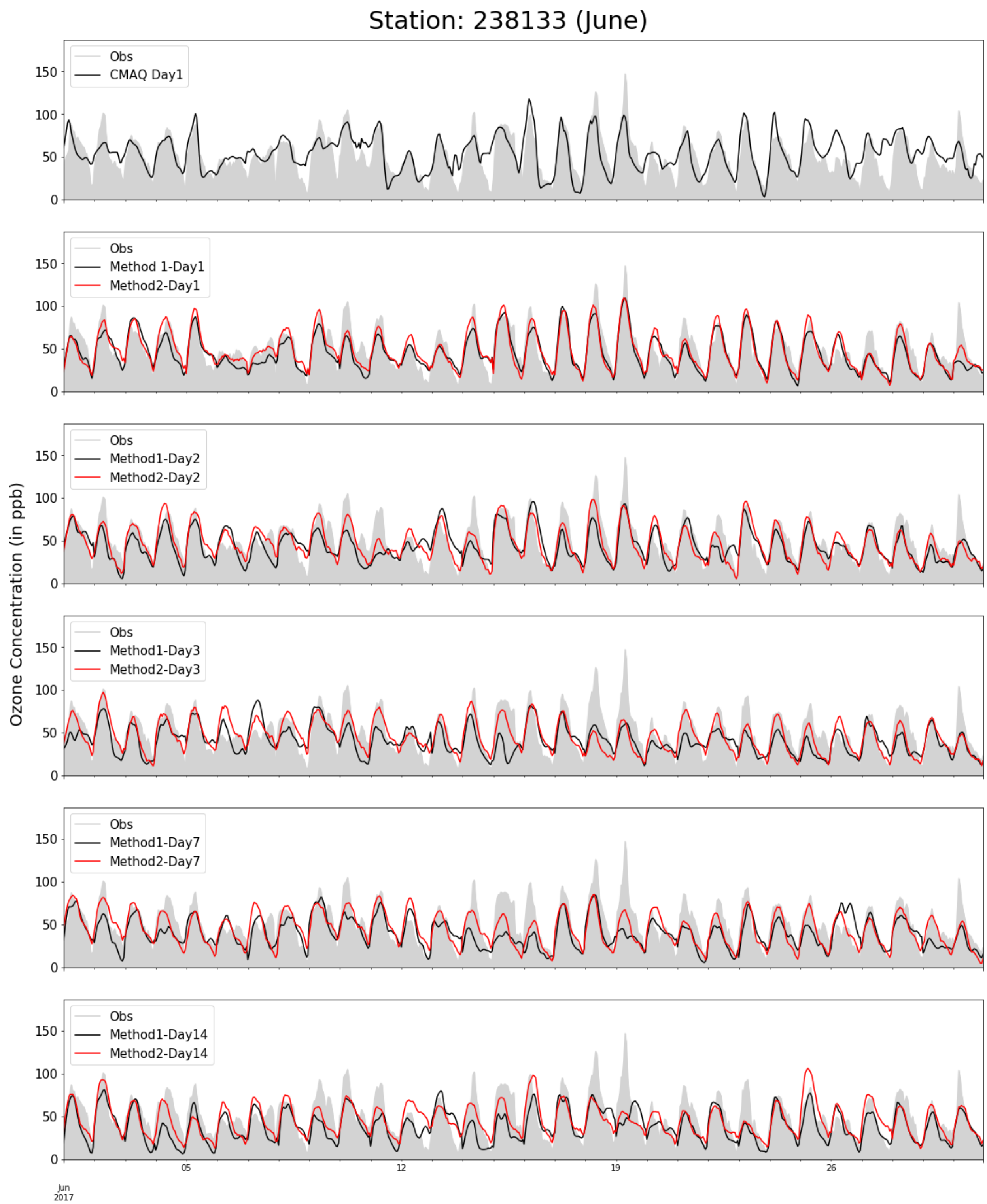


Figure F13: *Hourly time-series plot for station 238133 for the month of June 2017. The panels from top to bottom show a time-series comparison of the observed ozone concentration with the CMAQ day 1 forecast, the CNN-method 2 day1, day 2, day 3, day 7, and day 14 forecasts, respectively. The X-axis represents the days of the month and the y-axis represents the ozone concentration (in ppb).*

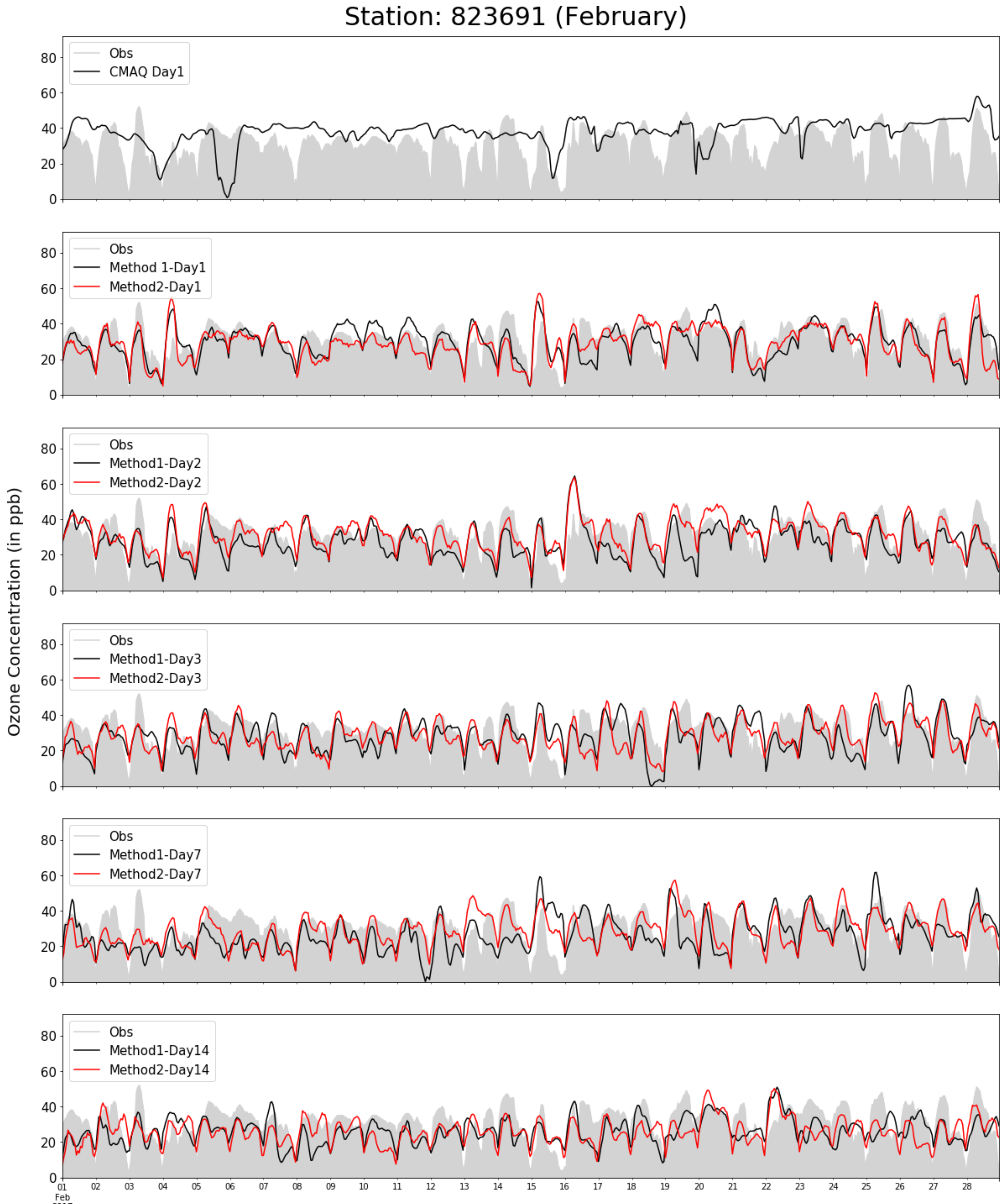


Figure F14: *Hourly time-series plot for station 823691 for the month of February 2017. The panels from top to bottom show a time-series comparison of the observed ozone concentration with the CMAQ day 1 forecast, the CNN-method 2 day1, day 2, day 3, day 7, and day 14 forecasts, respectively. The X-axis represents the days of the month and the y-axis represents the ozone concentration (in ppb).*

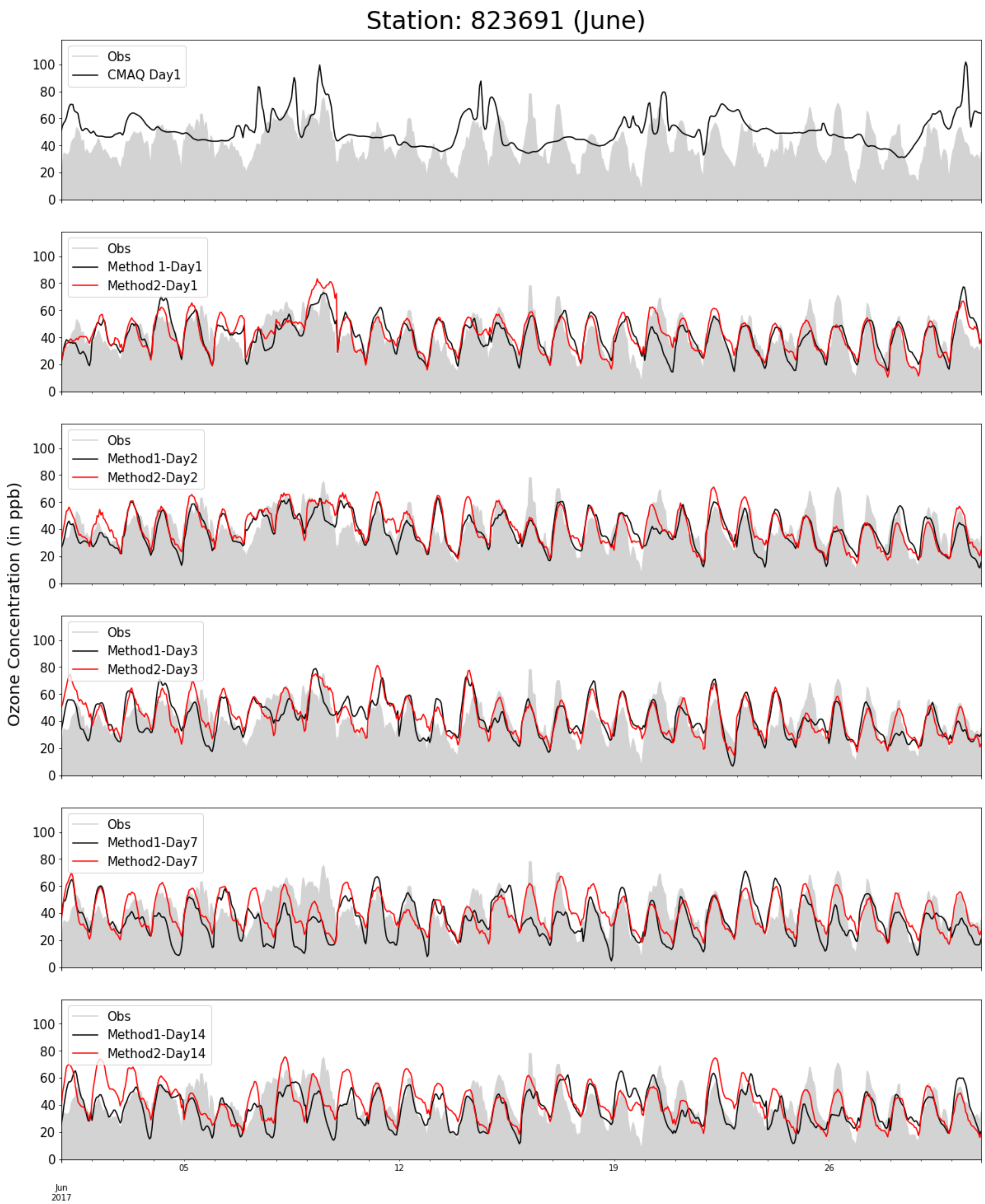


Figure F15: *Hourly time-series plot for station 823691 for the month of June 2017. The panels from top to bottom show a time-series comparison of the observed ozone concentration with the CMAQ day 1 forecast, the CNN-method 2 day1, day 2, day 3, day 7, and day 14 forecasts, respectively. The X-axis represents the days of the month and the y-axis represents the ozone concentration (in ppb).*

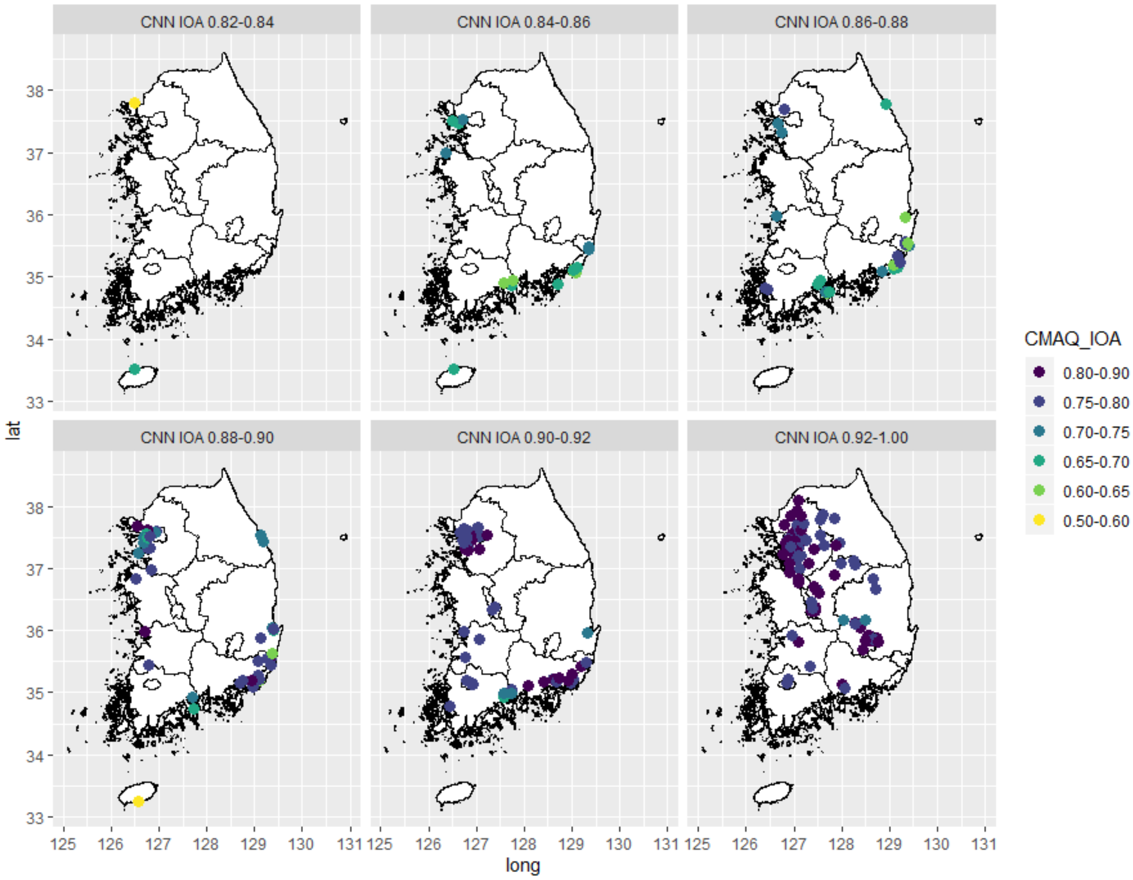


Figure F16: *Station-based CNN-IOA binned in specific ranges. A colored dot represents the location of the station, and a specific color represents the CMAQ-IOA. (Figures are created using R ggplot2* ("Create Elegant Data Visualisations Using the Grammar of Graphics," n.d.)*: https://ggplot2.tidyverse.org/)*

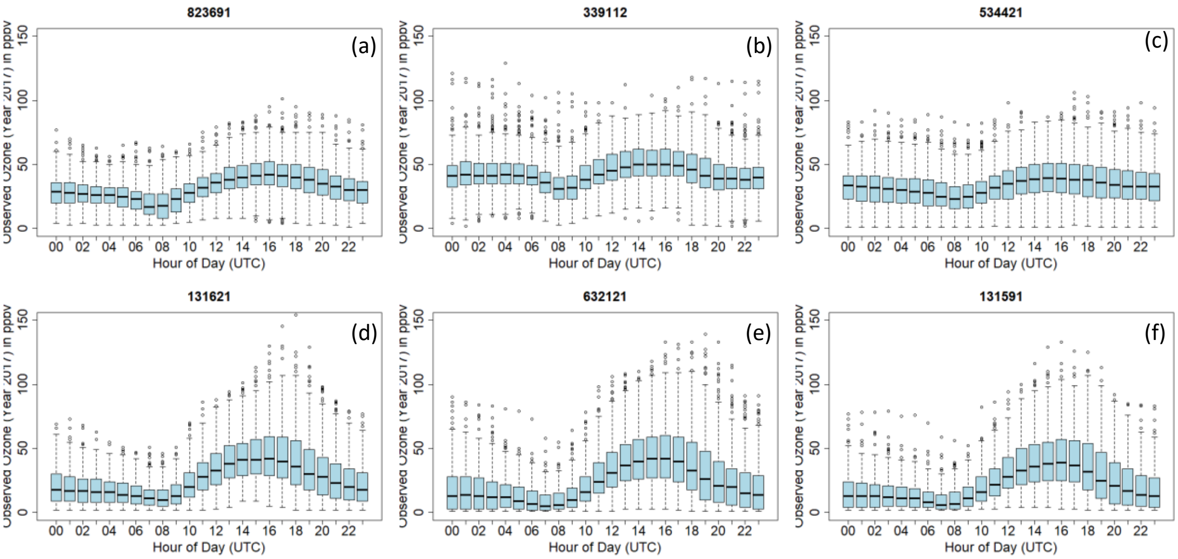


Figure F17: *Box and whisker plot 24-hour observed ozone concentration throughout the year 2017. a, b and c are the three worst-performing stations. d, e, and f are the best performing station.*

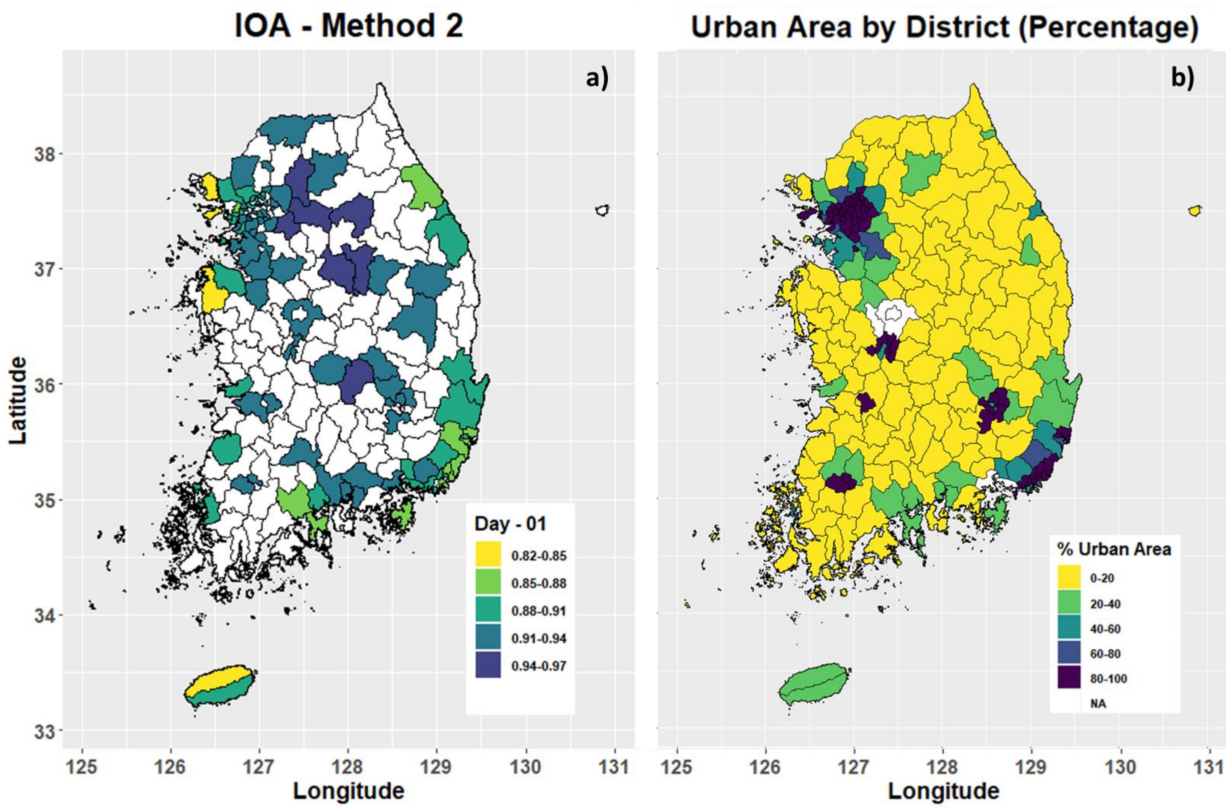


Figure F18: *a) District-wise IOA based on Method 2 of CNN. b) Percentage of Urban area*("Statistical Database | KOSIS KOrean Statistical Information Service," n.d.) *in each district of Korea. (Figures are created using R ggplot2* ("Create Elegant Data Visualisations Using the Grammar of Graphics," n.d.)*: https://ggplot2.tidyverse.org/*)

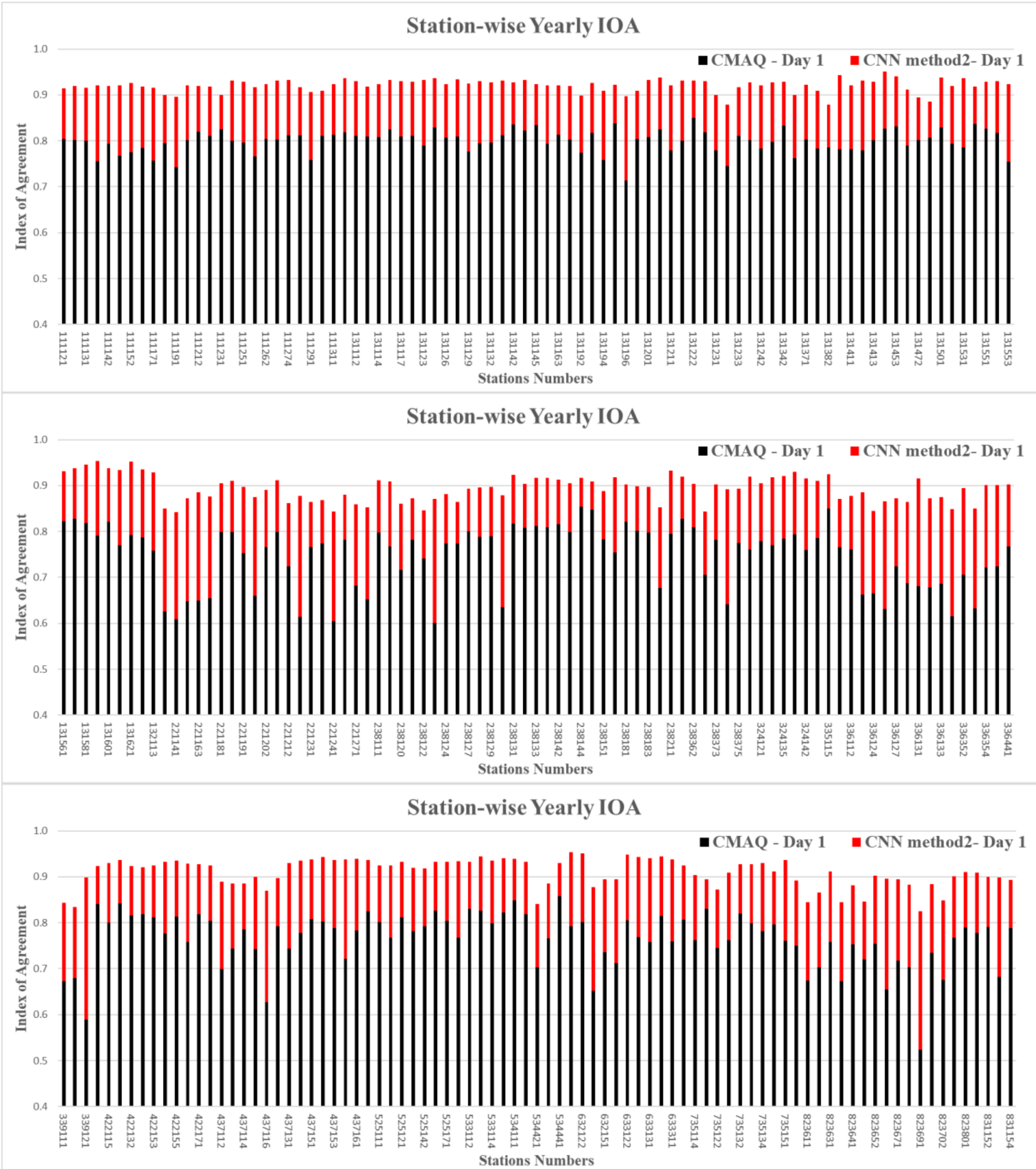


Figure F19: *Station-wise yearly index of agreement (IOA) for the CMAQ and the CNN-method 2 model for the day one forecast. The black bar represents the CMAQ models IOA. The sum of the black bar and red bar represents the IOA for the CNN-method 2 model. The red bar individually represents the absolute increase in the IOA from the CMAQ model. The X-axis represents IOA and the y-axis represents the station number.*

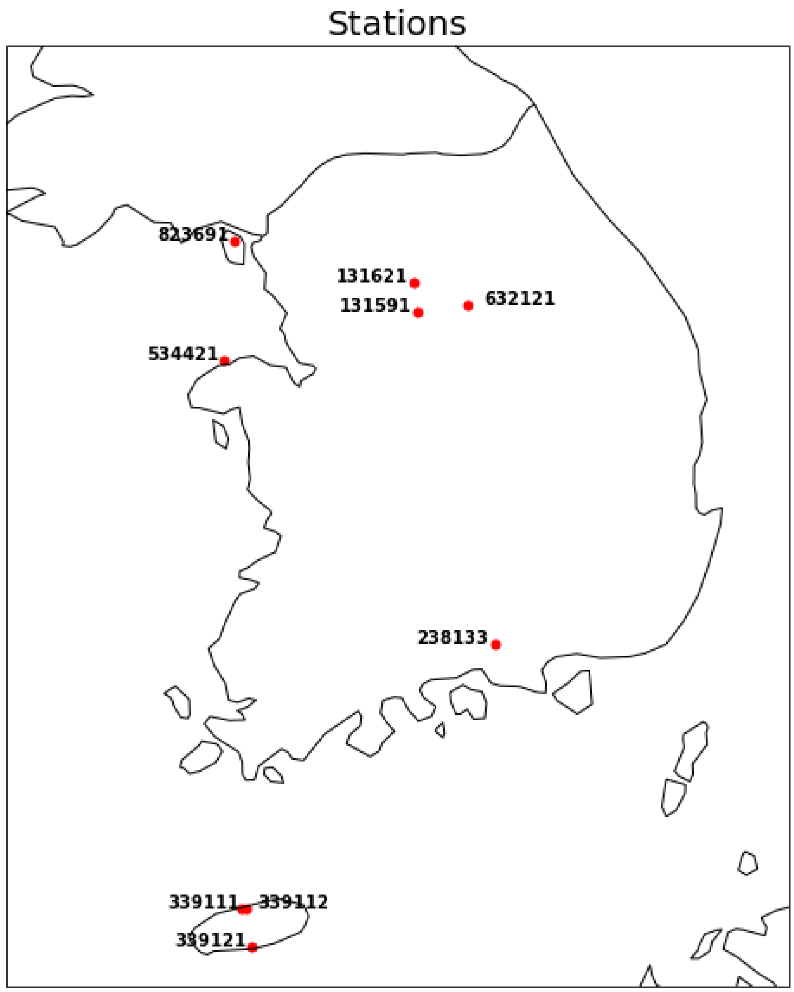


Figure F20: *Location of few stations specifically mentioned in the study. (Figures are created using python cartopy package*(“Introduction — cartopy 0.18.0 documentation,” n.d.)*: https://scitools.org.uk/cartopy/docs/latest/)*

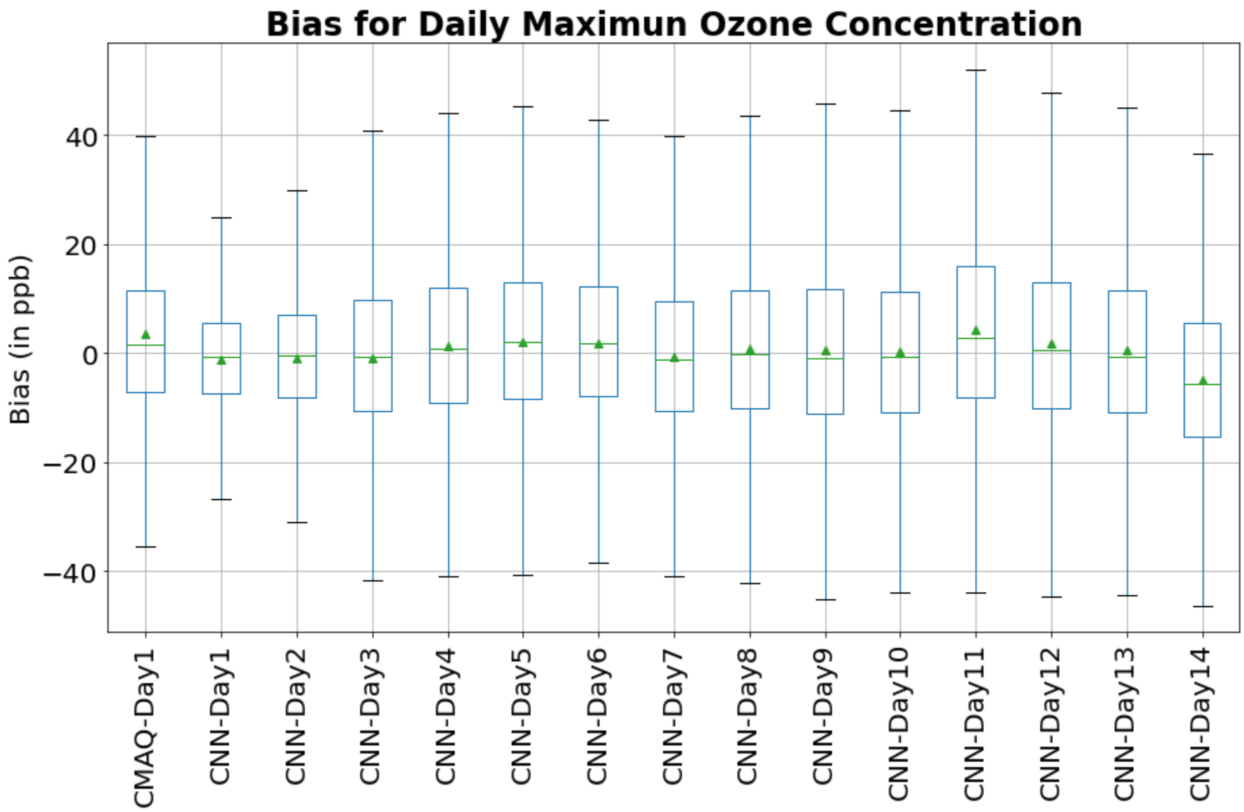


Figure F21: *Box plot of bias of daily maximum all stations combined. The x-axis represents the prediction days, and the y-axis represents the bias in ppb. The green line represents the median of bias, and the green triangle in each box represents the mean bias for that model—the extent of the box represents the interquartile range (IQR), i.e., 25th to 75th percentile value.*

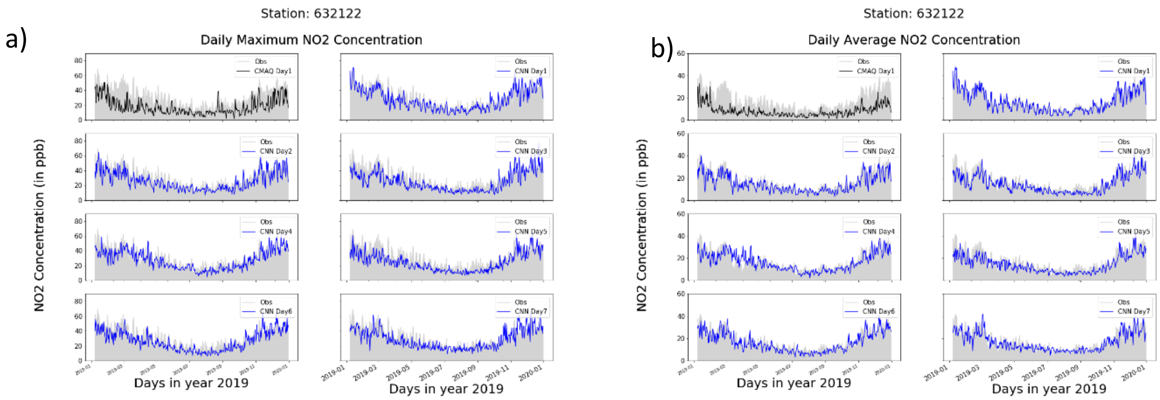


Figure F22: *a) Forecast of Daily Maximum of NO_2 concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 632122. b) Forecast of Daily Mean of NO_2 concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 632122.*

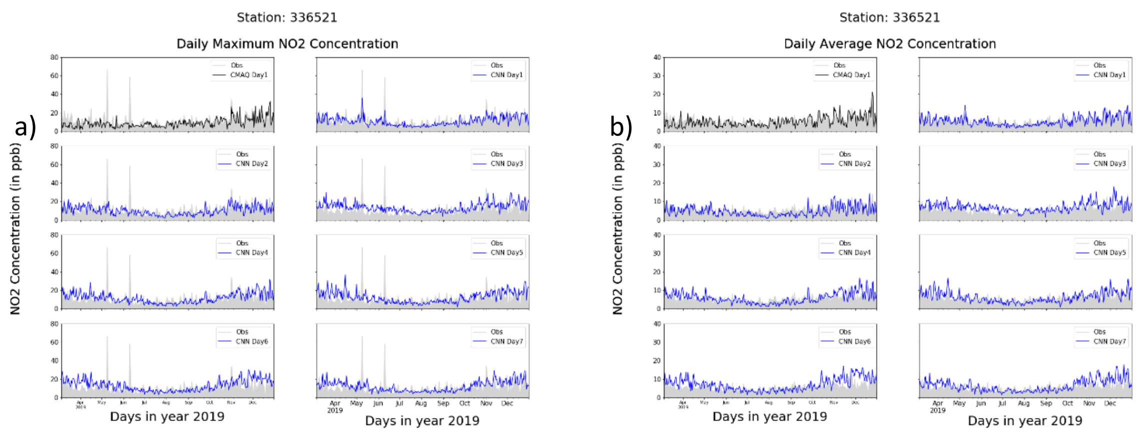


Figure F23: *a) Forecast of Daily Maximum of NO_2 concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 336521. b) Forecast of Daily Mean of NO_2 concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 336521.*

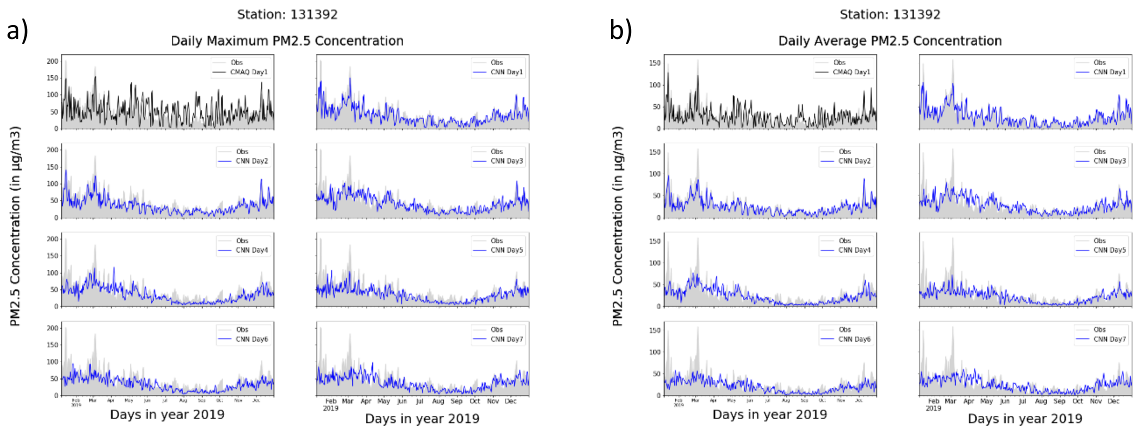


Figure F24: *a) Forecast of Daily Maximum of $PM_{2.5}$ concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 131392. b) Forecast of Daily Mean of $PM_{2.5}$ concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 131392.*

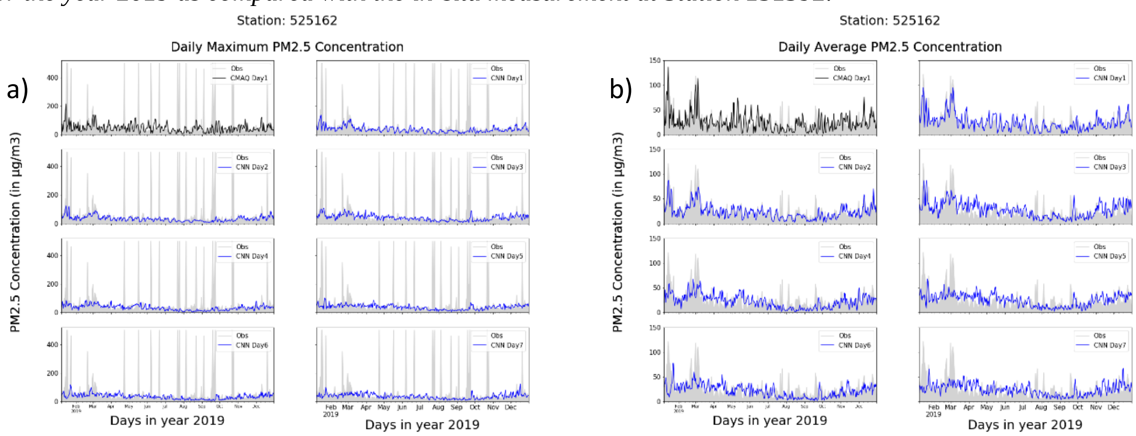


Figure F25: *a) Forecast of Daily Maximum of $PM_{2.5}$ concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 525162. b) Forecast of Daily Mean of $PM_{2.5}$ concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 525162.*

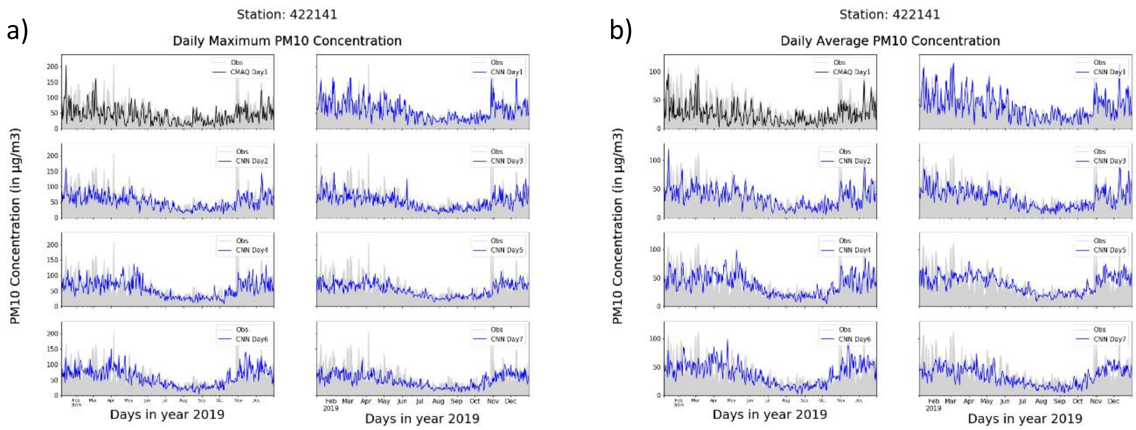


Figure F26: *a) Forecast of Daily Maximum of PM_{10} concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 422141. b) Forecast of Daily Mean of PM_{10} concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 422141.*

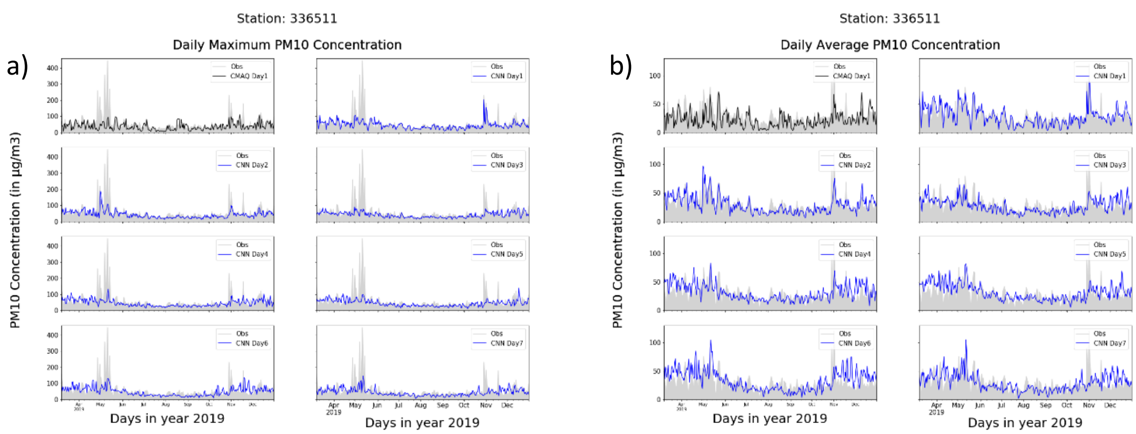


Figure F27: *a) Forecast of Daily Maximum of PM_{10} concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 336511. b) Forecast of Daily Mean of PM_{10} concentration 7-days in advance for the year 2019 as compared with the in-situ measurement at Station 336511.*

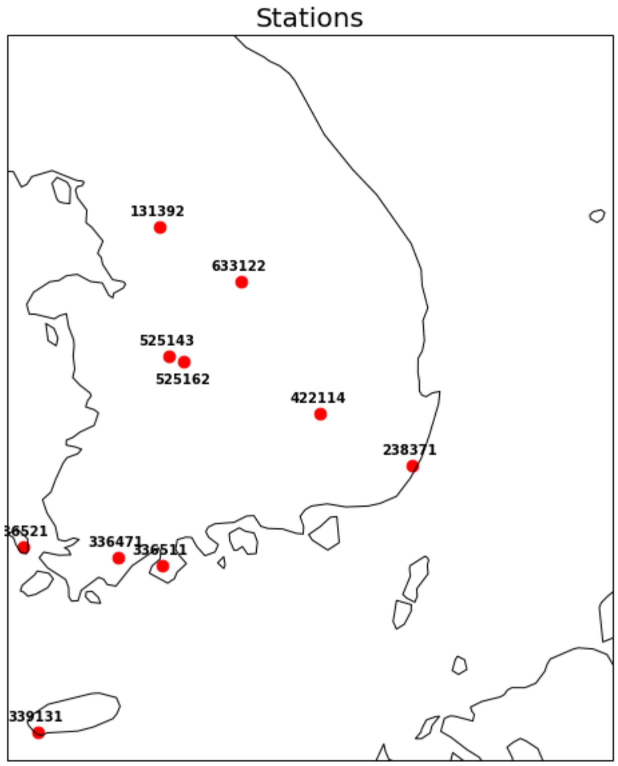


Figure F28: *Location of stations discussed in the study. These represent cases of unusually high concentrations of $PM_{2.5}$ due to measurement errors.*

		Input Feature (Normalized)				Output Target (Observed Meteorology)
		Meteorology Parameter 1	Meteorology Parameter 2	……	ith Meteorology Parameter	
Day 1	Hour 00					
	Hour 01					
	⋮					
	Hour 23					
Day 2	Hour 00					
	Hour 01					
	⋮					
	Hour 23					
……	Hour 00					
	Hour 01					
	⋮					
	Hour 23					
Nth Day	Hour 00					
	Hour 01					
	⋮					
	Hour 23					

a)

	Inputs (Normalised)								Output Target (Observed Hourly Rainfall)			
	Meteorology Parametr 1	Meteorology Parameter 2	……	(i-1)th Meteorology	WRF Simulated Hourly Rainfall							
Hours →	Daily Average	Daily Average	Daily Average	Daily Average	0	1	…	23	0	1	…	23
Day 1												
Day 2												
:												
Nth day												

b)

Figure F29: Model Architecture. *The Rain-RM model has 87 input (n=87) and 24 output (m=24) variables. All other models have 64 input and 1 output variable.* ***a)*** *The setup of inputs and outputs for all models except Rain-RM model; i=64.* ***b)*** *The setup of inputs and outputs for the Rain-RM model; i=64.*

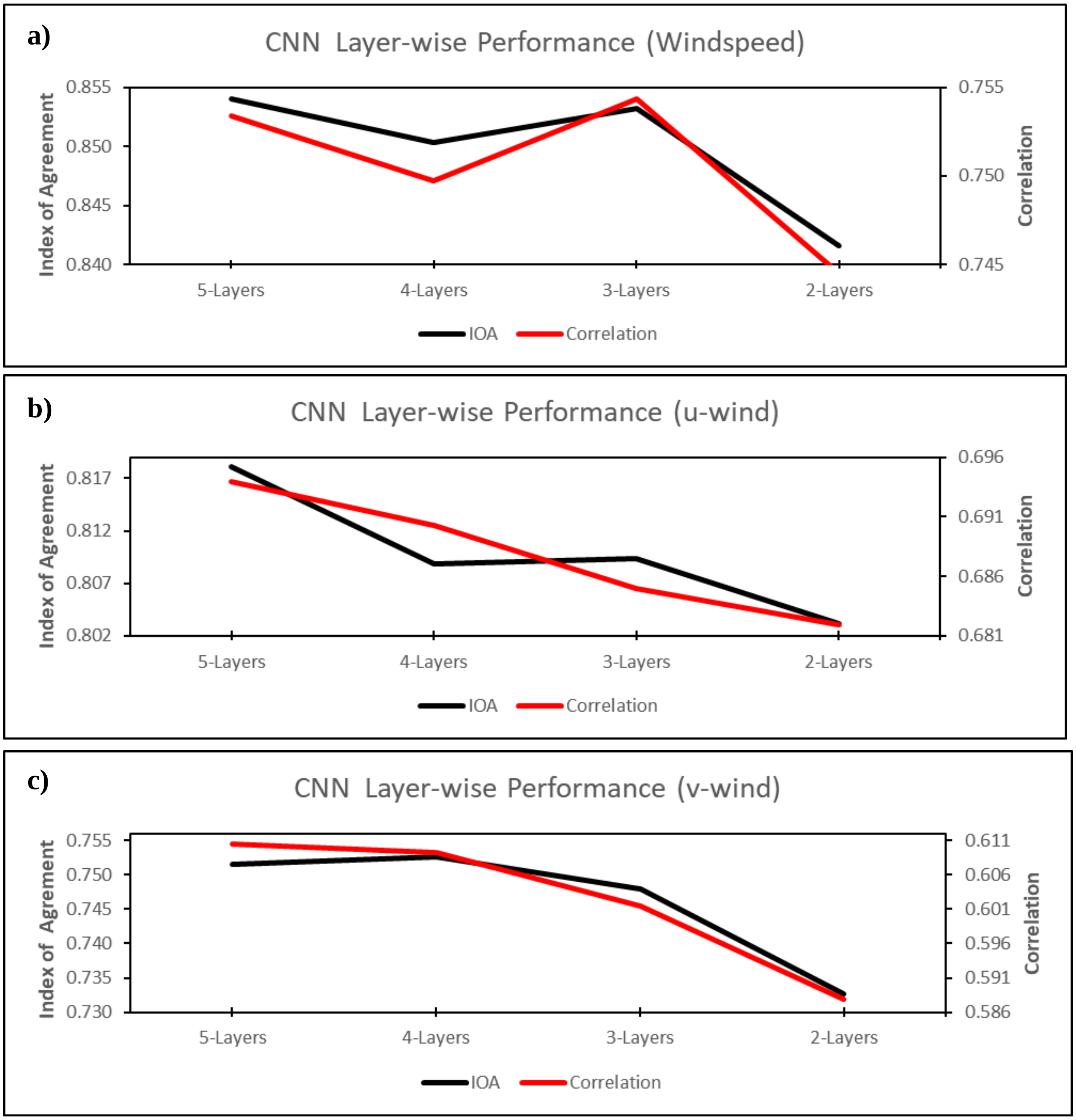


Figure F33: *The performance in terms of IOA and correlation of the CNN model with a different number of layers for a) windspeed, b) u-wind, and c) v-wind. The primary y-axis represents IOA and the secondary y-axis represents correlation, the x-axis represents the number of layers of the CNN model.*

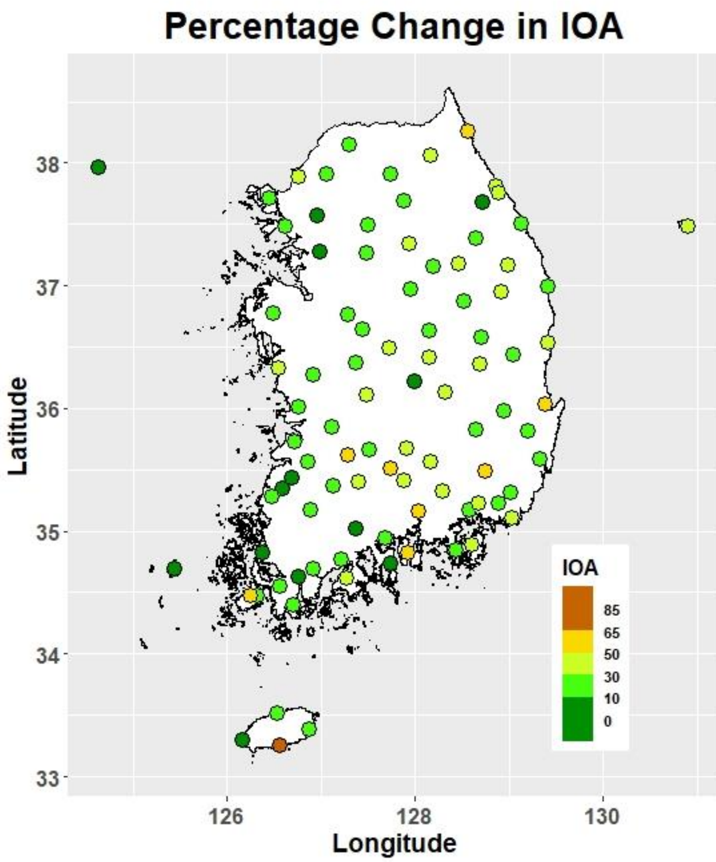


Figure F31. *Percentage change in the IOA from the WRF to the Weather-AI models for wind speed bias-correction.*

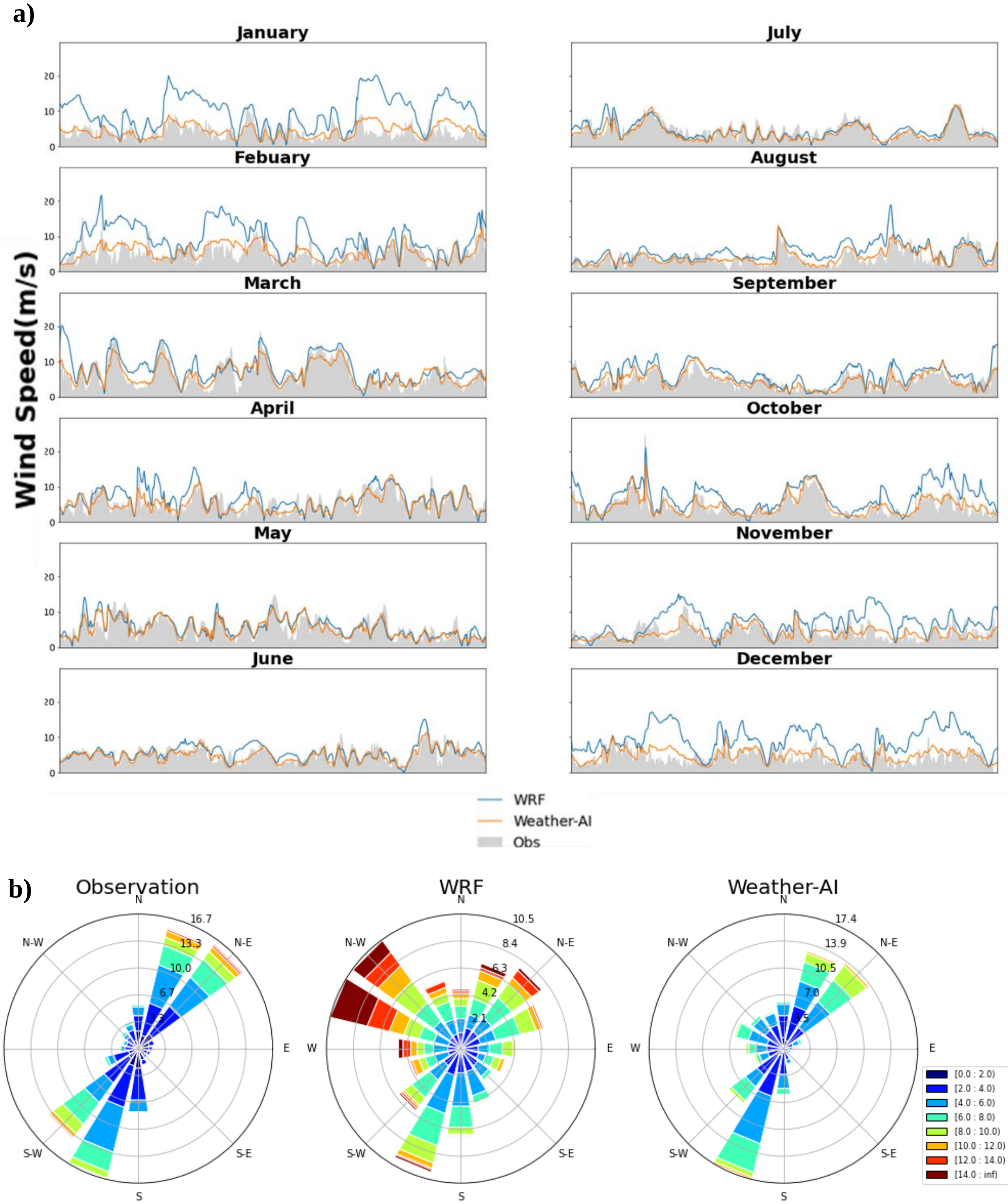
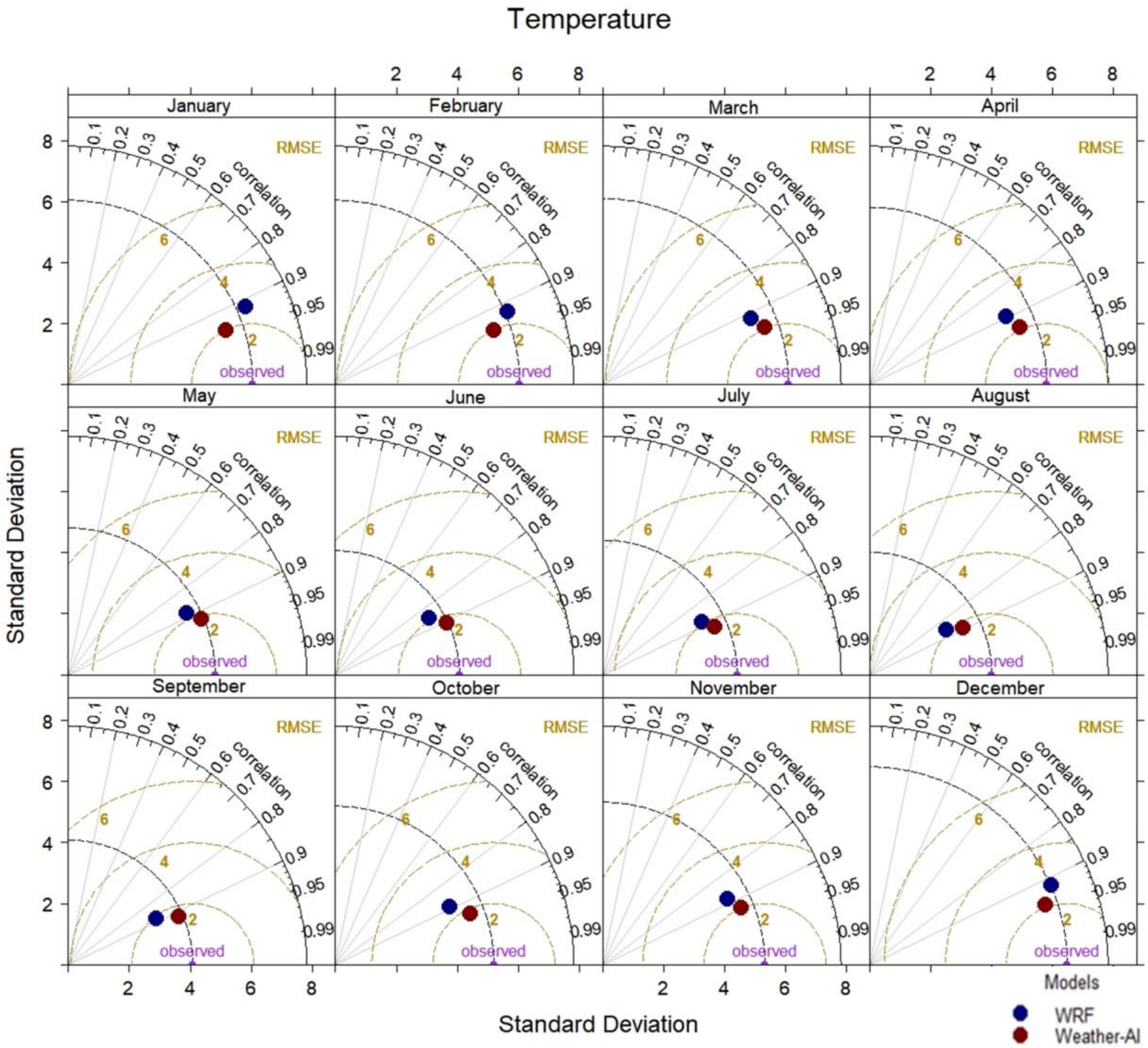
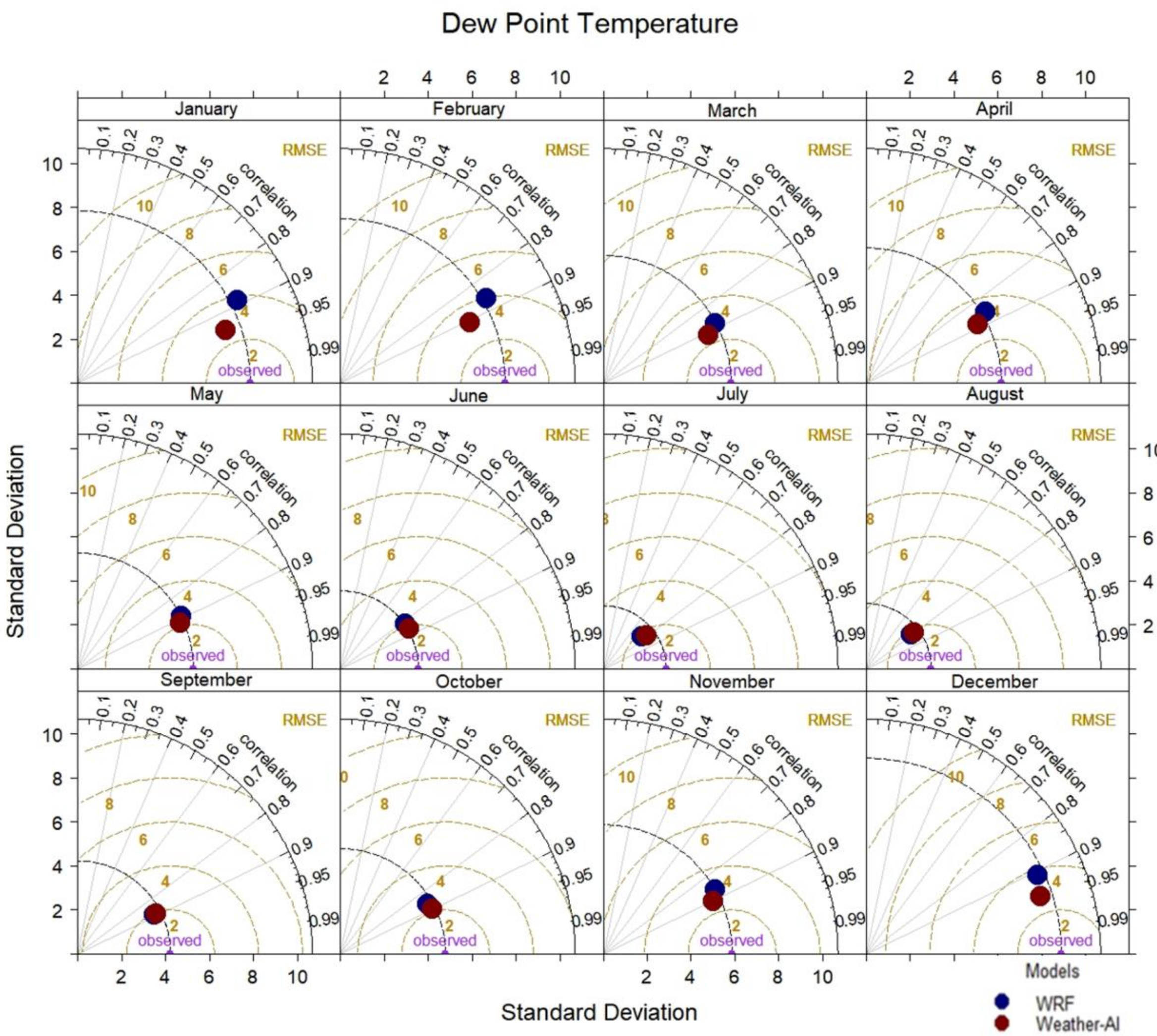


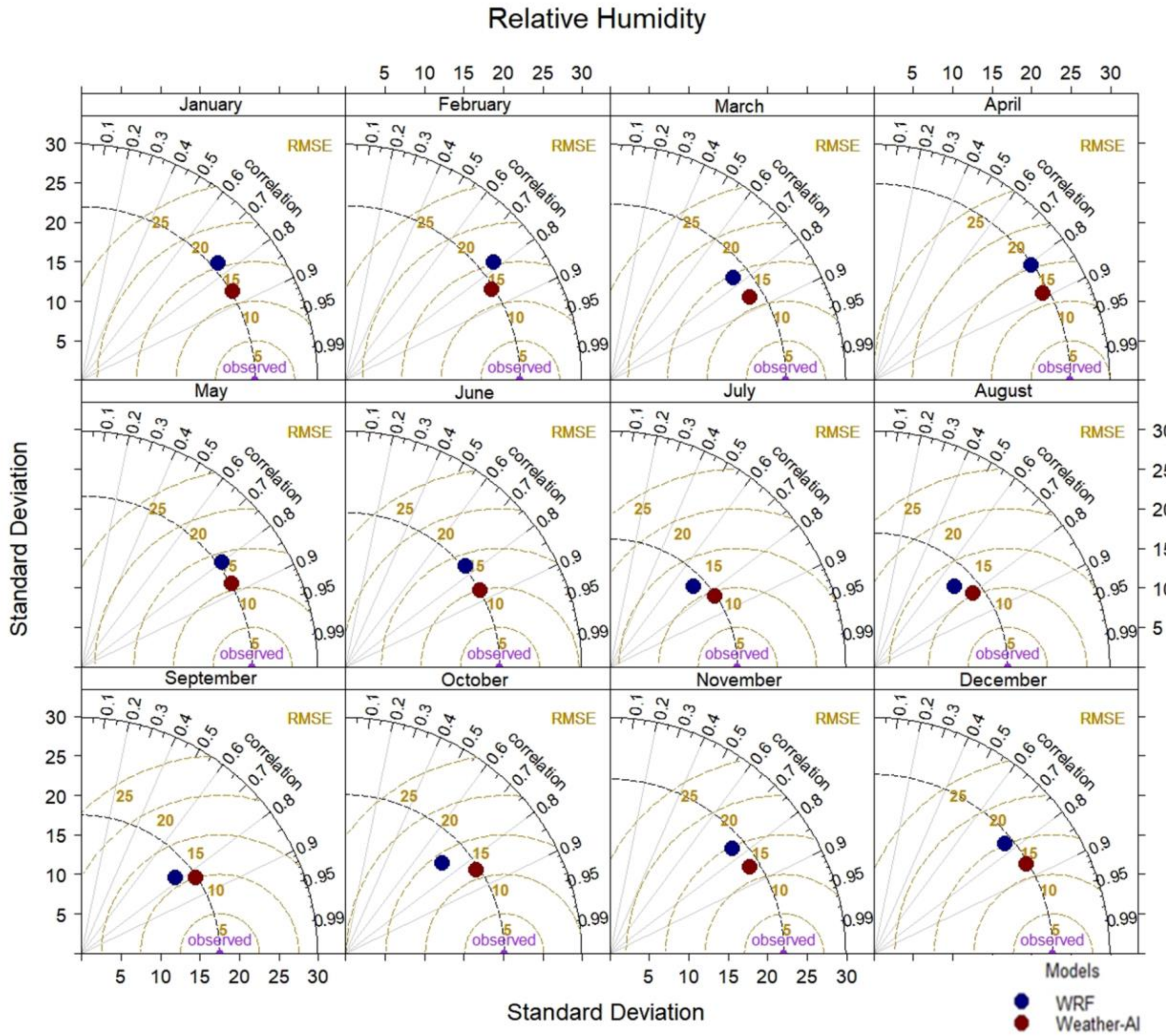
Figure F32. *a) Hourly wind-speed time-series for station 115 for the year 2018. Each subplot represents a month of the year; the X-axis represents hours of the day and the Y-axis the wind speed in m/s. b) Polar plot of hourly wind direction in 2018.*


Temperature
January
February
March
April
May
June
July
August
September
October
November
December
RMSE
correlation
observed
Standard Deviation
Standard Deviation
Models
WRF
Weather-AI

(a)


Dew Point Temperature
January
February
March
April
May
June
July
August
September
October
November
December
RMSE
correlation
observed
Standard Deviation
Models
WRF
Weather-AI

(b)



(c)

Figure F33. *Taylor diagrams comparing the WRF and Weather-AI models for a) temperature, b) dew-point temperature, c) relative humidity for each month in 2018. X- and Y-axis represent the standard deviation of wind speed. The black quarter circular axis represents the correlation. The golden circular axis represents the RMSE.*

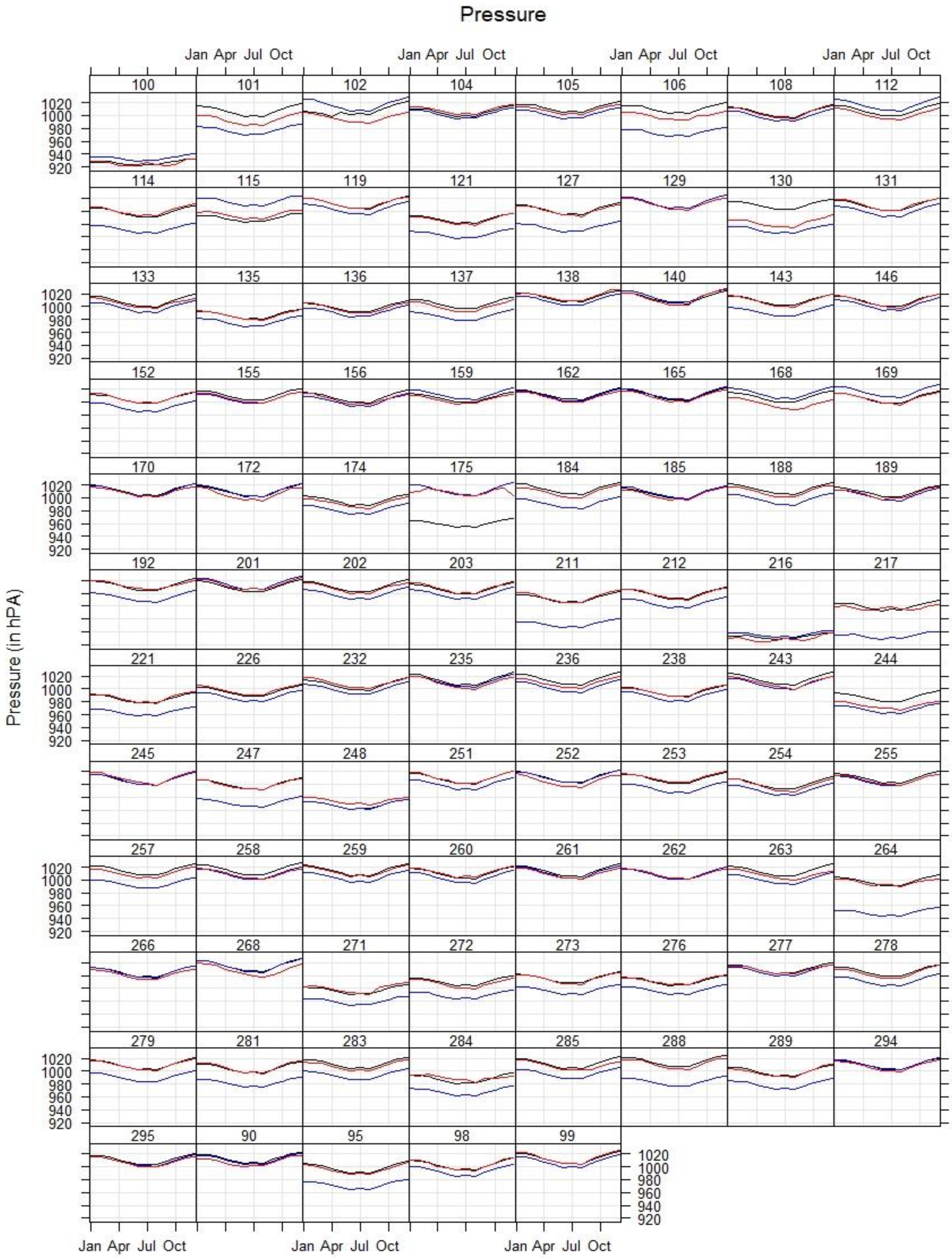


__Figure F34.__ Monthly mean of surface pressure (in hPa) of stations in South Korea. The X-axis represents the months for the year 2018 and Y-axis represents surface pressure.

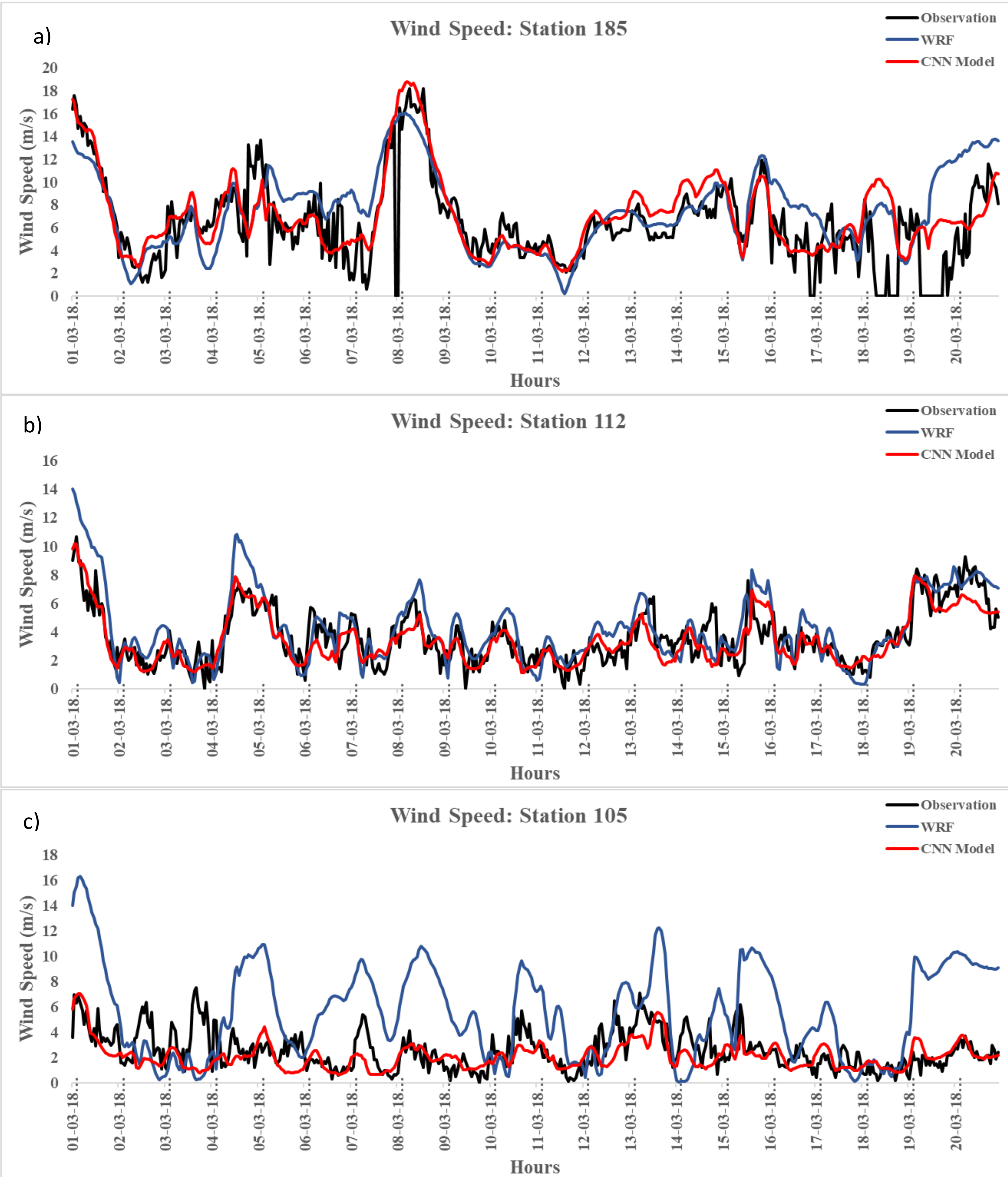


Figure F35. *Sample of time-series of wind-speed simulation, comparison between observations, WRF and CNN. a) Station with best IOA (Weather-AI model); b) Station with the median IOA(Weather-AI model); c) Station with least IOA(Weather-AI model).*

C: Experimental Set-up

Since the model requires each example of all input features in a one-dimensional array, the data obtained from TCEQ was needed to be converted into a format the CNN model can understand. To achieve this, all the meteorologies were first arranged, NOx and ozone in each column and each row having hourly data starting from January 1, 2014, 0000-hour local time (as shown in Table C1).

Table C1. *Input feature Description: Sample Arrangement of all inputs column-wise*

	Wind Speed (miles/hour)	Wind Direction (degrees compass)	Temperature (Fahrenheit)	Relative Humidity (percent)	……	NOx (ppb)	O_3 (ppb)
01-01-14 0:00							
01-01-14 1:00							
: :							
31-01-17 23:00							

Since there are some missing data in the observation, the next step involves the imputation (SOFT-IMPUTE by Mazumder et al. (2010)) of these values. Once imputed values were obtained, each input parameter was normalized between 0 and 1 until the last day before the prediction day. This was done because each feature has a different scale of measurement and to bring each scale to a normalized value (0-1). The process was required so that the model does not differentiate between the different scales of input features.

Now, the data was arranged as input and output features (Table C2). Each row contains a day, and each column contains an hour of the day. For example, the first 24 columns have 24-hour wind speed and the next 24 columns have 24-hour wind direction and so on. Current 24-hour observation data constitute the input features (i.e., i_1, i_2, i_3….i_k from Figure 3). The output features have next 24-hour ozone (this ozone was not normalized). Once the model was trained, it predicted the next day (for example, January 5, 2017) based on input features (i.e., meteorology and concentrations of air pollutants of the previous day (for this case, January 4, 2017)).

Table C2. The architecture of Training set for AI algorithm: An Example to show how input and output features are arranged to train the model. (Input features here are the same as in Figure 3)

	Input Feature (Normalized)																Output Target (Actual Observed Values)			
	Wind Speed				Wind Direction				……				Ozone				Next day Ozone			
Hours (UTC time)	0	1	…	23	0	1	…	23	0	1	…	23	0	1	…	23	0	1	…	23
Figure 3 Notations	i_1	i_2	…	…	…	…	…	…	…	…	…	…	…	…	…	i_k				
02-Jan-14		January 1, 2014															January 2, 2014			
03-Jan-14																				
:																				
n^{th} day		$(n-1)^{th}$ day															n^{th} day			

D: General Statistics

IOA varies between 0 and 1 and indicates the degree of model prediction error. A value of 1 indicates a perfect match and 0 indicates no agreement at all (Willmott et al., 1981). The IOA can be defined as

$$IOA = 1 - \frac{\sum(O_i - P_i)^2}{\sum(abs(O_i - \bar{O}) + abs(P_i - \bar{O}))^2}$$

where O_i and P_i represent the observed and predicted values, respectively. $\bar{O}$ was the mean of observed values for the entire observation sample.

E: Linear Model

The linear autoregressive model was used for station CAMS-003 with the configurations shown in Table E1. Before running the model, the augmented Dickey-Fuller (ADF) null hypothesis (Fuller, W. A., 1976) was tested for unit root and rejected. The rejection of the null hypothesis means that the time series does not have a unit root and was stationary.

In univariate models, only ozone concentrations were used to fit the model, while in the multivariate model, the exogenous variables were previous day meteorology and air pollutants (as in the CNN model). Model 1 to 4 were fitted from 2014 to 2016 hourly data and predicted the entire year of 2017 at once. In models 5 and 6, the model was fitted until the previous day and predicted the next 24 hours only.

Table E1. *A detailed description of each linear model.*

Model No.	Model	Order (p)	Computational Time	IOA
1	**Univariate Auto Regression**	0	4.9 s	1.72×10^{-14}
2	**Univariate Auto Regression**	24	6.55 min	0.037
3	**Multivariate Auto Regression**	0	6.03 s	0.38
4	**Multivariate Auto Regression**	24	19.92 min	0.35
5	**Univariate Auto Regression**	24	1.66 days	0.84
6	**Multivariate Auto Regression**	24	5.05 days	0.84

It was evident from Table E1 that the model needs a daily update of the previous days' data. Although a very good IOA was obtained for models 5 and 6, the time series was shifted 24 hours, as shown in Figure E2. This leads to the conclusion that the model was naïve in the prediction of next-day ozone concentrations.

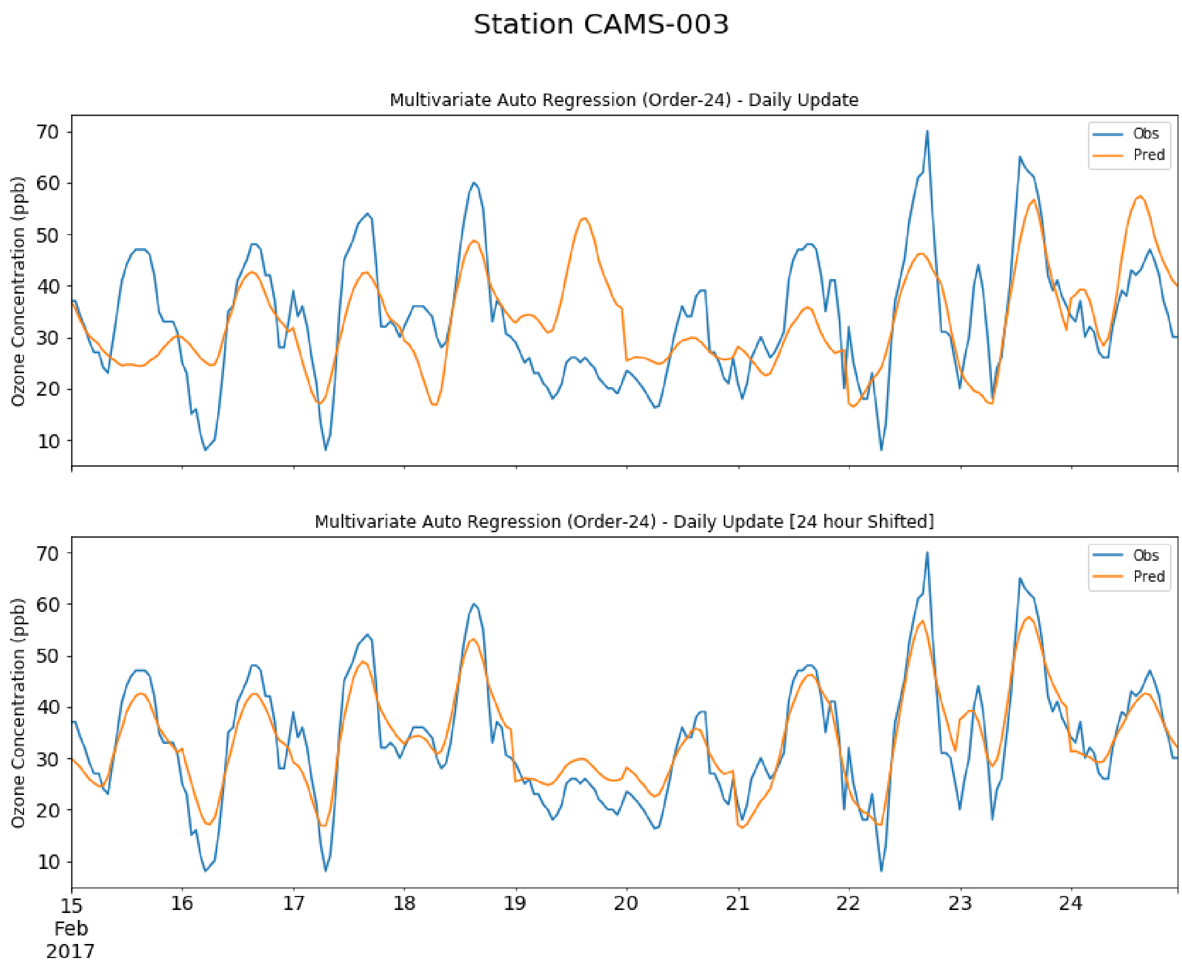


Figure E1: *The effect of shifting in the predicted time-series of multivariate autoregression. The top image was the originally predicted time series. The bottom image was the predicted time series when shifted 24 hours back.*

F: Time Delayed Neural Network

The time-delayed neural network (TDNN) model with configurations shown in Table F1 was used. For TDNN, the input was previous 7-day and 1-day ozone. The model process was fast and had good IOA, but the predicted time-series was 24-hour shifted. Figure F1 shows that the model was copying information from the previous 24-hour and representing it as a result.

Table F1. *Configuration of Time-delayed neural network (TDNN)*

Input	Input context	Computational Time	IOA
Previous 7- day	Layer 1 - [-15,15] Layer 2 - [-12,-12] Layer 3 - [-12,-12] Layer 4 - [-12,-12] Layer 5 - [-12,-12]	82.94 s	0.82
Previous 1-day	Layer 1 - [-15,15] Layer 2 - [-12,-12] Layer 3 - [-12,-12] Layer 4 - [-12,-12] Layer 5 - [-12,-12]	47.22 s	0.88

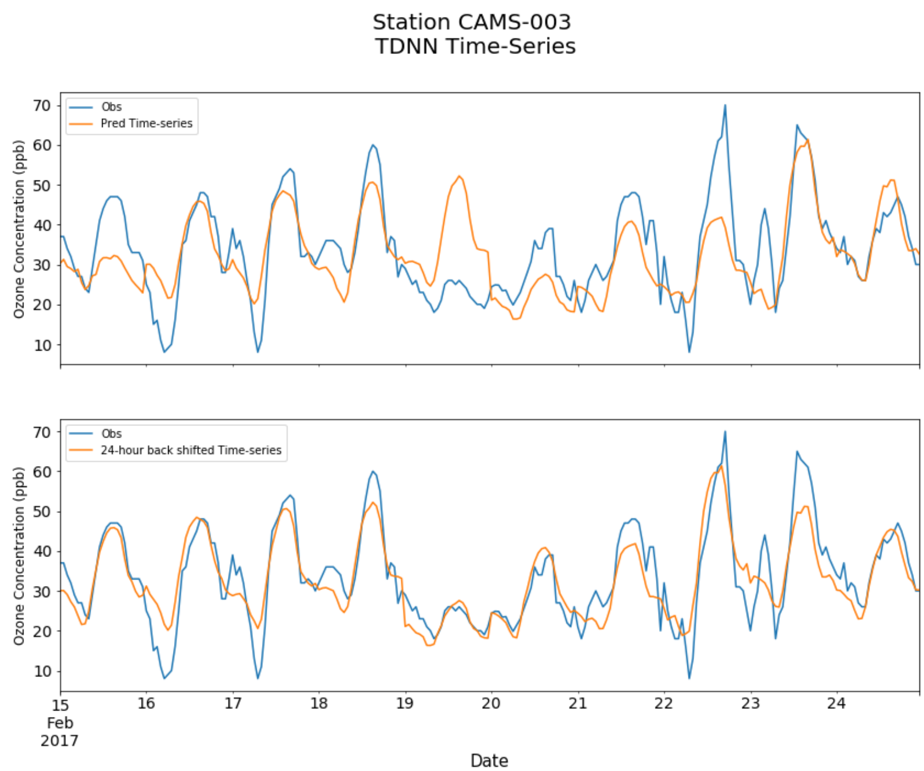


Figure F1. *The effect of shifting in the predicted time-series of time-delayed neural network (TDNN). The top image was the originally predicted time series. The bottom image was the predicted time series when shifted 24 hours back.*

BIBLIOGRAPHY

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: A system for large-scale machine learning 21.

Ambient (outdoor) air pollution [WWW Document], n.d. URL https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (accessed 6.23.21).

Ayres, J.G., Maynard, R.L., Richards, R.J., 2006. Air Pollution And Health. World Scientific.

Baklanov, A., Korsholm, U., Mahura, A., Petersen, C., Gross, A., 2008. ENVIRO-HIRLAM: on-line coupled modelling of urban meteorology and air pollution. Adv. Sci. Res. 2, 41–46.

Behera, S.N., Sharma, M., 2010. Investigating the potential role of ammonia in ion chemistry of fine particulate matter formation for an urban environment. Science of The Total Environment 408, 3569–3575.

Bell, M.L., Peng, R.D., Dominici, F., 2006. The Exposure–Response Curve for Ozone and Risk of Mortality and the Adequacy of Current Ozone Regulations. Environmental Health Perspectives 114, 532–536.

Bengio, Y., 2009. Learning Deep Architectures for AI. FNT in Machine Learning 2, 1–127.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., 2006. Greedy Layer-Wise Training of Deep Networks. Advances in Neural Information Processing Systems 19.

Bézenac, E. de, Pajot, A., Gallinari, P., 2019. Deep learning for physical processes: incorporating prior scientific knowledge*. J. Stat. Mech. 2019, 124009.

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G., Di Carlo, P., 2017. Recursive neural network model for analysis and forecast of PM10 and PM2.5. Atmospheric Pollution Research 8, 652–659.

Biancofiore, F., Verdecchia, M., Di Carlo, P., Tomassetti, B., Aruffo, E., Busilacchio, M., Bianco, S., Di Tommaso, S., Colangeli, C., 2015. Analysis of surface ozone using a recurrent neural network. Science of The Total Environment 514, 379–387.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G.R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J.L., Saide, P.E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., Seigneur, C., 2015. Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. Atmos. Chem. Phys. 15, 5325–5358.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. The Lancet 360, 1233–1242.

Byun, D., Schere, K.L., 2006. Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. Appl. Mech. Rev 59, 51–77.

Byun, D.W., Kim, S.-T., Kim, S.-B., 2007. Evaluation of air quality models for the simulation of a high ozone episode in the Houston metropolitan area. Atmospheric Environment 41, 837–853.

Cassola, F., Burlando, M., 2012. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. Applied Energy 99, 154–166.

Chai, T., Kim, H.-C., Lee, P., Tong, D., Pan, L., Tang, Y., Huang, J., McQueen, J., Tsidulko, M., Stajner, I., 2013. Evaluation of the United States National Air Quality Forecast Capability experimental real-time predictions in 2010 using Air Quality System ozone and NO_2 measurements. Geoscientific Model Development 6, 1831–1850.

Chan, C.H., Caine, E.D., You, S., Yip, P.S.F., 2015. Changes in South Korean urbanicity and suicide rates, 1992 to 2012. BMJ Open 5, e009451.

Chatani, S., Morikawa, T., Nakatsuka, S., Matsunaga, S., Minoura, H., 2011. Development of a framework for a high-resolution, three-dimensional regional air quality simulation and its application to predicting future air quality over Japan. Atmospheric Environment 45, 1383–1393.

Cho, K., van Merrienboer, B., Bahdanau, D., Bengio, Y., 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches, in: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Presented at the Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Association for Computational Linguistics, Doha, Qatar, pp. 103–111.

Choi, Y., Kim, H., Tong, D., Lee, P., 2012. Summertime weekly cycles of observed and modeled NO_x and O_3 concentrations as a function of satellite-derived ozone production sensitivity and land use types over the Continental United States. Atmos. Chem. Phys. 12, 6291–6307.

Choi, Y., Kim, J., Eldering, A., Osterman, G., Yung, Y.L., Gu, Y., Liou, K.N., 2009. Lightning and anthropogenic NOx sources over the United States and the western North Atlantic Ocean: Impact on OLR and radiative effects. Geophysical Research Letters 36.

Choi, Y., Souri, A.H., 2015. Chemical condition and surface ozone in large cities of Texas during the last decade: Observational evidence from OMI, CAMS, and model analysis. Remote Sensing of Environment 168, 90–101.

Chollet, F., others, 2015. Keras.

CMAQ version 5.2beta (February 2017 release) Technical Documentation - CMASWIKI [WWW Document], n.d. URL https://www.airqualitymodeling.org/index.php/CMAQ_version_5.2beta_(February_2017_release)_Technical_Documentation (accessed 2.21.20).

Costoya, X., Rocha, A., Carvalho, D., 2020. Using bias-correction to improve future projections of offshore wind energy resource: A case study on the Iberian Peninsula. Applied Energy 262, 114562.

Create Elegant Data Visualisations Using the Grammar of Graphics [WWW Document], n.d. URL https://ggplot2.tidyverse.org/ (accessed 2.11.21).

Crétat, J., Pohl, B., Richard, Y., Drobinski, P., 2012. Uncertainties in simulating regional climate of Southern Africa: sensitivity to physical parameterizations using WRF. Clim Dyn 38, 613–634.

Czader, B.H., Percell, P., Byun, D., Kim, S., Choi, Y., 2015. Development and evaluation of the Screening Trajectory Ozone Prediction System (STOPS, version 1.0). Geosci. Model Dev. 8, 1383–1394.

Damo, R., Icka, P., n.d. Assessment of Air Pollution Using the Global Pollution Index for City of Korça, Albania□ 1.

Davidson, C.I., Phalen, R.F., Solomon, P.A., 2005. Airborne Particulate Matter and Human Health: A Review. Aerosol Science and Technology 39, 737–749.

de Gouw, J.A., Brock, C.A., Atlas, E.L., Bates, T.S., Fehsenfeld, F.C., Goldan, P.D., Holloway, J.S., Kuster, W.C., Lerner, B.M., Matthew, B.M., Middlebrook, A.M., Onasch, T.B., Peltier, R.E., Quinn, P.K., Senff, C.J., Stohl, A., Sullivan, A.P., Trainer, M., Warneke, C., Weber, R.J., Williams, E.J., 2008. Sources of particulate matter in the northeastern United States in summer: 1. Direct emissions and secondary formation of organic matter in urban plumes. J. Geophys. Res. 113, D08301.

Díaz-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson, J.G., Moncada-Herrera, J.A., 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. Atmospheric Environment 42, 8331–8340.

Eder, B., Kang, D., Mathur, R., Yu, S., Schere, K., 2006. An operational evaluation of the Eta–CMAQ air quality forecast model. Atmospheric Environment 40, 4894–4905.

Erdem, E., Shi, J., 2011. ARMA based approaches for forecasting the tuple of wind speed and direction.

Eslami, E. 1985-, 2020. Applications of Deep Learning in Atmospheric Sciences: Air Quality Forecasting, Post-Processing, and Hurricane Tracking (Thesis).

Eslami, E., Choi, Y., Lops, Y., Sayeed, A., 2019a. A real-time hourly ozone prediction system using deep convolutional neural network. Neural Comput & Applic.

Eslami, E., Choi, Y., Lops, Y., Sayeed, A., Salman, A.K., 2020. Using wavelet transform and dynamic time warping to identify the limitations of the CNN model as an air quality forecasting system. Geoscientific Model Development 13, 6237–6251.

Eslami, E., Salman, A.K., Choi, Y., Sayeed, A., Lops, Y., 2019b. A data ensemble approach for real-time air quality forecasting using extremely randomized trees and deep neural networks. Neural Comput & Applic.

Fernando, H.J.S., Mammarella, M.C., Grandoni, G., Fedele, P., Di Marco, R., Dimitrova, R., Hyde, P., 2012. Forecasting PM10 in metropolitan areas: Efficacy of neural networks. Environmental Pollution 163, 62–67.

Ghahremanloo, M., Lops, Y., Choi, Y., Mousavinezhad, S., 2021. Impact of the COVID-19 outbreak on air pollution levels in East Asia. Science of The Total Environment 754, 142226.

Glorot, X., Bengio, Y., n.d. Understanding the difficulty of training deep feedforward neural networks 8.

Hinton, G.E., 2014, What's wrong with convolutional nets? MIT Brain and Cognitive Sciences – Fall Colloquium Series, URL http://techtv.mit.edu/collections/bcs/videos/30698-what-s-wrong-with-convolutional-nets.

Hinton, G.E., 1989. Connectionist learning procedures. Artificial Intelligence 40, 185–234.

Hodan, W., Barnard, W., 2004. Evaluating the Contribution of PM2.5 Precursor Gases and Re-entrained Road Emissions to Mobile Source PM2.5 Particulate Matter Emissions 58.

Holman, B.P., Lazarus, S.M., Splitt, M.E., 2017. A Fetch-Based Statistical Method to Bias Correct and Downscale Wind Speed over Unresolved Water Bodies. Weather and Forecasting 32, 1637–1657.

Hong, Y.-Y., Satriani, T.R.A., 2020. Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network. Energy 209, 118441.

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., Brasseur, O., 2005. A neural network forecast for daily average PM10 concentrations in Belgium. Atmospheric Environment 39, 3279–3289.

Hoshyaripour, G., Brasseur, G., Andrade, M.F., Gavidia-Calderón, M., Bouarar, I., Ynoue, R.Y., 2016. Prediction of ground-level ozone concentration in São Paulo, Brazil: Deterministic versus statistic models. Atmospheric Environment 145, 365–375.

Hur, S.-K., Ho, C.-H., Kim, J., Oh, H.-R., Koo, Y.-S., 2021. Systematic bias of WRF-CMAQ PM10 simulations for Seoul, Korea. Atmospheric Environment 244, 117904.

Introduction — cartopy 0.18.0 documentation [WWW Document].

Jacobson, M.Z., n.d. Fundamentals of Atmospheric Modeling, Second Edition 829.

Jeon, W., Choi, Y., Percell, P., Souri, A.H., Song, C.-K., Kim, S.-T., Kim, J., 2016. Computationally efficient air quality forecasting tool: implementation ofSTOPS v1.5 model into CMAQ v5.0.2 for a prediction of Asian dust. Geosci. Model Dev. 9, 3671–3684.

Jung, J., Souri, A.H., Wong, D.C., Lee, S., Jeon, W., Kim, J., Choi, Y., 2019. The Impact of the Direct Effect of Aerosols on Meteorology and Air Quality Using Aerosol Optical Depth Assimilation During the KORUS-AQ Campaign. Journal of Geophysical Research: Atmospheres 124, 8303–8319.

Kim, K.-H., Kabir, E., Kabir, S., 2015. A review on the human health impact of airborne particulate matter. Environment International 74, 136–143.

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization.

Kitayama, K., Morino, Y., Yamaji, K., Chatani, S., 2019. Uncertainties in O3 concentrations simulated by CMAQ over Japan using four chemical mechanisms. Atmospheric Environment 198, 448–462.

Kotsakis, A., Choi, Y., Souri, A.H., Jeon, W., Flynn, J., Kotsakis, A., Choi, Y., Souri, A.H., Jeon, W., Flynn, J., 2019. Characterization of Regional Wind Patterns Using Self-Organizing Maps: Impact on Dallas–Fort Worth Long-Term Ozone Trends. Journal of Applied Meteorology and Climatology.

Koulouri, E., Saarikoski, S., Theodosi, C., Markaki, Z., Gerasopoulos, E., Kouvarakis, G., Mäkelä, T., Hillamo, R., Mihalopoulos, N., 2008. Chemical composition and sources of fine and coarse aerosol particles in the Eastern Mediterranean. Atmospheric Environment 42, 6542–6550.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90.

Lawrence, S., Giles, C.L., Ah Chung Tsoi, Back, A.D., 1997. Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Netw. 8, 98–113.

Lecun, Y., Bengio, Y., 1995. Convolutional networks for images, speech, and time-series. The handbook of brain theory and neural networks.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

LeNail, A., 2019. NN-SVG: Publication-Ready Neural Network Architecture Schematics. Journal of Open Source Software 4, 747.

Li, W., Wang, C., Wang, H., Chen, J., Yuan, C., Li, T., Wang, W., Shen, H., Huang, Y., Wang, R., Wang, B., Zhang, Y., Chen, H., Chen, Y., Tang, J., Wang, X., Liu, J., Coveney, R.M., Tao, S., 2014. Distribution of atmospheric particulate matter (PM) in rural field, rural village and urban areas of northern China. Environmental Pollution 185, 134–140.

Li, X., Choi, Y., Czader, B., Roy, A., Kim, H., Lefer, B., Pan, S., 2016. The impact of observation nudging on simulated meteorology and ozone concentrations during DISCOVER-AQ 2013 Texas campaign. Atmos. Chem. Phys. 16, 3127–3144.

Liu, X.-H., Zhang, Y., Cheng, S.-H., Xing, J., Zhang, Q., Streets, D.G., Jang, C., Wang, W.-X., Hao, J.-M., 2010. Understanding of regional air pollution over China using CMAQ, part I performance evaluation and seasonal variation. Atmospheric Environment 44, 2415–2426.

Lops, Y., Choi, Y., Eslami, E., Sayeed, A., 2019. Real-time 7-day forecast of pollen counts using a deep convolutional neural network. Neural Comput & Applic.

Lu, C., Liu, Y., Niu, S., Krueger, S., Wagner, T., 2013. Exploring parameterization for turbulent entrainment-mixing processes in clouds: ENTRAINMENT-MIXING PARAMETERIZATION. J. Geophys. Res. Atmos. 118, 185–194.

Marr, D., 1976. Artificial Intelligence -- A Personal View.

Mazumder, R., Hastie, T., Tibshirani, R., 2010. Spectral Regularization Algorithms for Learning Large Incomplete Matrices 36.

Memarianfard, M., Hatami, A.M., Memarianfard, M., 2017. Artificial neural network forecast application for fine particulate matter concentration using meteorological data. Global Journal of Environmental Science and Management 3, 333–340.

Mills, G., Buse, A., Gimeno, B., Bermejo, V., Holland, M., Emberson, L., Pleijel, H., 2007. A synthesis of AOT40-based response functions and critical levels of ozone for agricultural and horticultural crops. Atmospheric Environment 41, 2630–2643.

Misenis, C., Zhang, Y., 2010. An examination of sensitivity of WRF/Chem predictions to physical parameterizations, horizontal grid spacing, and nesting options. Atmospheric Research 97, 315–334.

Morino, Y., Chatani, S., Hayami, H., Sasaki, K., Mori, Y., Morikawa, T., Ohara, T., Hasegawa, S., Kobayashi, S., 2010. Evaluation of Ensemble Approach for O_3 and $PM_{2.5}$ Simulation. Asian Journal of Atmospheric Environment 4, 150–156.

Mukherjee, A., Agrawal, M., 2017. World air particulate matter: sources, distribution and health effects. Environ Chem Lett 15, 283–309.

Nair, V., Hinton, G.E., 2010. Rectified Linear Units Improve Restricted Boltzmann Machines 8.

Nie, Y., Bo, H., Zhang, W., Zhang, H., 2020. Research on Hybrid Wind Speed Prediction System Based on Artificial Intelligence and Double Prediction Scheme. Complexity 2020, 1–22.

Noxon, J.F., 1978. Tropospheric NO2. Journal of Geophysical Research: Oceans 83, 3051–3057.

Pan, S., Choi, Y., Jeon, W., Roy, A., Westenbarger, D.A., Kim, H.C., 2017. Impact of high-resolution sea surface temperature, emission spikes and wind on simulated surface ozone in Houston, Texas during a high ozone episode. Atmospheric Environment 152, 362–376.

Pandis, S.N. (Ed.), 2004. Atmospheric aerosol processes. Cambridge University Press, Cambridge, UK.

Pleim, J.E., Gilliam, R., 2009. An Indirect Data Assimilation Scheme for Deep Soil Temperature in the Pleim–Xiu Land Surface Model. Journal of Applied Meteorology and Climatology 48, 1362–1376.

Pleim, J.E., Xiu, A., 2003. Development of a Land Surface Model. Part II: Data Assimilation. Journal of Applied Meteorology and Climatology 42, 1811–1822.

Pouyaei, A., Choi, Y., Jung, J., Sadeghi, B., Song, C.H., 2020. Concentration Trajectory Route of Air pollution with an Integrated Lagrangian model (C-TRAIL model v1.0) derived from the Community Multiscale Air Quality Modeling (CMAQ model v5.2). Geoscientific Model Development Discussions 1–30.

Prasad, K., Gorai, A.K., Goyal, P., 2016. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. Atmospheric Environment 128, 246–262.

Rao, S.T., Luo, H., Astitha, M., Hogrefe, C., Garcia, V., Mathur, R., 2020. On the limit to the accuracy of regional-scale air quality models. Atmospheric Chemistry and Physics 20, 1627–1639.

Sadeghi, B., Choi, Y., Yoon, S., Flynn, J., Kotsakis, A., Lee, S., 2020. The characterization of fine particulate matter downwind of Houston: Using integrated factor analysis to identify anthropogenic and natural sources. Environmental Pollution 262, 114345.

Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A.K., Lee, J.-B., Park, H.-J., Choi, M.-H., 2021a. A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. Sci Rep 11, 1–8.

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., Jung, J., 2020a. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. Neural Networks.

Sayeed, A., Choi, Y., Jung, J., Lops, Y., Eslami, E., Salman, A.K., 2020b. A Deep Convolutional Neural Network Model for improving WRF Forecasts. arXiv:2008.06489 [physics].

Sayeed, A., Lops, Y., Choi, Y., Jung, J., Salman, A.K., 2021b. Bias correcting and extending the PM forecast by CMAQ up to 7 days using deep convolutional neural networks. Atmospheric Environment 253, 118376.

Scarpa, G., Gargiulo, M., Mazza, A., Gaetano, R., 2018. A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. Remote Sensing 10, 236.

Schaefer, J.T., 1990. The Critical Success Index as an Indicator of Warning Skill. Weather and Forecasting 5, 570–575.

Shah, A.S., Langrish, J.P., Nair, H., McAllister, D.A., Hunter, A.L., Donaldson, K., Newby, D.E., Mills, N.L., 2013. Global association of air pollution and heart failure: a systematic review and meta-analysis. The Lancet 382, 1039–1048.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W., WOO, W., 2017. Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 5617–5627.

sklearn.model_selection.GridSearchCV — scikit-learn 0.24.1 documentation [WWW Document], n.d. URL https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (accessed 2.2.21).

Statistical Database | KOSIS KOrean Statistical Information Service [WWW Document], n.d. URL https://kosis.kr/eng/statisticsList/statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ETITLE&parmTabId=M_01_01&statId=2008001&themaId=#SelectStatsBoxDiv (accessed 2.11.21).

Stensrud, D.J., 2007. Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models. Cambridge University Press, Cambridge.

Strode, S.A., Ziemke, J.R., Oman, L.D., Lamsal, L.N., Olsen, M.A., Liu, J., 2019. Global changes in the diurnal cycle of surface ozone. Atmospheric Environment 199, 323–333.

Su, Z., Wang, J., Lu, H., Zhao, G., 2014. A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting. Energy Conversion and Management 85, 443–452.

Taylan, O., 2017. Modelling and analysis of ozone concentration by artificial intelligent techniques for estimating air quality. Atmospheric Environment 150, 356–365.

Taylan, O., 2013. Assessing air quality in Jeddah by modeling suspended PM10 concentration. Journal of International Environmental Application and Science 8, 326–335.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106, 7183–7192.

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

Tikhonov, A.N., 1998. Nonlinear Ill-Posed Problems, Applied Mathematical Sciences. Springer Netherlands.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M., 2015. Learning Spatiotemporal Features with 3D Convolutional Networks, in: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 4489–4497.

Trieu, T.T.N., Goto, D., Yashiro, H., Murata, R., Sudo, K., Tomita, H., Satoh, M., Nakajima, T., 2017. Evaluation of summertime surface ozone in Kanto area of Japan using a semi-regional model and observation. Atmospheric Environment 153, 163–181.

US EPA, O., 2016. Particulate Matter (PM) Basics [WWW Document]. US EPA. URL https://www.epa.gov/pm-pollution/particulate-matter-pm-basics (accessed 6.23.21).

van der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Kasibhatla, P.S., Arellano, A.F.J., 2006. Interannual variability in global biomass burning emissions from 1997 to 2004. Atmospheric Chemistry and Physics 6, 3423–3441.

Wang, Z., Zhang, J., Zhang, Y., Huang, C., Wang, L., 2020. Short-Term Wind Speed Forecasting Based on Information of Neighboring Wind Farms. IEEE Access 8, 16760–16770.

Wilgan, K., Rohm, W., Bosy, J., 2015. Multi-observation meteorological and GNSS data comparison with Numerical Weather Prediction model. Atmospheric Research 156, 29–42.

Willmott, C.J., 1981. On the Validation of Models. Physical Geography 2, 184–194.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. Journal of Geophysical Research: Oceans 90, 8995–9005.

Yeo, I., Choi, Y., Lops, Y., Sayeed, A., 2021. Efficient PM2.5 forecasting using geographical correlation based on integrated deep learning algorithms. Neural Computing and Applications.

Yienger, J.J., Levy, H., 1995. Empirical model of global soil-biogenic NO $_{\chi}$ emissions. J. Geophys. Res. 100, 11447.

Zhang, Q., Geng, G., Wang, S., Richter, A., He, K., 2012. Satellite remote sensing of changes in NOxemissions over China during 1996–2010. Chin. Sci. Bull. 57, 2857–2864.

Zhang, R., Wang, G., Guo, S., Zamora, M.L., Ying, Q., Lin, Y., Wang, W., Hu, M., Wang, Y., 2015. Formation of Urban Fine Particulate Matter. Chem. Rev. 115, 3803–3855.