Applications of Deep Learning in Atmospheric Sciences: Air Quality Forecasting, Post-Processing, and Hurricane Tracking

by
Ebrahim Eslami

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: Prof. Yunsoo Choi

Committee Member: Prof. Robert Talbot

Committee Member: Prof. Xun Jaing

Committee Member: Prof. Christoph Eick

University of Houston
May 2020

DEDICATION

This dissertation is dedicated to my wife, Hanie, whose 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, Prof. Yunsoo Choi, for the continuous support of my Ph.D. studies and related research, for his patience, motivation, and immense knowledge. His guidance has helped me in all the time of research and writing of this thesis. I never would 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: Prof. Robert Talbot, Prof. Xun Jaing, and Prof. Christoph Eick, for their insightful comments and encouragement which incented me to widen my research from various perspectives.

I especially thank my wife, Dr. Hanie Joodat. I married the best person out there for me. Hanie has been a true and great supporter and has unconditionally loved 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 neverending support.

I also thank my late dad, mom, and brothers. My hard-working parents have sacrificed their lives for my brothers and myself and provided unconditional love and care. I would not have made it this far without them. I especially thank my oldest brother, Prof. Ali Eslami, and my sister-in-law, Prof. Preetika Kumar, for all their advice and support in my academic and personal life. I know I always have my family to count on when times are rough.

I thank my fellow lab-mates for the sleepless nights we were working together before deadlines, and for all the fun we had in the last four years. In particular, I am grateful to our AI team members, Yannic Lops, Alqamah Sayeed, and Ahmed Khan Salman for the countless stimulating discussions we had regarding various exciting research topics in the field of machine learning. Additionally, I would like to thank Dr. Anirban Roy, Dr. Shuai Pan, Jia jung, Bavand Sadeghi, and Arman Pouyaie for working together. I express my thanks to the other EAS faculty members and graduate students for discussions on different interesting research topics.

Last but not the 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

This study employs deep learning-based models for developing: fast, real-time air quality forecasting systems; a post-processing tool for bias-correcting the chemical transport model; and a reliable hybrid hurricane tracking model. A deep convolutional neural network (CNN) algorithm, which is an advanced deep learning algorithm, was employed to predict the hourly ozone concentrations each day (24 hours in advance) for the entire year using several meteorological variables and air pollution concentrations from the previous day. The CNN model showed a reasonable performance with an average index of agreement (IOA) of 0.84-0.89 and a Pearson correlation coefficient of 0.74-0.81. Although the CNN model successfully captured daily trends of the ozone concentrations, it notably underpredictd high ozone peaks during the summer. To address this issue, six generalized machine leaning ensemble models were developed to regularize low- and high-ozone episodes. By resampling the training dataset based on the daily peaks, the 'best' ensemble model reduced the ozone peak prediction error by 5 to 30 ppb during summer. Another deep CNN model was developed to post-process the results of the Community Multiscale Air Quality (CMAQ) model. The CNN model significantly improved the performance of the CMAQ model by improving its absolute correlation coefficient by 0.16 and reducing its prediction bias by more than 20 ppb on average. To improve the prediction of hurricane models, a novel, hybrid approach was proposed using a CNN and an ensemble Kalman filter (EnKF). First, a three-step CNN ensemble model was developed to predict direction, distance traveled, and intensity of a hurricane. Then, an EnKF was applied as a post-processing step. The results of the hybrid model for 17 tropical storms in 2017 showed statistical advantages over official National Hurricane Center (NHC) 24-hour-ahead forecasts (i.e., ~13% and ~34% improvement in track and intensity forecast biases, respectively).

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Ambient ozone is an essential phytotoxic pollutant. As one of the most harmful secondary air pollutants, surface ozone is mainly formed by the photochemical reactions between nitrogen oxides (NOx) and volatile organic compounds (VOCs) under certain meteorological circumstances (1). With increased awareness of the health effects of ozone, having an accurate system for real-time ozone forecasting can be a significant benefit to public health by identifying adverse impacts of ozone (2-6). In recent years, many research studies have focused on improving air quality models, specifically real-time ozone forecasting (7-16). Real-time air pollution concentrations can be estimated using two different models: deterministic models (e.g., chemical transport models) and statistical methods (e.g., machine learning techniques).

Air quality forecasting is commonly carried out using three-dimensional Eulerian chemical transport models, such as the United States Environmental Protection Agency (U.S. EPA)'s Community Multiscale Air Quality (CMAQ) model (17) to forecast extreme events and take necessary precautions to prevent extensive damage, especially in heavily populated areas. These models often report a significant model-measurement error that results from uncertainties in the treatment of physical processes and require higher runtime (6, 11, 18, 19). Hence, statistical models, which are more computationally efficient, are also currently used for forecasting purposes. They include neural networks, regression methods, the fuzzy logic (FL) method, classification and regression trees (CART), and decision trees (11, 12).

One common class of statistical models is a neural network technique. The most popular of these models are multilayer perceptrons (MLP), recurrent neural networks (RNN), stacked autoencoders (SAE), and convolutional neural networks (CNN) (20-22). There are two general neural network models in terms of their computational complexity: shallow and deep neural networks. The difference between these networks is their depth, i.e., the number of 'hidden' computational layers. These techniques have been incorporated into multiple approaches for air quality forecasting (23-29). Certain concerns remain regarding the performance of such methods. For one, they require the determination of the optimum network structure (e.g., the number of hidden layers/units, input variables). In theory, nonlinear physics can be approximated using a hidden layer with a large number of units (21, 22), which can lead to "overfitting", necessitating the use of an adaptive structure (21). These methods also require a proper input dataset using suitable initial weights for accelerating learning convergence and avoiding stoppage at local error minima. Because of these limitations, the aforementioned "shallow" neural networks produce substandard results, necessitating the use of "deep" neural networks (30). Introduced by Hinton et al. (31), deep learning is a subfield of machine learning and neural networks, which consists of several hidden computational layers. Using these hidden layers, deep learning utilizes a hierarchical level of artificial neural networks to carry out the process of learning from a large, nonlinear dataset. Deep learning allows machines to solve complex problems even when using a dataset that is very diverse, unstructured, and inter-connected.

To use deep neural networks, Hinton et al. (31) developed a deep belief net (DBN) based on a fast, deep learning algorithm. In DBN, a layer-wise unsupervised learning algorithm is first used to pre-train the initial weights (32, 33). These weights are then fine-tuned by a global supervised learning approach with which more accurate architectures of hidden layers can be used for data-driven forecasting (34, 35). Previous studies (35-37) have indicated that for a set of nonlinear input variables, deep learning algorithms have reported higher forecasting precision than the conventional neural networks and the traditional autoregressive integrated moving average (ARIMA) time series model. The deep learning algorithm has been implemented in numerous applications in various fields such as finance, biology, and physics (20, 38). Kuremoto et al. (36) used a three-layer deep network to forecast a sample time-series. Hraskoa et al. (39) implemented a hybrid neural network to predict time series obtained from three databases. Li et al. (40) designed four deep learning models using feed-forward, energy-based, and recurrent architectures to predict the torsion angle of proteins. All of their models resulted in comparable accuracy.

For the application of the deep learning approach to weather and air pollution forecasting, Zhang et al. (29, 41) applied a deep learning algorithm to predict short-term wind speed for up to two hours at one location. Li et al. (29) proposed a deep learning architecture for predicting PM2.5 concentrations up to 12 hours using only previous concentrations. Despite achieving a relatively high prediction accuracy, the average mean absolute error of their predictions was around 9 $\mu g/m^3$, compared to the average concentration of 83 $\mu g/m^3$. Li et al. (42) used long short-term memory (LSTM) along with several machine learning techniques for the same application. Their results, however, showed a significant error for more than a four-hour prediction. Li et al. (43) applied a deep learning approach to improve the estimation of surface PM2.5 using satellite data and surface observation.

In recent years, the convolutional neural network (CNN) (44), which was widely used in this study, has been acknowledged as the most successful and widely used deep learning approach (20). The CNN is a biologically inspired, multistage architecture composed of convolutional, pooling, and fully-connected layers that can be efficiently trained in a completely supervised manner. The key attribute of this model is the use of multiple processing units that can yield an effective representation of local data features. The deep architecture allows the stacking of multiple layers of these processing units that enable the characterization of data properties over several scales. Thus, the features extracted by the CNN are task-dependent and "non-handcrafted"; meaning there is no need to use an external feature detection algorithm.

In order to obtain better understanding of capabilities of deep learning in the field of atmospheric sciences, I developed different modeling approaches for three exclusive applications during my Ph.D. studies: real-time air quality forecasting, post-processing the chemical transport models, and hurricane tracking. The rest of this dissertation is organized as follows: Chapter 2 describes the CNN model and its characteristics. Chapter 3 employs a CNN model as a real-time ozone forecasting system in Seoul, South Korea. In Chapter 4, several generalized machine learning-based data ensemble techniques are proposed to address several issues with the CNN model discussed in Chapter 3. Chapter 5 proposes a new generation of post-processing method for chemical transport models, in particular CMAQ, using a CNN model. The advantages of using such a model as an ozone forecasting system in the United States is discussed. Chapter 6 employs the CNN ensemble modeling approach and ensemble Kalman filter to introduce a hybrid hurricane forecasting model for the north Atlantic tropical cyclones. Finally, Chapter 7 concludes the findings of this research study and discusses a number of potential benefits of the current work in the field of atmospheric sciences.

CHAPTER 2. METHODS

2.1 Machine learning

Machine learning, particularly deep neural networks and CNNs, has been at the foundation of this study. A general background information of CNN model being used in this study is described in this chapter. The specific modeling configuration is expressed based on each application being discussed here in this study. A general introduction to machine learning is explained, and then an explanation is provided on how CNNs works for regression problems (e.g., forecasting model).

Machine learning (ML) is a subfield of artificial intelligence (AI) which is becoming increasingly more popular, and is widely used out in the industry to solve various tasks. Currently, there is a significant interest for applying ML algorithms in the field of atmospheric sciences, especially in air quality forecasting, remote sensing data retrieval, hurricane tracking, etc. (45). ML is a technique in development of data-driven algorithms that learn to mimic human behavior on the basis of prior example or experience. Machine learning is often considered as a tool for increasing the performance of systems for coping with knowledge-intensive problems in complex domains (46). This happens in a way of learning that involves gathering information from a training dataset for purposefully detecting a pattern in a certain logic. Thus, the fundamental goal of ML models is to generalize such a knowledge detection beyond the examples in the training set. The generalization by ML models provides a scope of improvement in vast variety of physical applications (47). The growing interest in applying machine learning is evidenced by the rapid increase in scientific publications in this area, illustrated in Figure 2.1.

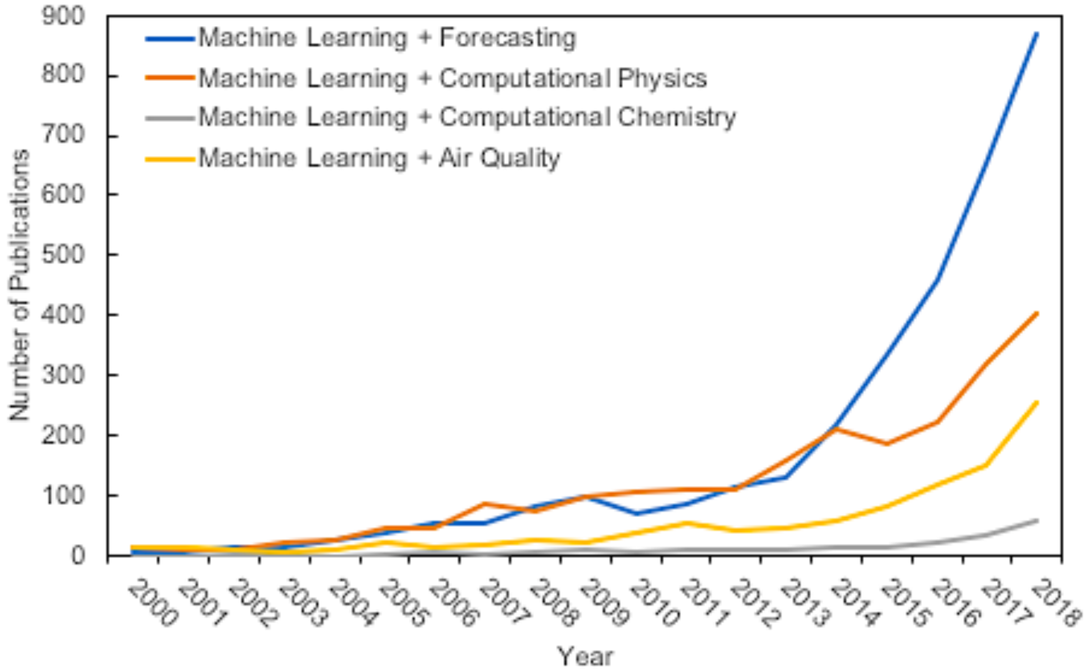


Figure 2.1. Number of machine learning-related publications in different topics from the Web of Science database.

ML has proven to be a valuable tool as an advanced forecasting model and now is a successful part of several data-driven physical approaches. Several methods including neural networks, regression methods, classification and regression trees (CART), and Fuzzy Logic Method (FL) (11, 12) have been employed in this space. Among which, neural networks are most commonly used method in both prediction and classification problems.

Fully-connected neural networks, shown in Figure 2.2, consist of a system of simple interconnected nodes or ‘neurons’ within a set of ‘hidden’ layers. These neurons are representing a nonlinear mapping between an input (array) and an output (array). An input layer distributes input variables to the next layer, which is a hidden layer. Each unit in the hidden layer sums its inputs, then processes then by applying a transfer function (i.e., linear or nonlinear activation function). The results of such process are then distributed the next layer (hidden or output layer). Due to this direction of processing (from input to hidden layer to output layer), the fully-connected neural network is also known as a feed-forward neural network. The superposition of many small linear or nonlinear processes enables a fully-connected network to model extremely non-linear relationships (22). Due to its easily computed derivative a commonly used activation function is the Rectified Linear Unit (33), shown in Figure 2.3.

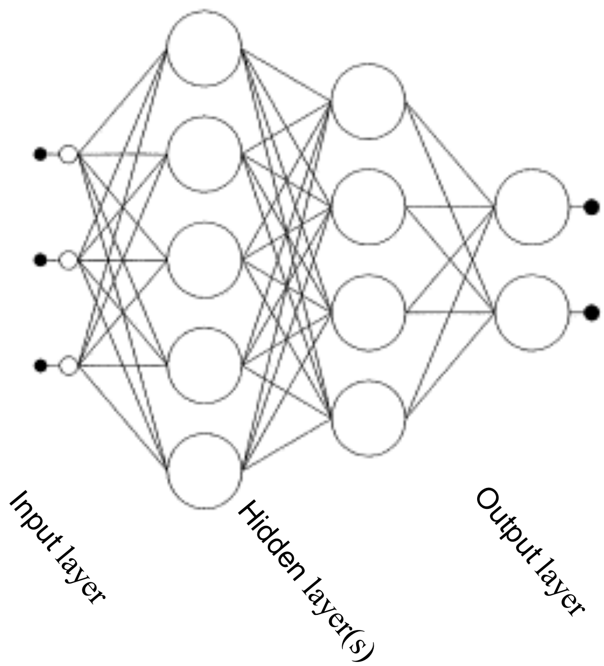


Figure 2.2. A fully-connected neural network with two hidden layers.

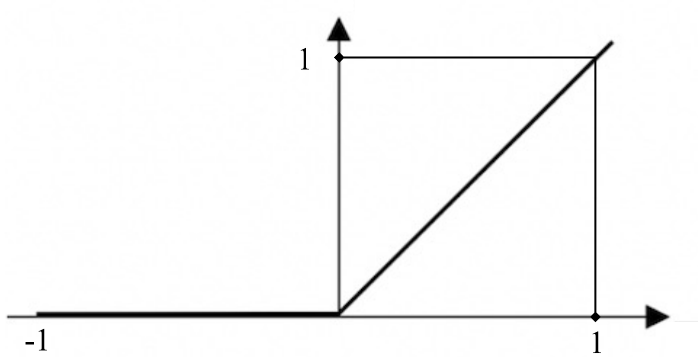


Figure 2.3. Rectified Linear Unit.

2.2 Deep Learning

The backpropagation algorithm and gradient-based optimization technique are widely used for training neural networks (22). Deeper networks (with more computation layers) with large initial weights usually lead to poor local minima. Those with small initial weights, however, produce shallow gradients in the later layers, which decrease the applicability of training networks with numerous hidden layers (31). To resolve this issue, Bengio et al. (32) used a greedy layer-wise learning technique to train deep networks effectively. As the training strategy for this technique, the first layer learned the simpler concepts, and then the next layer learned more abstract features using the feature representation provided by the previous layer. Hence, the objective was to train the deep network layer-by-layer and use the backpropagation algorithm to fine-tune all of the network parameters (32). This is the basis different between the 'deep' and 'shallow' neural networks.

Common deep learning architecture for regression problems are the fully-connected network (i.e., multilayer perceptrons, or MLP) and LSTM. The MLP uses the feedforward approach with backpropagation. Unfortunately, this approach results in inaccurate predictions of global error minima and consumes a great deal of computational time when dealing with large size of input variables and nonlinearity (48). In addition, as hidden units can be very inefficient, especially in networks with multiple hidden layers, the learning process of earlier hidden layers could be significantly slower than that of latter layers (32). The LSTM, on the other land, is a type of recurrent neural network and is designed to recognize patterns in sequences of data, such as times series (49, 50). Since LSTM uses sequential processing over time steps, it has a temporal dimension which is essential in a robust time-series prediction. However, LSTM is a complicated network, requires a tremendous trial for the fine-tuning process, and is computationally expensive (see Section 3.3). Because of these issues, this study proposed using a deep CNN architecture as its selected deep learning algorithm.

2.3 CNNs

Unlike other methods, CNNs, capable of joint feature and classifier learning, can achieve greater classification accuracy on large-scale datasets (44). Deep CNNs can be trained to approximate smooth, highly nonlinear functions, rendering them appropriate for forecasting nonlinear processes related to air quality (51, 52). In addition, feature extraction using deep learning algorithms is more efficient than using other neural network methods, particularly when multiple hidden layers are structured. Although numerous studies have applied CNN, very few have used it for regression problems (51, 52).

A schematic for the deep CNN used in this paper appears in Figure 2.4. The figure shows that the CNN algorithm (53) has the input layer receives the time series of all input variables that are normalized to avoid a steep cost function. Each unit of a layer receives inputs from a set of units located in a small neighborhood in the previous layer. With local receptive fields, neurons can extract elementary features of inputs that are then combined with those of the higher layers. The outputs of such a set of neurons constitute a feature map (see Figure 2.4). At each position, various types of units in different feature maps compute different types of features. A sequential implementation of this procedure for each feature map would be used to scan the input data with a single neuron with a local receptive field and to store the states of this neuron at corresponding locations in the feature map. The constrained units in a feature map perform the same operation on different instances in a time series, and several feature maps (with different weight vectors) can comprise one convolutional layer; thus, multiple features can be extracted at each instance (53). Once a feature is detected, its exact "location" becomes less important as long as its approximate position relative to the other features is preserved (54).

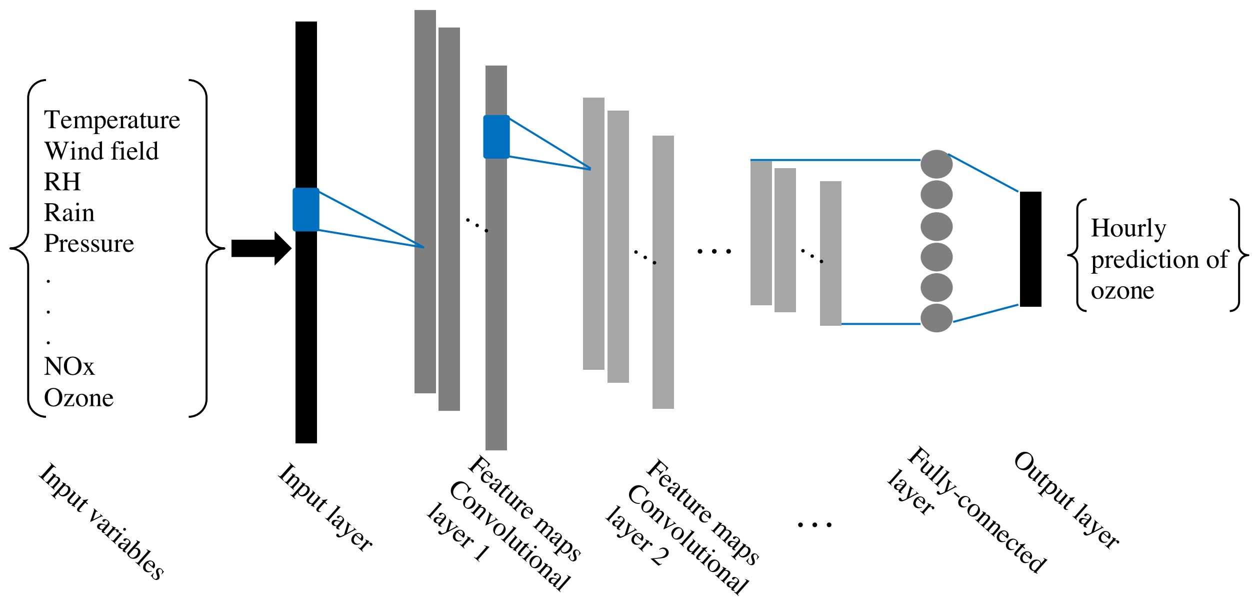


Figure 2.4. Schematic of the deep CNN in our approach for a real-time ozone forecasting system.

The CNN captures changes in the temporal variation of the input data by sweeping through time series using a kernel of a given size. The various sections of the data are represented by feature maps. An additional layer performs a local averaging, called "pooling," and a subsampling reduces the resolution of the feature map and the sensitivity of the output to possible shifts and distortions. This step could potentially discard important information (e.g., sudden ozone peaks). Hence, this study uses the convolution layer without pooling. The feature maps are connected to a fully-connected layer, which helps map each feature for multiple inputs to hourly ozone output (see Figure 2.4).

The deep CNN model is a common deep learning architecture with a long history in numerous applications. Unlike other methods, the CNN model, which is capable of joint feature and classifier learning, and achieves greater accuracy with large-scale datasets. In addition, the use of deep learning algorithms for feature extraction is more efficient than the use of other neural network methods, particularly when multiple hidden layers are structured. Unlike conventional machine learning methods, deep CNNs are capable of automatically identifying the most informative required features, which facilitates predictions from input/output information of the air quality forecasting system.

Compared to fully connected MLPs that have been extensively used as regression models, CNNs are attractive for several reasons. Firstly, MLPs are not explicitly designed to model variance within an ozone concentration that results from a complex interaction between several inputs. While MLPs of sufficient size could indeed capture invariance, they require large networks with a large training set. On the other hand, CNNs are more suitable for small-scale datasets than MLPs because they generally use a smaller number of parameters compared to fully-connected MLPs. Since input can be presented in any order without affecting the performance of the network, MLPs ignore input topology (53). However, temporal variations of the ozone concentration have strong correlations, and modeling these local correlations with Unlike the CNN model used in this study, recurrent neural networks (RNN)—the most common algorithms for time-series forecasting— urged several challenges for the air quality forecasting purpose. As Chapter 3 discussed, their implementation is challenging and their cost is prohibitively high; nevertheless, they offer no significant benefits in terms of accuracy (45, 52).

CHAPTER 3. A REAL-TIME HOURLY OZONE PREDICTION SYSTEM USING A DEEP CONVOLUTIONAL NEURAL NETWORK

3.1 Introduction

To protect citizens from unhealthy air, real-time air quality forecasting systems are required to forecast the concentrations of pollutants of special health concern such as O_3, and PM (2). Such forecast will be used to issue early air quality alerts that allow governments and individuals to take precautionary measures to reduce air pollution and avoid or limit their exposures to unhealthy levels of air pollution (11). Accurate forecasting system can, therefore, offer tremendous societal and economic benefits by enabling advanced planning for individuals, organizations, and communities in order to reduce pollutant emissions and their adverse health impacts.

There are two major air quality forecasting modeling approaches: chemical transport models (CTMs), and statistical models. Despite their physical advantages (taking into account chemical and meteorological processes) over statistical models, chemical transport models (CTM) are limited for real-time air pollution forecasting in several ways. For one, they require a specific computational configuration to deal with all of the nonlinearities in the complex atmosphere, resulting in prolonged computational time. CTM models also encounter significant uncertainties in input fields as well as in the models themselves (e.g., uncertain emissions, simplified physical processes, limited access to measurement data and meteorological fields) (1, 55, 56).

An alternative method for forecasting pollutant concentrations in real-time is statistical methods. Parametric or non-parametric statistical methods for air quality are based on the fact that weather and air quality variables are statistically related. Commonly-used statistical methods are NNs, which use simplified mathematical models to mimic human's brain to enable a structure to find a pattern in the data. NNs usually require a large quantity of historical measured data under a variety of atmospheric conditions (e.g., several years of observed O_3 concentrations). They are generally more suitable for descriptions of complex site-specific relationships between concentrations of air pollutants and potential predictors. When a specific geographical location or city is considered, NNs often have higher accuracy than deterministic models (25, 45).

Conventional neural networks have several common drawbacks. First, they cannot properly capture the contribution of a weather-dependent sources that are important in the formation of a secondary pollutant (e.g., O_3) since they only take 'local' relationship into account. Second, they are unable to address the temporal relations between the time-series of a weather and air pollution variables. This is important when the predicting multiple hours (e.g., 24 hours) is required (e.g., operational real-time, air quality forecasting system). Third, in operational level, they are either unreliable in capturing high air quality episodes (e.g., decision tree models), or computationally inefficient in predicting such events (e.g., LSTM model).

Given the computational efficiency of CNN in processing complex data, a CNN model can be potentially used as a real-time air quality forecasting system. However, real-time hourly ozone forecasting is challenging because of its nonlinear chemistry and the highly varying and complex behavior of the atmosphere. Since ozone chemistry is significantly influenced by meteorology, which adds to the complexity of the problem, providing high prediction accuracy using conventional methods such as chemical transport modeling is challenging. This study introduces a deep learning technique for predicting hourly ozone concentrations for the entire year of 2017 over the city of Seoul, South Korea. Ozone is a secondary pollutant formed by reactions between primary pollutants such as NO_2. In Seoul, these pollutants are emitted by various sources and in various locations that are influenced by industry, automobiles, and biogenic sources (57, 58). This study uses a deep convolutional neural network to predict hourly ozone concentrations based on previous day observations of species and meteorological variables over multiple locations in Seoul.

3.2 Materials and Methods

To model the ozone concentration time-series, we used several predictors including hourly observed values of O_3 and NO_x concentrations (as recorded by South Korea's National Institute of Environmental Research, or NIER), surface temperature, relative humidity, wind speed, and direction, dew-point temperature, surface pressure, and precipitation (as recorded by the Korean Meteorological Administration, or KMA). All of these parameters were measured each hour in the Seoul area since 2014. Thus, each of these nine input parameters had a size of 24, representing the previous day's hourly values. All available data in 25 air quality monitoring stations and one meteorological station was used to construct the input and target dataset for training the CNN model. The location of air and meteorology stations can be found in Figure 3.1.

This study used predictive deep learning techniques to forecast hourly surface ozone concentrations for the year 2017 and selected historical surface measurement data from 2014 to 2016 for training the model. Such a training period provides a broad history to fit a relationship between input variables and ozone concentration. For treating the missing observation data, SOFT-IMPUTE by Mazumder et al. (59) was applied to the raw measured data. SOFT-IMPUTE is a missing data treatment approach that iteratively replaces missing elements with those obtained from a soft-thresholded singular-value decomposition by taking all available data (spatially and temporally) into account.

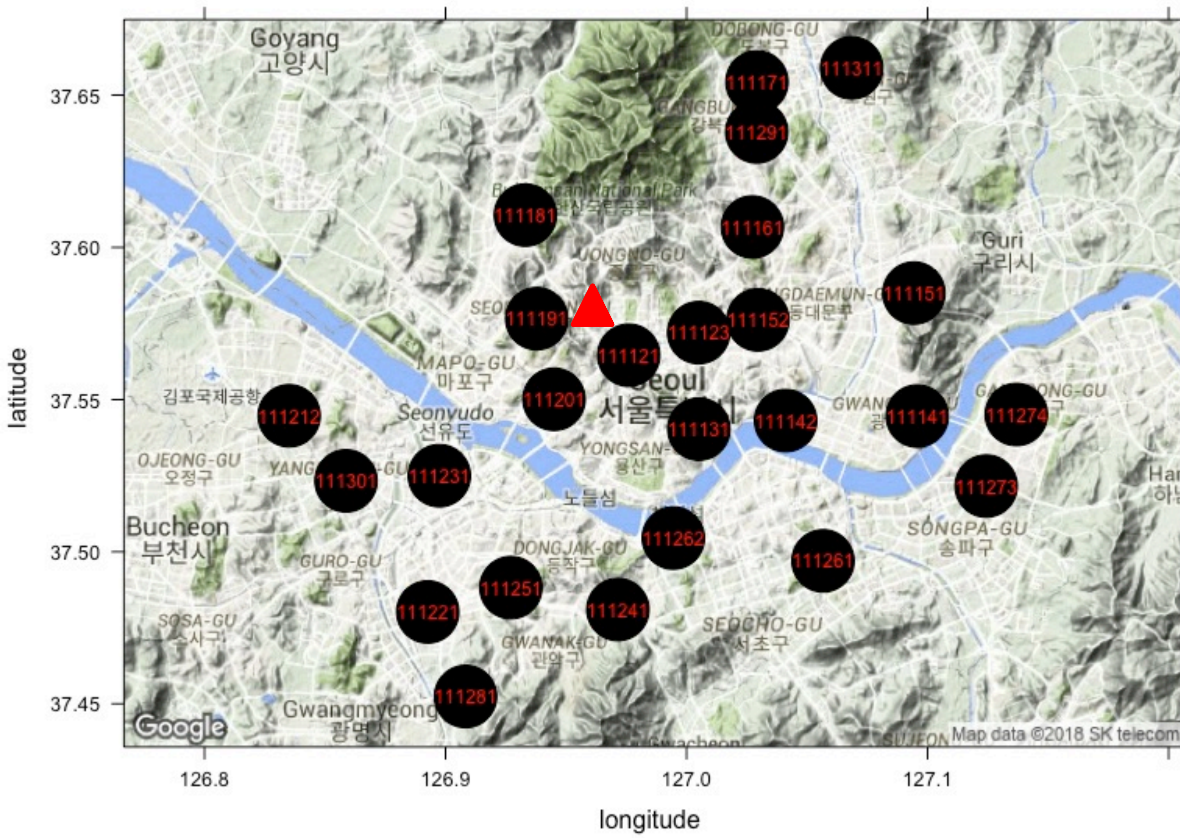


Figure 3.1. Location of the ozone monitoring network by managed by Korea's NIER. The red triangle is the location of meteorology station managed by KMA.

I used a deep convolutional neural network with five convolutional layers, followed by a fully-connected layer before the output layer. Each convolutional layer is characterized by 32 filters and a kernel size of 2, while the fully-connected layer features 256 hidden units. I used a rectified linear unit (ReLU) as the activation function in each layer applied to the normalized input data (since ReLU only passes values greater than zero). The algorithm was implemented in the Keras environment with the TensorFlow backend (60). For a given station, I predicted each day's ozone concentrations in 2017 based on the observations from the previous day. At the end of the day of forecasting, we modified the weight matrix. The CNN algorithm is able to adjust its weight matrix through the backpropagation process with introducing a small modification in weights or bias (a sharp change might lead to the instability of the learning process). I used 80% of the randomly selected data samples for training the model, and the remaining 20% for the validation process; the ratio was the result of a trial/error experience within the model configuration tuning. After each epoch (an epoch is one complete presentation of the sample dataset to train a machine learning model), I monitored the performance of the model to make sure that the model stopped training at a minimum validation loss to avoid the possibility of overfitting.

3.3 Results and Discussion

3.3.1 General statistical analysis

Table 3.1 compares the model-measurement statistics for all of the NIER stations in this study: the index of agreement (IOA), the correlation coefficient (r), the mean bias (MB), the mean absolute error (MAE), and the root mean square error (RMSE). IOA is a standardized measure of the degree of model prediction error and varies between 0 and 1. A value of 1 indicates a perfect match, and 0 indicates no agreement at all. The IOA can be defined as

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

where O_i and P_i represents the observed and predicted values, respectively. $\bar{O}$ is the mean of observed values for the entire observation samples.

The results of the deep learning technique demonstrate acceptable accuracy for the real-time prediction of the hourly ozone concentration for all stations. The yearly averaged IOA and r values of the CNN model were 0.87 and 0.79, respectively, while the values varied between 0.84 and 0.89 for the IOA and 0.74 and 0.81 for r. The yearly averaged MB, MAE, and RMSE were about 1 ppb, 9 ppb, and 12 ppb, respectively. This indicates that the CNN model, overall, underestimated the ozone concentration while missing very high-peak ozone episodes. Only three of the stations showed slightly positive MB (less than 0.3 ppb) while around half experienced negative MB of more than 1 ppb. The reason behind the relatively high MB is that the CNN model is unable to capture high-peak ozone concentrations specifically during the summertime.

Table 3.1. Statistical analysis of the CNN ozone forecasting system for the entire year of 2017.

Station ID	IOA*	r*	MB* (ppb)	MAE* (ppb)	RSME* (ppb)
111121	0.85	0.76	-1.41	9.90	13.30
111123	0.85	0.77	-2.56	9.56	12.91
111131	0.86	0.77	-0.45	7.56	10.40
111141	0.86	0.78	-1.45	9.64	12.80
111142	0.87	0.77	-0.38	7.86	10.78
111151	0.87	0.78	-0.83	7.96	10.67
111152	0.88	0.80	-0.57	7.44	10.07
111161	0.86	0.77	-1.38	8.34	11.29
111171	0.87	0.79	-0.53	9.17	12.23
111181	0.84	0.76	-2.23	9.86	13.38
111191	0.85	0.74	0.19	8.76	11.43
111201	0.87	0.78	-0.37	9.12	12.22
111212	0.88	0.81	-1.31	9.71	13.34
111221	0.89	0.81	-0.49	7.84	10.52
111231	0.87	0.81	-0.78	8.80	12.28
111241	0.88	0.80	-1.50	9.28	12.67
111251	0.89	0.81	-0.31	8.65	11.69
111261	0.87	0.78	0.31	7.49	10.12
111262	0.87	0.80	-1.38	8.87	11.88
111273	0.88	0.80	-0.57	9.22	12.43
111274	0.88	0.80	-1.10	8.89	12.05
111281	0.88	0.80	-1.39	9.17	12.35
111291	0.85	0.76	-2.18	10.93	14.46
111301	0.88	0.79	0.30	8.08	11.04
111311	0.87	0.79	-1.12	10.48	13.99
Average	0.87	0.79	-0.94	8.90	12.01

* IOA is the index of agreement, r is the Pearson correlation coefficient, MB is the mean bias, MAE is the mean absolute error, and RMSE is the root mean squared error.

The CNN model generally predicted the ozone concentrations of the stations located south of the Han River with higher IOA and r (see the map plots of each statistical analysis in Figure 3.2). The topology of the region south of the river is typically flat while that north of the region has a few elevated areas; thus, ozone formation in the southern region is more closely related to the variability of meteorological parameters than the northern region. The CNN model, therefore, is able to map a more accurate function to predict the ozone concentration in the southern region. In addition, the dominant wind direction is from the West and the Southwest (see the monthly pollution-rose diagram of Figure 3.3), both of which are influenced by the Yellow Sea. This dominant pattern results in a meteorology-dependent condition of the ozone concentrations in the southern part of the river and more variable ozone concentrations within the northern region.

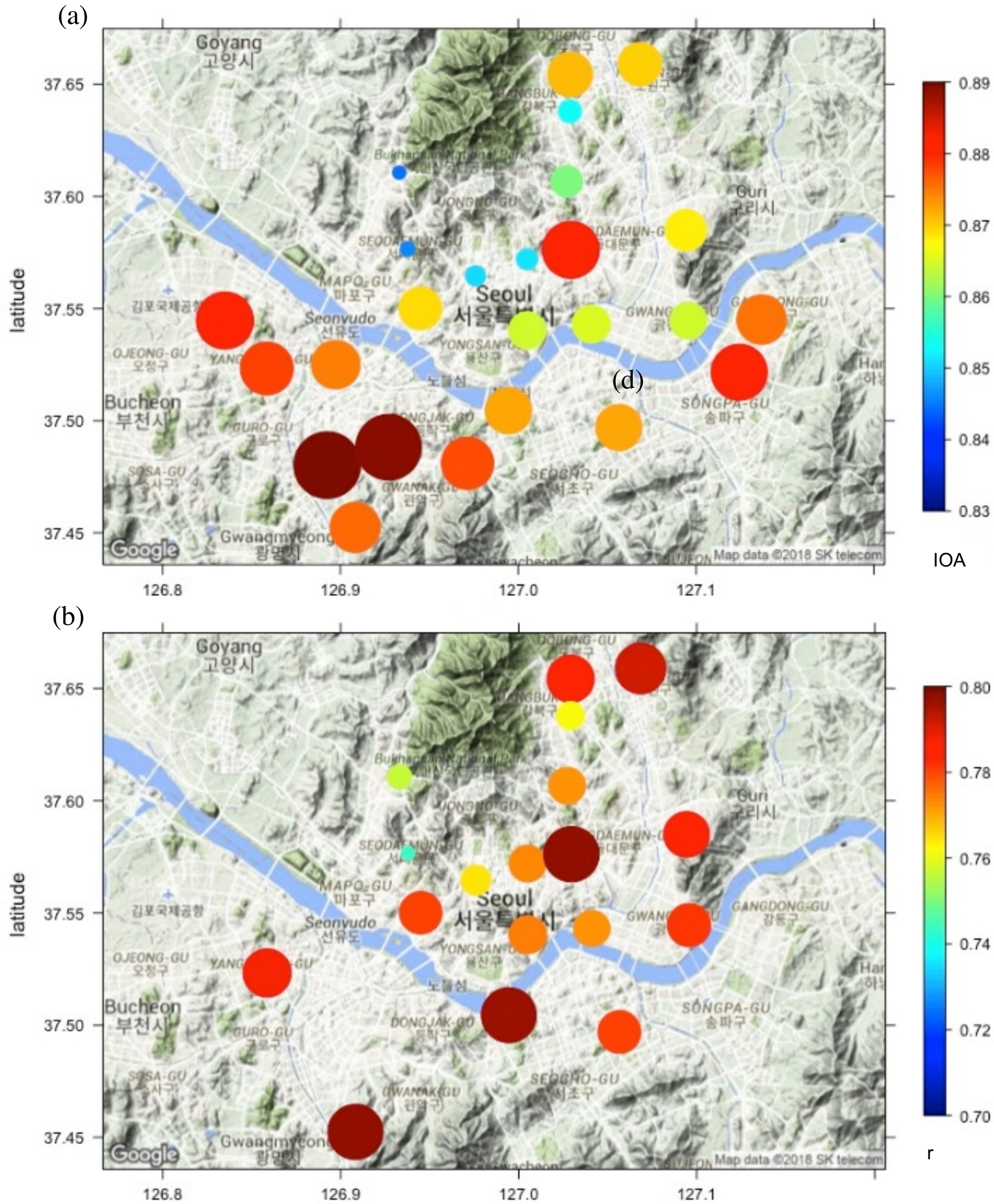


Figure 3.2. Spatial distribution of (a) Index of Agreement (IOA), (b) Pearson correlation coefficient (r) of CNN prediction for NIER stations in Seoul.

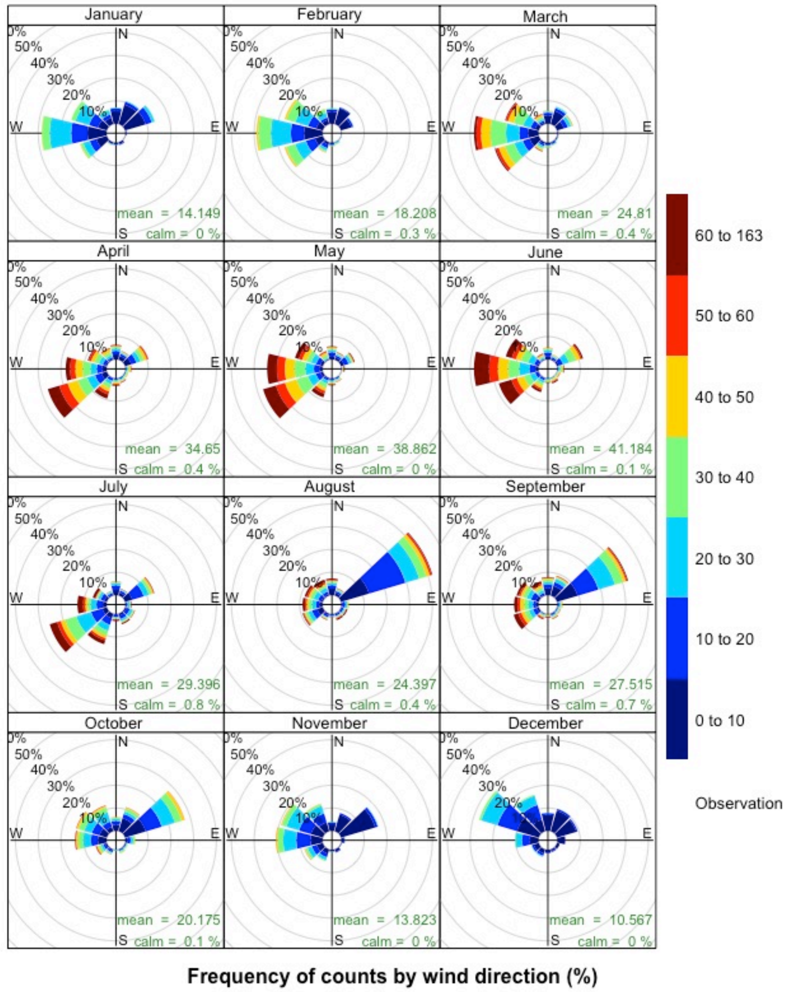


Figure 3.3. Pollution rose of ozone observation in a different month of the year 2017 averaged over all Seoul stations.

Figure 3.4 represents the daily prediction bias in different months of all 25 stations. It is observed that the model bias varied more widely during the warm months (June-September) with more outliers. One explanation for this finding is that the wind pattern changed during these months (see Figure 3.3) with relatively hot and humid days and occasional precipitation events. Most of the precipitation in Seoul occurs during the summer monsoon period between June and September, as a part of the East Asian monsoon season (see Figure 3.5 showing weekly percipitation level in 2017). Variable wind patterns along with scattered rain showers account for uncertainty in the CNN predictions during these months, leading to a larger bias range. Another explanation is that daytime ozone concentrations differed significantly from nighttime concentrations (see Figure 3.6 showing the difference between daytime and nighttime ozone in different months of 2017). This large difference caused a gap in the training process of the CNN model as it had less time to adjust the training process under such circumstances.

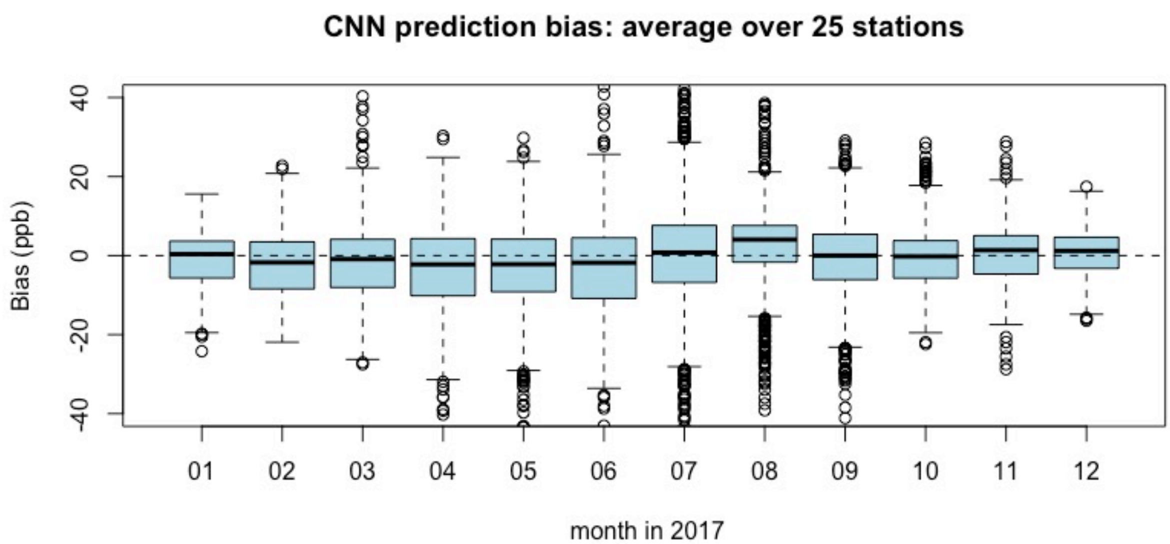


Figure 3.4. Difference between the ozone observation and the results of CNN ozone prediction systems for all 25 stations in Seoul.

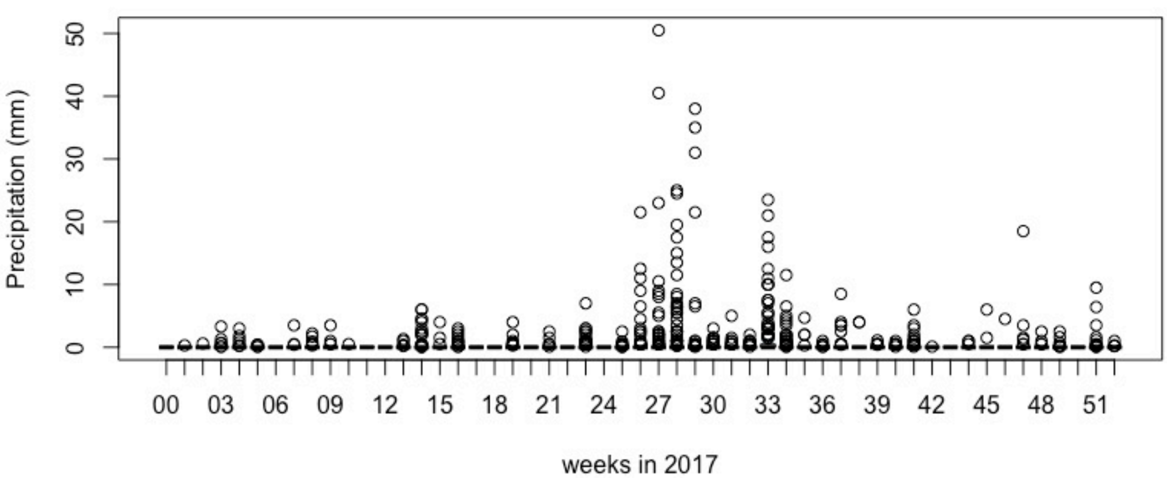


Figure 3.5. Precipitation levels in different weeks of the year 2017.

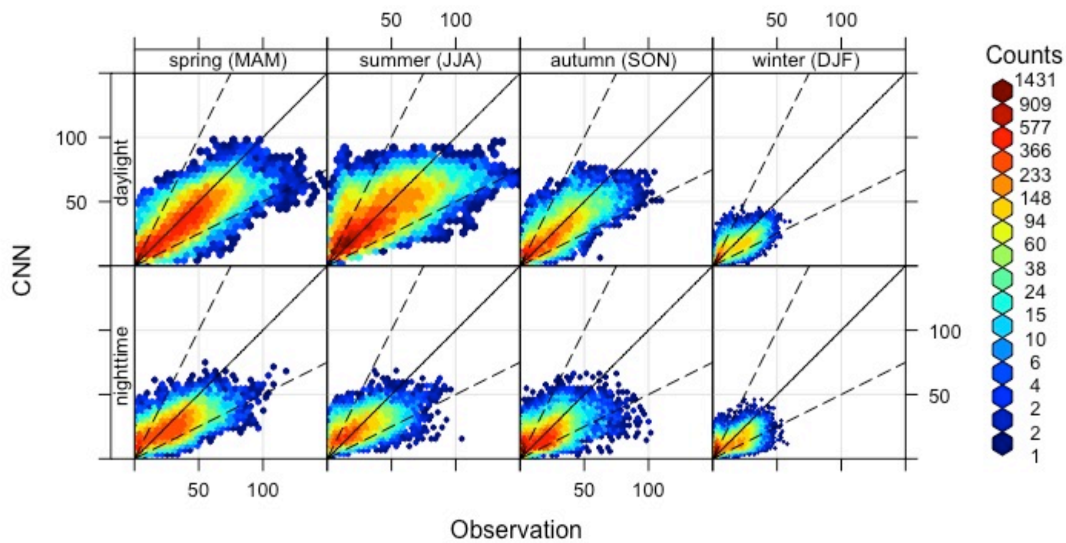


Figure 3.6. Scatter plot of comparison between CNN prediction and observation in different seasons and during day and night averaged over all Seoul stations.

3.3.2 CNN model prediction performance analysis

The daily variations of the IOA and r of the CNN forecasting system are shown in Figure 3.7. The figure shows that the variability of both parameters was higher during the cold months as well as in July. During the winter, most of the input parameters, the ozone concentrations, and the weights inside the CNN model that were trained to capture these extreme examples are at their lowest levels. During the cold months of the winter, the low (or near minimum) values in the weight matrix accounted for variable prediction performance. For July, however, the reason behind the variable prediction performances was the higher ozone concentrations during that month. Maximum concentrations in nearly one-third of the days in July 2017 exceeded 90 ppb level – the highest of all months in 2017.

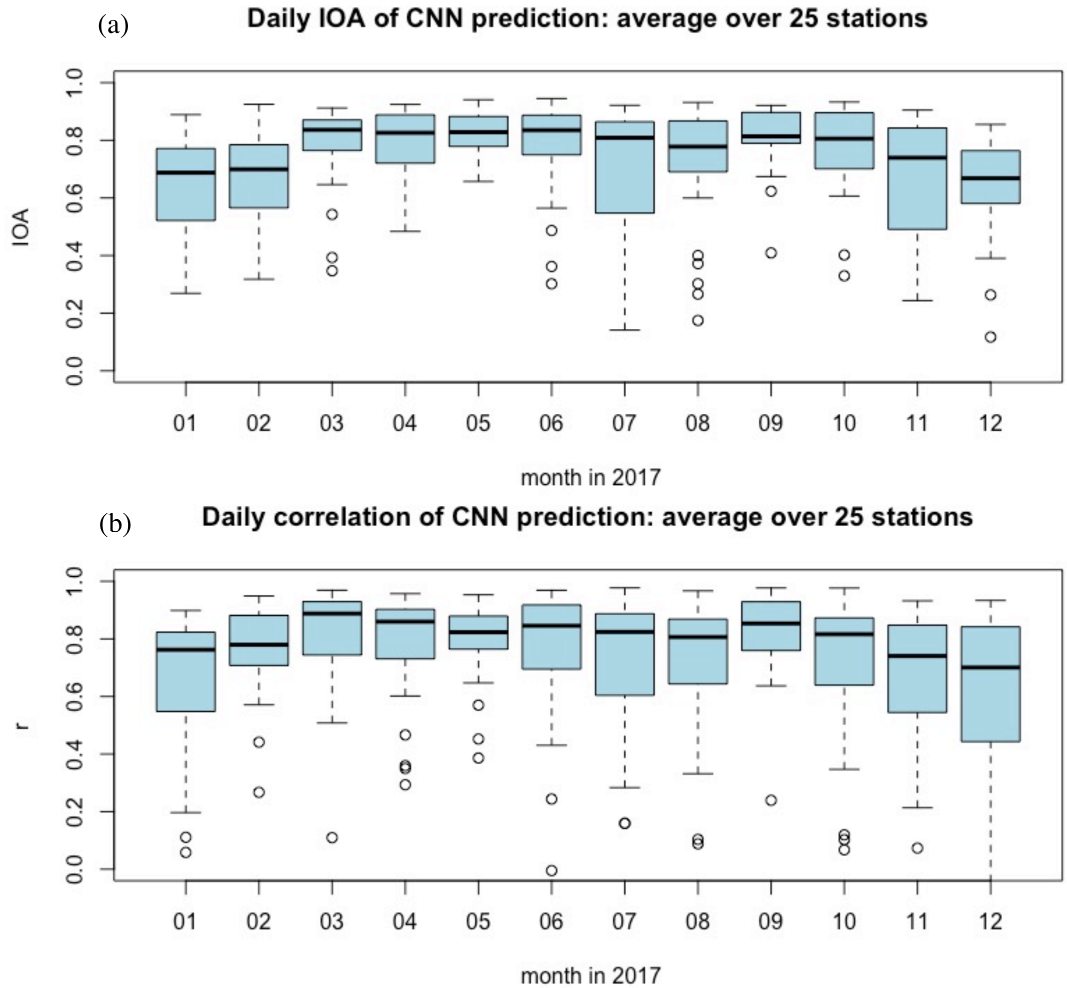


Figure 3.7. (a) Daily index of agreement and (b) the correlation coefficient of CNN forecasting systems averaged over the 25 stations in Seoul, South Korea.

Figure 3.8. shows the time-series comparison of daily mean and daily maximum ozone concentrations between the observed and predicted CNN values. The CNN successfully captured a trend in ozone for the entire year for all stations despite the noticeably different ozone trends (see Figure 3.8(a)). For example, the ozone concentration at stations 111131 (the third station from top left) and 111311 (the last station at the bottom right) followed different trends throughout the year 2017. Nevertheless, the CNN model predicted acceptable, similar trends for both stations (see Table 3.1). The CNN model, however, significantly underpredicted

the maximum daily ozone concentrations during high ozone months (see Figure 3.8(a)). The model also "mispredicted" (predicting high maximum ozone concentrations even though observations revealed relatively low concentrations, and vice versa) concentrations on many days during the warm months because of the rapid change in weather conditions. Since we trained the CNN model based on the information of the previous day, the model was unable to capture the aforementioned change. Nevertheless, for the training for the CNN model, we did not use several important meteorological parameters such as cloud fraction and solar radiation, which could represent the mentioned weather changes; continuous measurements of these parameters were unavailable for the city of Seoul for the time period of this study.

The results of the monthly IOA for all NIER stations is shown in Figure 3.9. They show that the model was generally consistent in its predictions throughout the year. For each station, however, model performance varied from month-to-month. The reason for this variation was the availability of only one meteorological station (KMA station #108) in the city of Seoul for use as input, indicating that the meteorological inputs of the model were the same for all stations, resulting in variability in the prediction performance. In addition, we used only one ozone precursor (NO_2) as the emission representation for the input parameters, indicating that the CNN model had to rely on the quality of meteorological data from one meteorology station for the entire city.

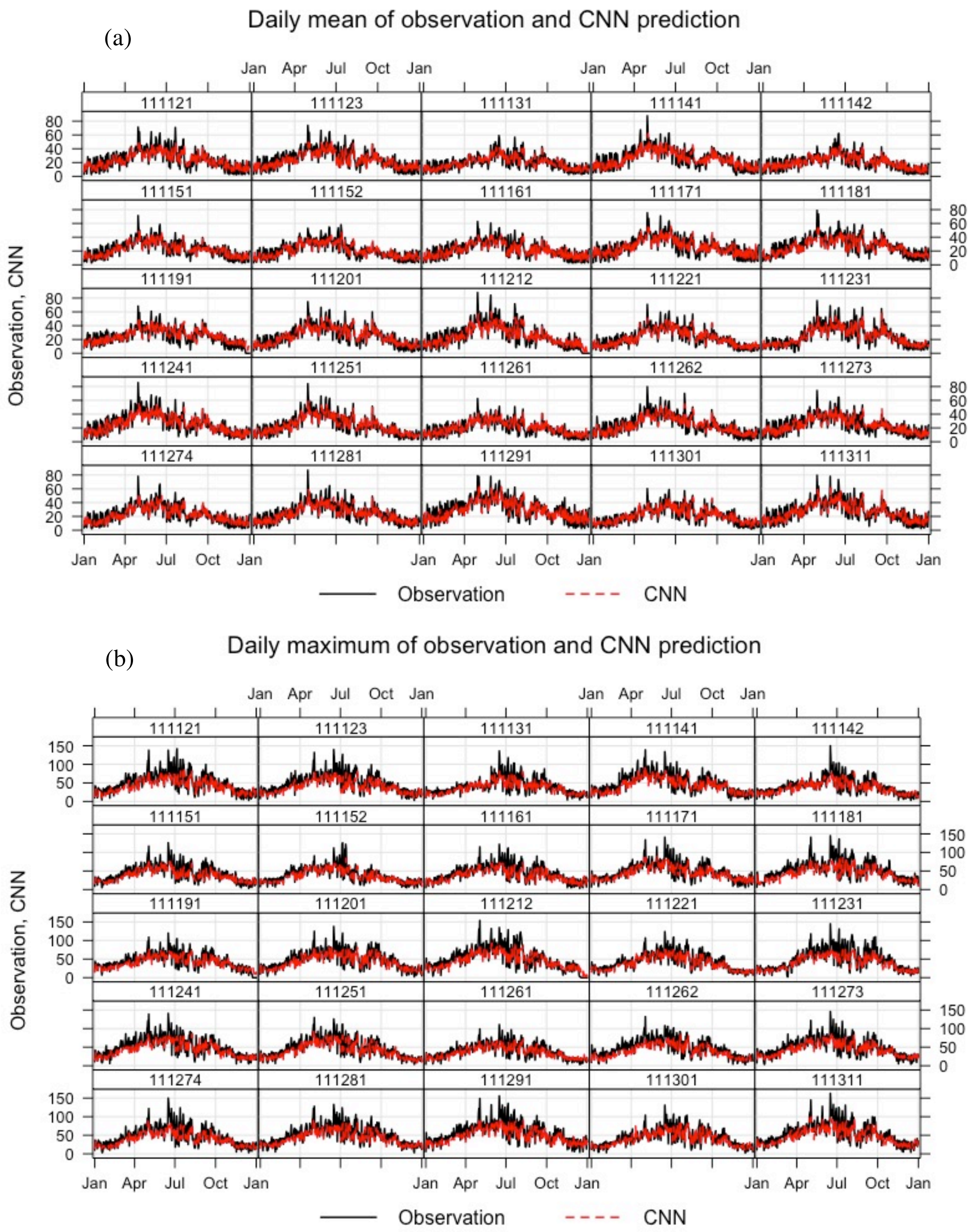


Figure 3.8. (a) Daily mean and (b) daily maximum of the observations and the CNN prediction for the NIER stations in Seoul.

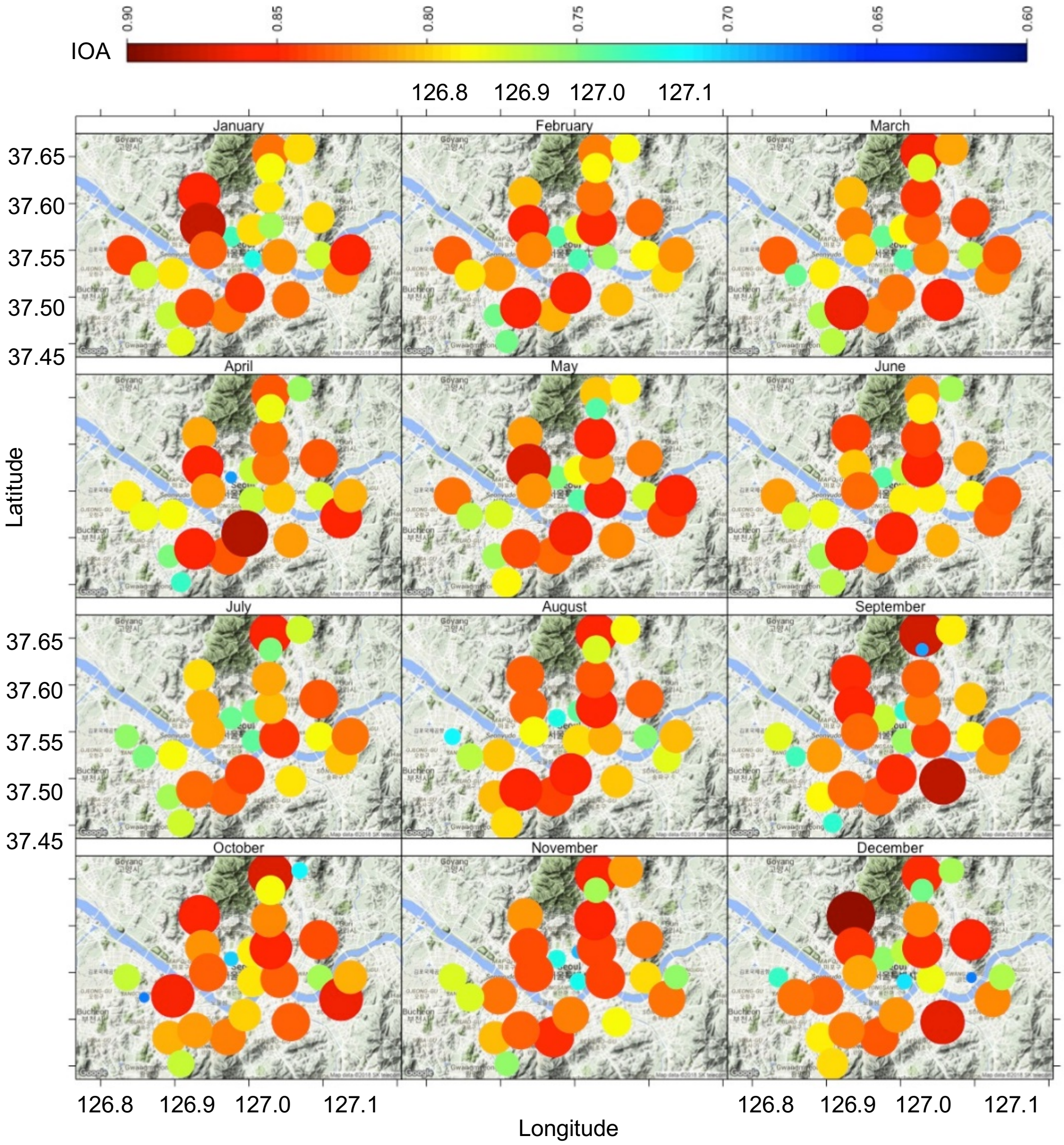


Figure 3.9. A monthly index of agreement for the NIER stations in Seoul; the larger the size of representing a circle, the higher the value is.

3.3.3 Comparison of CNN model performance to other neural networks

This study discussed a successful application of the deep learning approach, deep CNNs, in real-time prediction of an air pollutant using the previous day's air quality and weather conditions. However, there are no available or a limited number of studies which implemented deep CNNs for time series forecasting. Here, the performance of common neural network models was compared with the CNN model. The neural network models being compared include a long short-term memory (LSTM) model as an RNN, an MLP model as a traditional Artificial Neural Network (ANN), and a Stacked Autoencoder (SAE) model. Table 3.2 shows the specifications of all neural network models being used in this study.

Table 3.3 compares the performance of the neural network models for ozone forecasting with four different performance measures. These measures include overall accuracy (IOA and r), prediction bias (MAE and standard deviation) at ozone peaks, prediction bias (MAE) at hourly ozone concentrations during daytime and nighttime, and computational run time for training and predicting. For real-time ozone forecasting, the CNN model performed better with yearly IOA and r greater than the other neural network models. In terms of the relative run time, the CNN model was surpassed by ANN. The LSTM model was the second best performing relatively better than both SAE and ANN for the accuracy and bias measures. However, the LSTM model was the slowest model to be trained, more than 60 times slower than the CNN model used for this study. Details on the advantages of using the CNN model can be found in terms of overall accuracy (Figure 3.10), capturing better daily ozone peaks in most months (Figure 3.11), and predicting the nighttime ozone concentrations with better prediction biases (Figure 3.12).

Table 3.2. Specifications of neural network models compared in this study.

Model	Hidden/convolutional layer(s) structure†	Number of epochs†	Optimizer‡
CNN	32/32/32/32/32/256*	100	Adam
ANN	64	100	SGD
SAE	64/32/16/32/64	100	Adam
LSTM	64/32	350	Adam

† Optimized using trial and error tests.
‡ Both Adam and stochastic gradient descent (SGD) were explained in [58].
* Convolutional layers with filter size 32 and kernel size 2, following with a fully connected hidden layer with size 264.

Table 3.3. Comparison of four neural network models as a real-time hourly ozone prediction model. All values are based on average for all 25 NIER stations across Seoul. The bold font indicates the best performing model in each column.

Model	Yearly IOA/r	MAE/SD† of bias at ozone peaks in ppb	MAE in ppb daytime/nighttime	Relative run time‡
CNN	**0.87/0.79**	**12.7/11.5**	**10.3/ 7.5**	33.4
ANN	0.79/0.70	14.9/12.9	11.8/10.1	**1.0**
SAE	0.81/0.72	14.8/13.1	11.6/ 9.6	39.4
LSTM	0.82/0.74	14.3/12.6	11.5/ 9.1	2227.3

† Standard Deviation

‡ Unitless relative computational run time compared with the fastest model, here ANN.

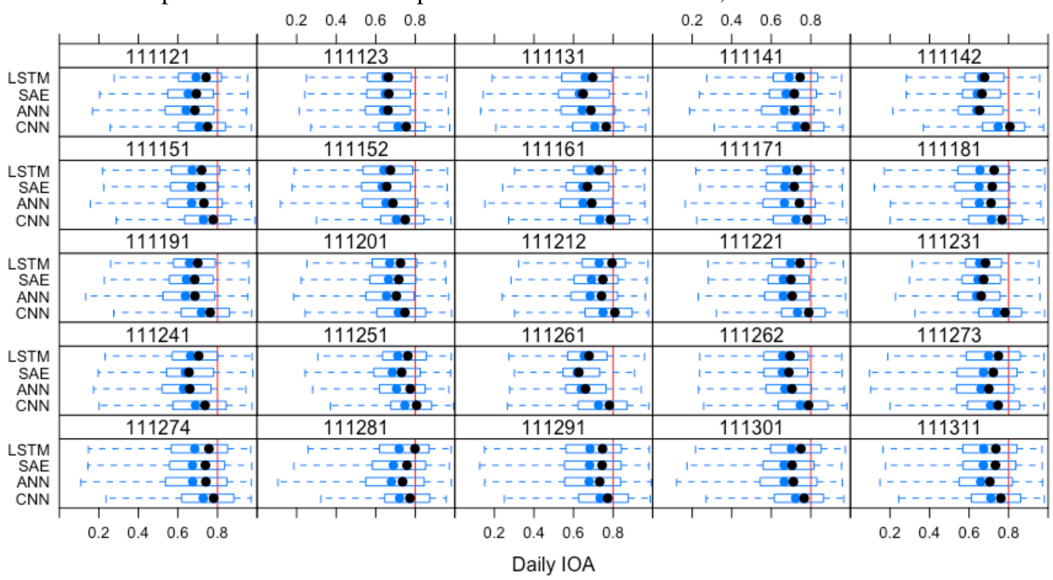


Figure 3.10. Box plot of daily IOA of different NIER stations across Seoul comparing four neural network models. The red vertical lines indicate IOA=0.8 as a comparison measure.

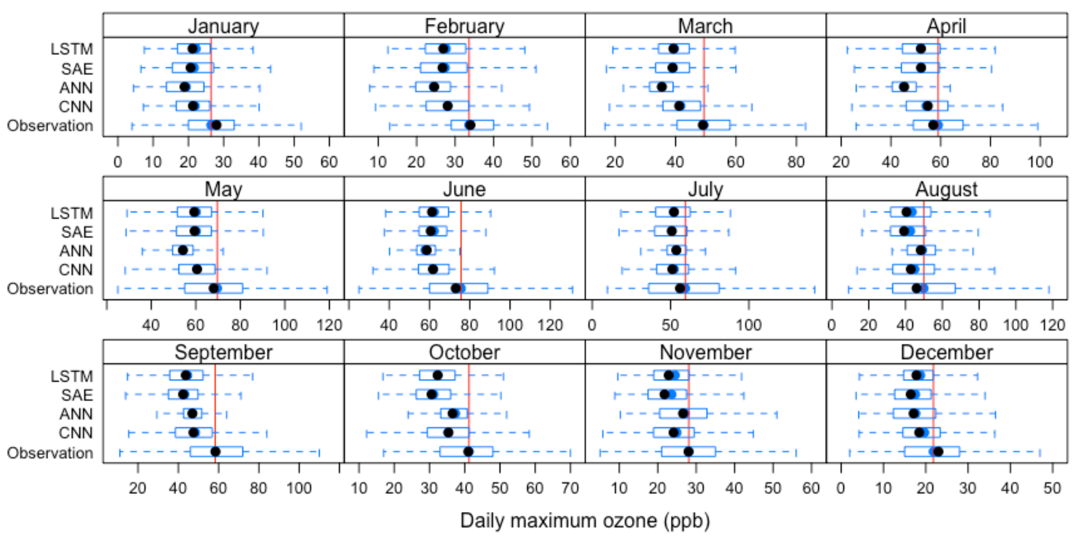


Figure 3.11. Box plot of daily maximum ozone in different months of the year 2017 comparing four neural network models, averaged over all NIER stations. The red vertical lines indicate monthly mean of observed daily maximum ozone concentrations as a comparison measure.

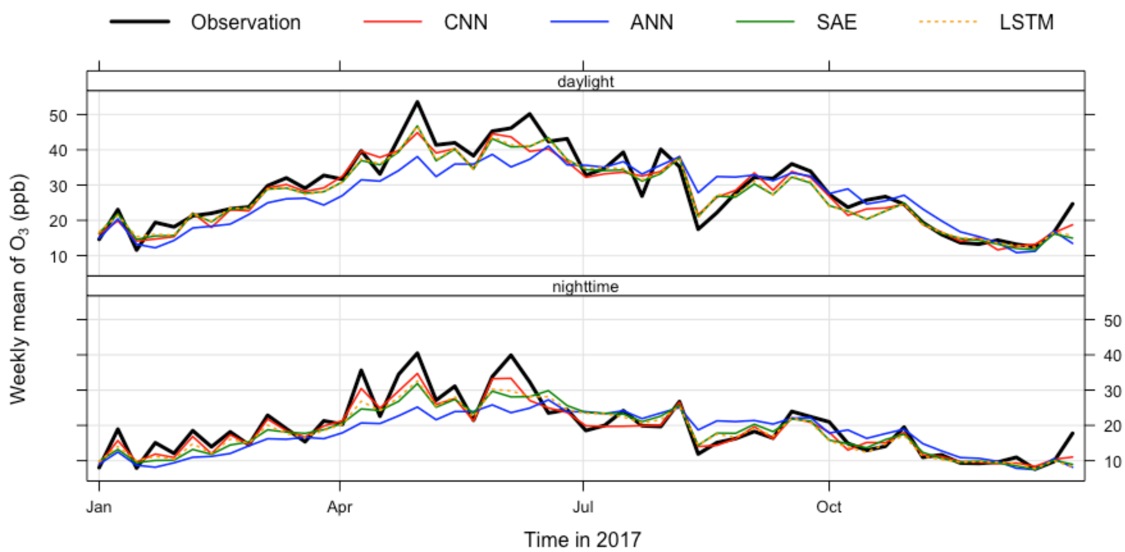


Figure 3.12. Weekly mean time-series of ozone concentrations in different NIER stations across Seoul comparing four neural network models.

3.3.4 CNN model diagnosis

Although the CNN model can successfully predict hourly ozone concentrations with reasonable accuracies and within less than a minute of processing time, several issues deserve further investigation. One limitation of this study was the underprediction of the peak ozone. As shown in Figure 3.8(a), the CNN model underpredicted all observed ozone concentrations over 80 ppbv. This large ozone bias could be attributed to the local emissions or the meteorological parameters that were not incorporated into the model. These biases could be mitigated using big data (e.g., using a significantly larger period for training the model) combined with deep learning. Using big data, a more efficient learning environment (including more training examples, more input variables, larger input size, and so on) in which to efficiently train the deep learning algorithm would be available (61). Another shortcoming of this approach is the lack of standard procedures for determining an optimal network architecture (e.g., the number of hidden layers/units, learning parameters), which is typically determined through trial and error and can significantly affect the performance of the model. As these hyper-parameters have internal dependencies, tuning them is prohibitively costly (62).

Another characteristic of CNN, like any statistical approach, is data sensitivity. Data sensitivity indicates that the quality of output directly depends on the input parameters. In light of the experience presented in this study, in which the ozone concentrations during the daytime and nighttime were separated, the performance of the CNN model differed. The monthly IOA of NIER stations calculated during the daytime and the nighttime is shown in Figures 3.13 (daytime) and 3.14 (nighttime). The differences between the IOA values for the daytime and the nighttime ranged from 0.05 to more than 0.3. The reason behind this notable difference is related to the difference in the values of input parameters during the day and night. All inputs (except for the relative humidity, NO_2, and the wind field) were at their daily minimum levels. Because of the normalization process, these values were usually close to zero in the input data. Thus, the model relied only on the relative humidity, NO_2, and the wind field for its predictions, which led to a significant reduction in the model performance. For instance, the model had a significant underprediction during the nighttime when the wind blew from the South, while it generally overpredicted the ozone for the same wind directions during the daytime (see Figure 3.15 showing categorial comparison of daytime and nighttime ozone with respect to wind speed).

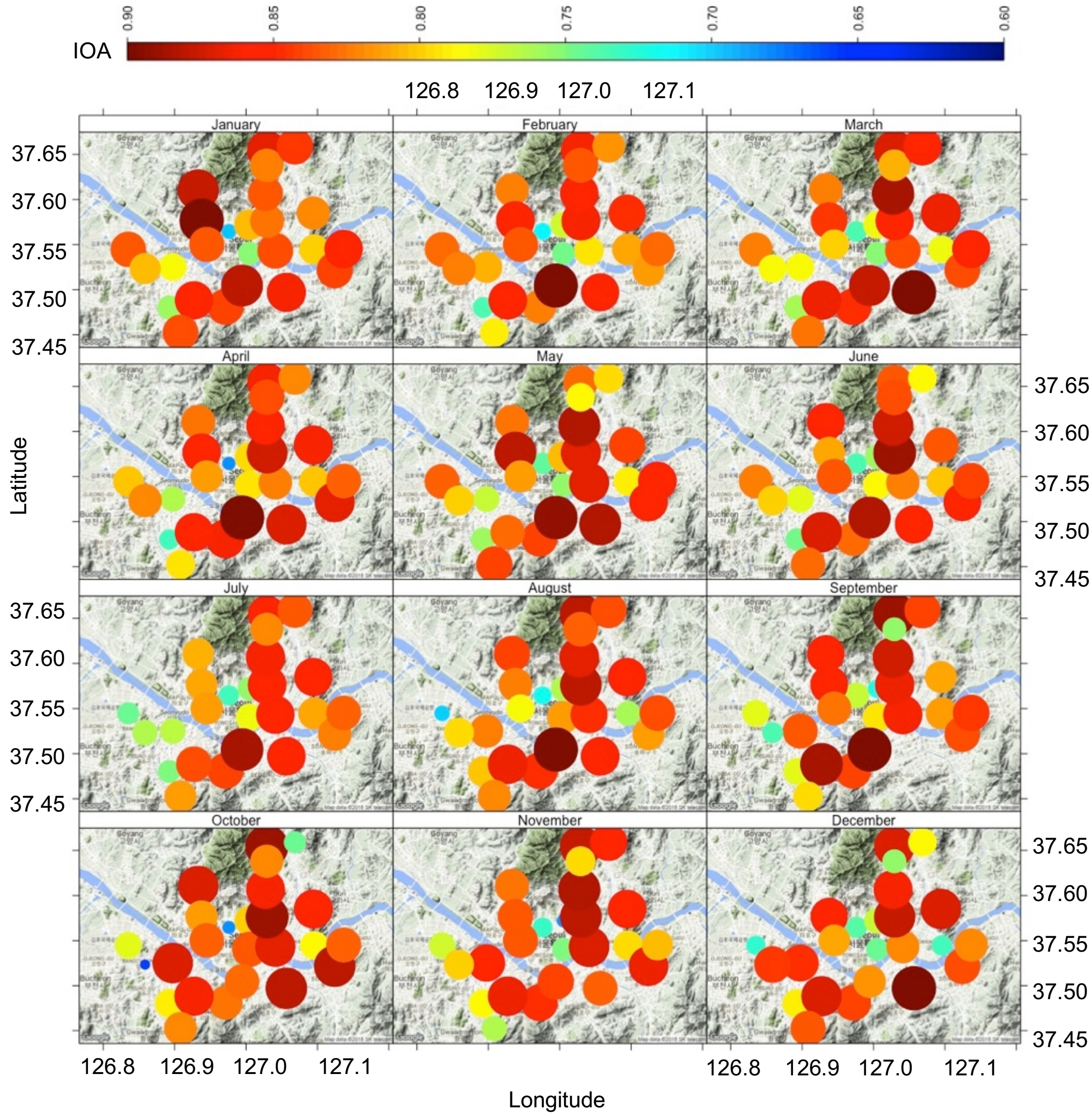


Figure 3.13. A monthly index of agreement for the NIER stations in Seoul during the daytime.

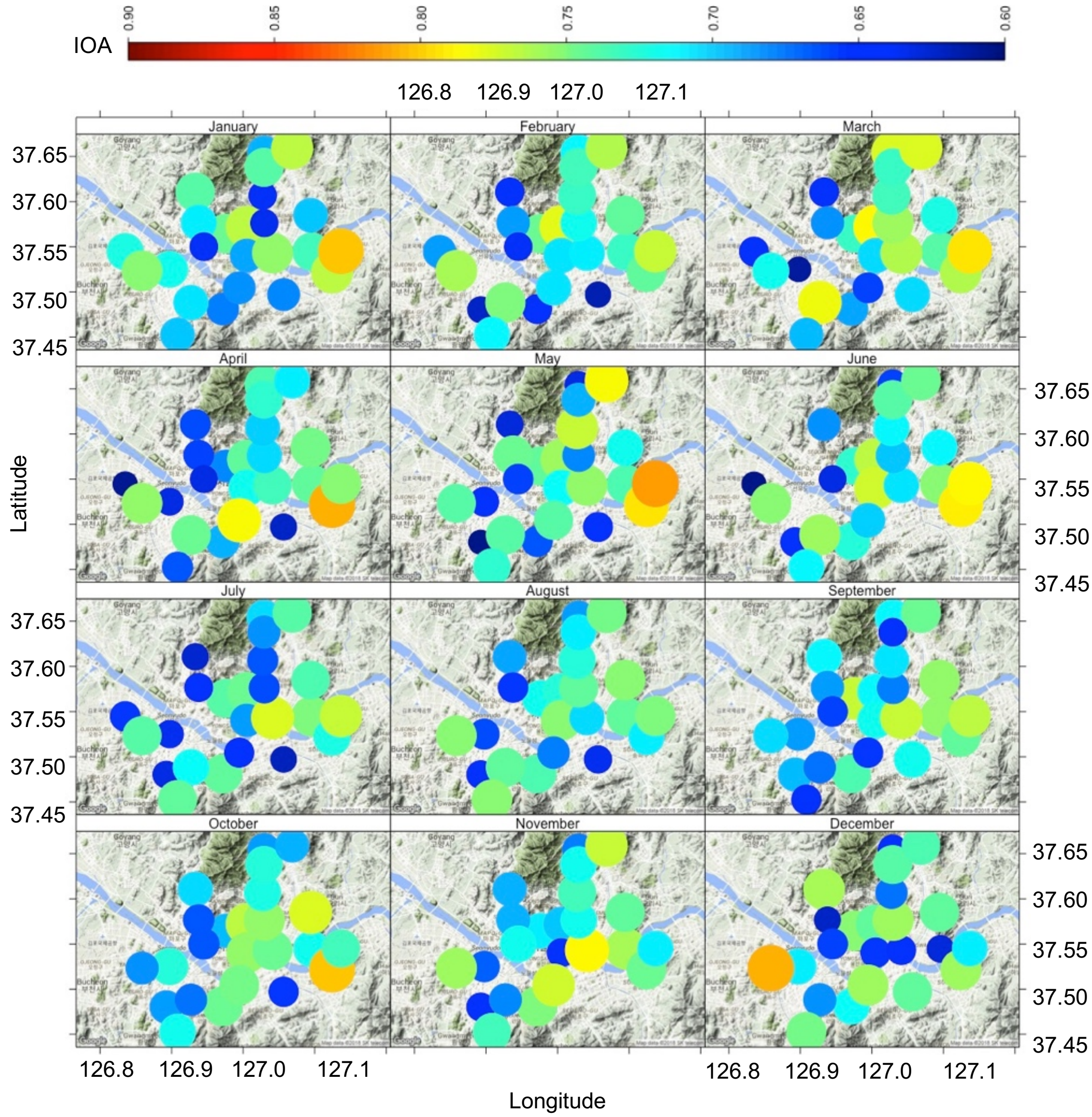


Figure 3.14. A monthly index of agreement for the NIER stations in Seoul during the nighttime.

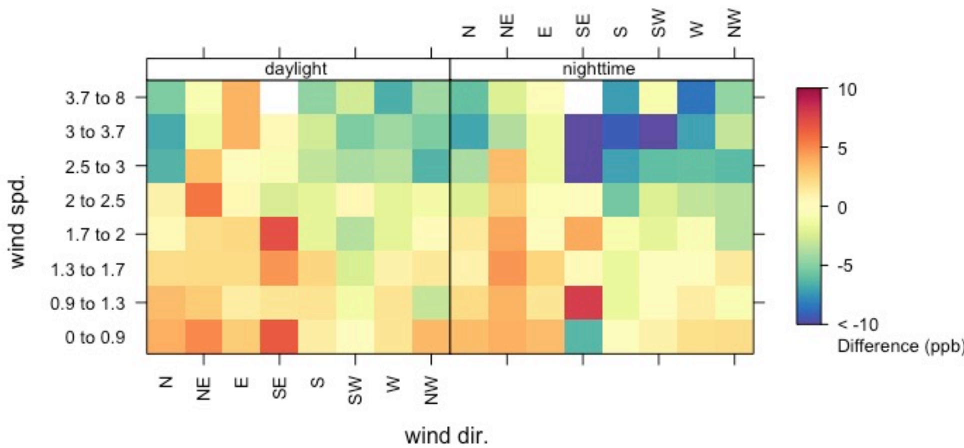


Figure 3.15. Mean bias of CNN prediction in categorial comparison during daytime and nighttime. The wind speed is in m/s.

Another drawback of this study is that we trained the CNN model using a limited number of variables from the previous day. Therefore, unlike the physical models, the CNN model was unable to take a physical phenomenon into account unless a proper indicator was used within the inputs. For example, Figure 3.16 shows the mean bias of the CNN prediction in a categorical comparison using the relative humidity and the surface temperature. In a hot (high temperature) and humid (high relative humidity) condition, the model showed a significant overestimation (red areas in Figure 3.16). This condition could have reduced the rate of the ozone formation and expressed by the presence of cloud fraction, a parameter that is lacking in the current CNN model. By contrast, a hot and relatively dry (lower than average relative humidity) condition can contribute to ozone formation during the daytime with the presence of its precursors (from local sources). In this condition, the model noticeably underestimated the ozone (blue areas in Figure 3.16) owing to the lack of information about the local sources.

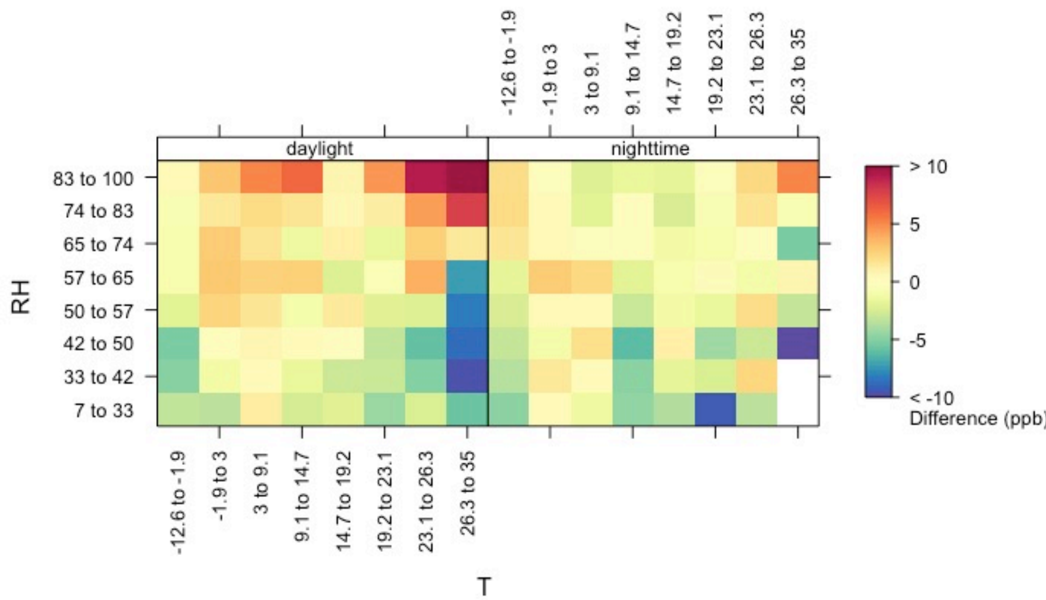


Figure 3.16. Mean bias of the CNN prediction in a categorical comparison between the daytime and the nighttime. The relative humidity is in percentages, and the surface temperature is in degrees Celsius.

The optimization algorithm in the CNN model develops an adaptive architecture to represent data and shows the influence of each feature (63). A deep learning model, however, can only be trained with historical data for its feature extraction process. Thus, the sensitivity of the input parameters of the CNN model to the output might have been imbalanced because of the significant difference in the historical examples, one of which appears in Figure 3.17, indicating that the trend in ozone concentration changed with the wind direction from 2014 to 2017. Such anomalies in the trend can result in imbalanced prediction sensitivity at different levels of input parameters. Figure 18 shows the annual mean of the CNN prediction at various category levels of four input parameters: wind speed, temperature, NO_2, and relative humidity. Because of the aforementioned limitation, the model prediction with respect to a change in the category level is less sensitive than the observation on the same day of the week.

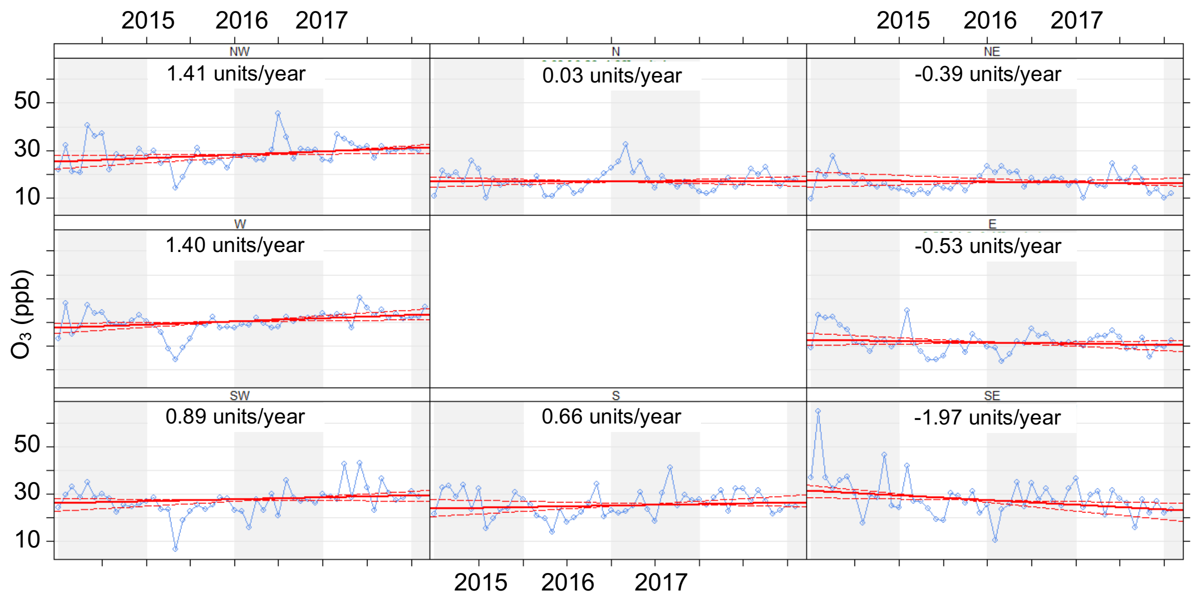


Figure 3.17. The change in ozone concentration trends in different wind directions in station 111121.

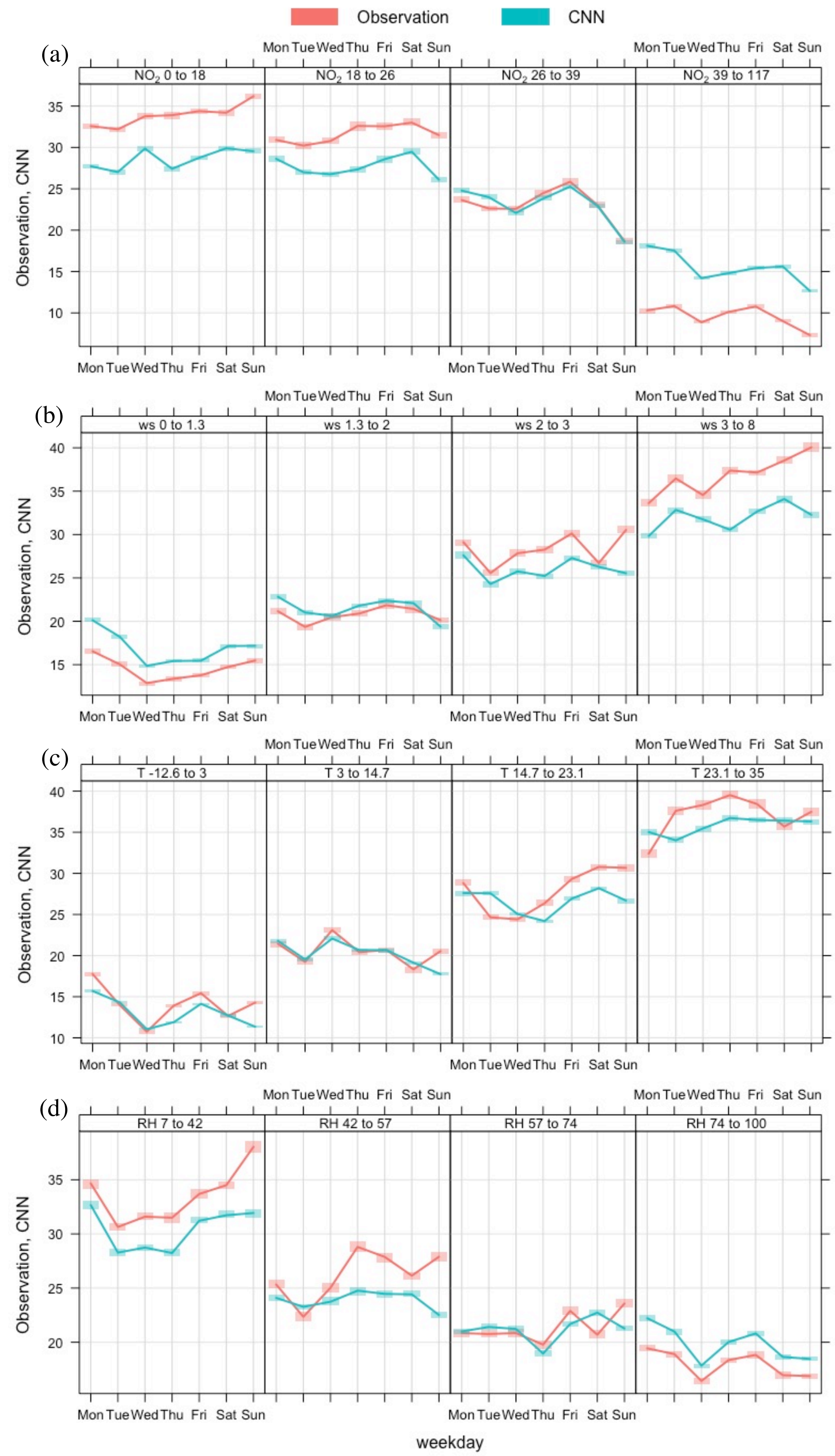


Figure 3.18. The annual mean of the CNN prediction in a categorical comparison between days of the week. NO_2 is in ppb (a), the wind speed is in m/s (b), the surface temperature is in degrees Celsius (c), and the relative humidity is in percentages (d).

3.3.5 CNN limitations

While the model maintained a proper level of the prediction accuracy, it was prone to two main limitations: (i) varying performance in different time of the year (see Figure 3.19 showing boxplot of daily IOA of CNN model in different weeks); and (ii) higher relative bias (as shown in times-series in Figure 3.20) and lower modeling performance during the nighttime as compared with the daytime (see Figures 3.13-14). In general, the varying, time-dependent modeling performance can be explained by wavelet transform (64), while significant difference between the modeling performance during daytime and nighttime is regarding the undertrained CNN model.

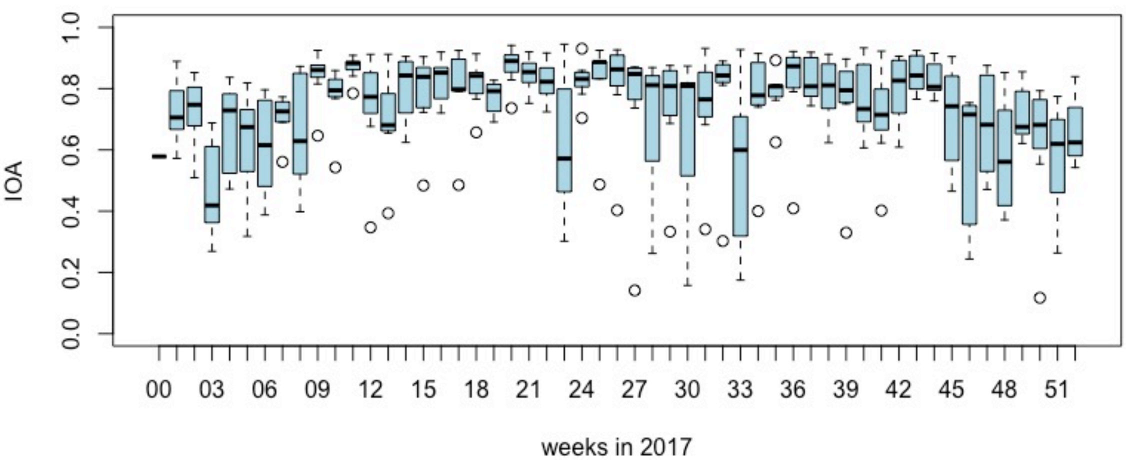


Figure 3.19. Box plot of daily IOAs of the CNN model in different weeks of 2017, averaged over 25 stations in Seoul.

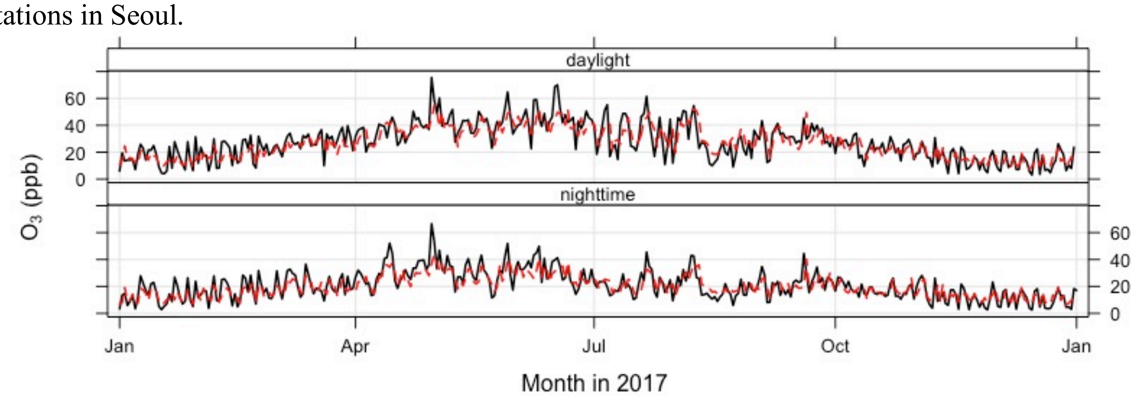


Figure 3.20. Time series showing daily CNN model predictions (in red) and observed ozone concentrations (in black) during the daytime and the nighttime, averaged over 25 stations in Seoul, South Korea.

3.3.5.1 Time-dependent model performance:

The performance of the CNN model is directly dependent on well the model understands the relationship between the inputs (meteorology and ozone precursors) and output (ozone concentration). While emission sources from volatile organic compounds (VOCs) and NOx are relatively constant in time, meteorological variables governs the variation of the ozone in different through the year. Temperature, wind speed, and relative humidity (RH) are among the most important meteorological parameters affecting ozone variation.

In order to obtain better understanding of model performance in different time of the year, wavelet transform was used to explain the importance of the different 'modes' in the time-series. Wavelet transformation decomposes a signal into a scale frequency space, allowing the determination of the relative contributions of the different temporal scales present within a signal (65). Wavelet decompositions are powerful tools in analyzing the variation in signal properties across different resolutions of geophysical variables (65-67). The wavelet transform overcomes the inability of the Fourier transform to represent a signal in the time and frequency domain at the same time by using a fully scalable modulated window that is shifted along the signal (see Figure 3.21 for schematics of different decomposing methods). For every position, the spectrum is calculated. After repeating the process, each time with a different window size, the result is a collection of time-frequency representations of the signal, all with different resolutions. Data are separated into multiresolution components; each component is studied with a resolution that matches its scale (65-67). The high-resolution components capture the fine scale features in the signal, whereas the low-resolution components capture the coarse scale features.

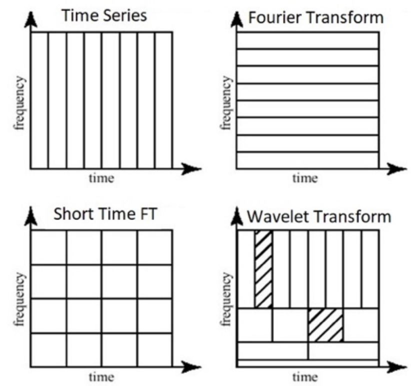


Figure 3.21. Schematics comparing wavelet transform to Fourier transform (FT).

Wavelet analysis represents any arbitrary (nonlinear) function by a linear combination of a set of wavelets or alternative basis functions, making them very suitable to be used both as an integration kernel for analysis to extract information about the process and as a basis for representation or characterization of the processes (68). Figure 3.22 shows the hourly ozone time series of a monitoring station in downtown Seoul (NIER stations #111121) with its the wavelet transform for the year 2017. Here, wavelet transform shows a strong wavelet power levels associated with period=24 and period=168 in the middle of the year indicating dominant daily (24 hours) and weekly variation (168 hours).

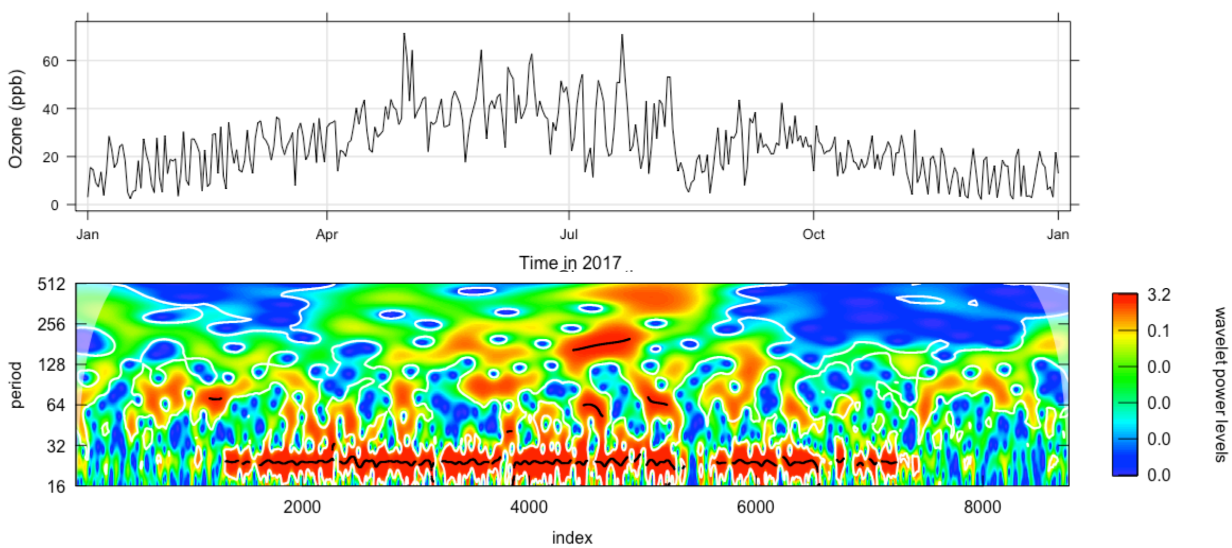


Figure 3.22. Time series and wavelet transform analysis of ozone concentrations from the NIER stations #111121 in Seoul, South Korea. Index at the x-axis as well as the period are in hours.

Figure 3.23 shows the wavelet power transform of aforementioned meteorological variables for 2017. In this figure, both index and period are in hour since the hourly time-series was used to calculate the wavelet powers. The figure also locates five different time periods indicating significant performance variations. From Figure 3.19, the CNN model underperformed during weeks 3-9 as well as weeks 44-51, showed as 'Worst CNN results' in Figure 3.23. For weeks 14-22 and 42-44, the CNN model showed the best forecasting results. Between weeks 29 and 33, the CNN model predictions encountered a significant under-estimation indicated as 'Large under-prediction' in Figure 3.23. This figure shows strong wavelet powers for 24-hourly (daily) period for all variables. This is due to strong diurnal variation of such parameters controlled by the sunlight. The wavelet powers for wind speed were generally larger than RH, while the temperature showed lower, but more consistent daily modes. This is important since the CNN model can detect the specific 'patterns' in temperature better than wind speed and RH. Thus, when the daily modes are stronger in temperature, it is likely that the CNN model performs better. In contrast, when the daily modes of meteorological variables are relatively weak, the CNN model performs poorly (see Figure 3.23).

Another reason for the poor performance of the CNN model in the selected time period is the relatively large coarse modes (period > 24 hours). The CNN model was only received the information from the last day, hence, was unable to address the bidaily and weekly trends within the input data. For instance, for the time period with large underpredictions, the coarse modes in wind speed were even larger than daily mode. Thus, employing longer history is required to properly explain the relation between wind speed and ozone. Figure 3.24 compares the average wavelet powers in different periods (from daily to weekly modes) of CNN prediction and observation data. It shows, the powers for both time-series match for the periods around 24 hours. However, after 32 hours, the wavelet power of the CNN model shrinks to a relatively constant power, while, for the observation, it reaches local extremums in around 3, 5, and 7 days.

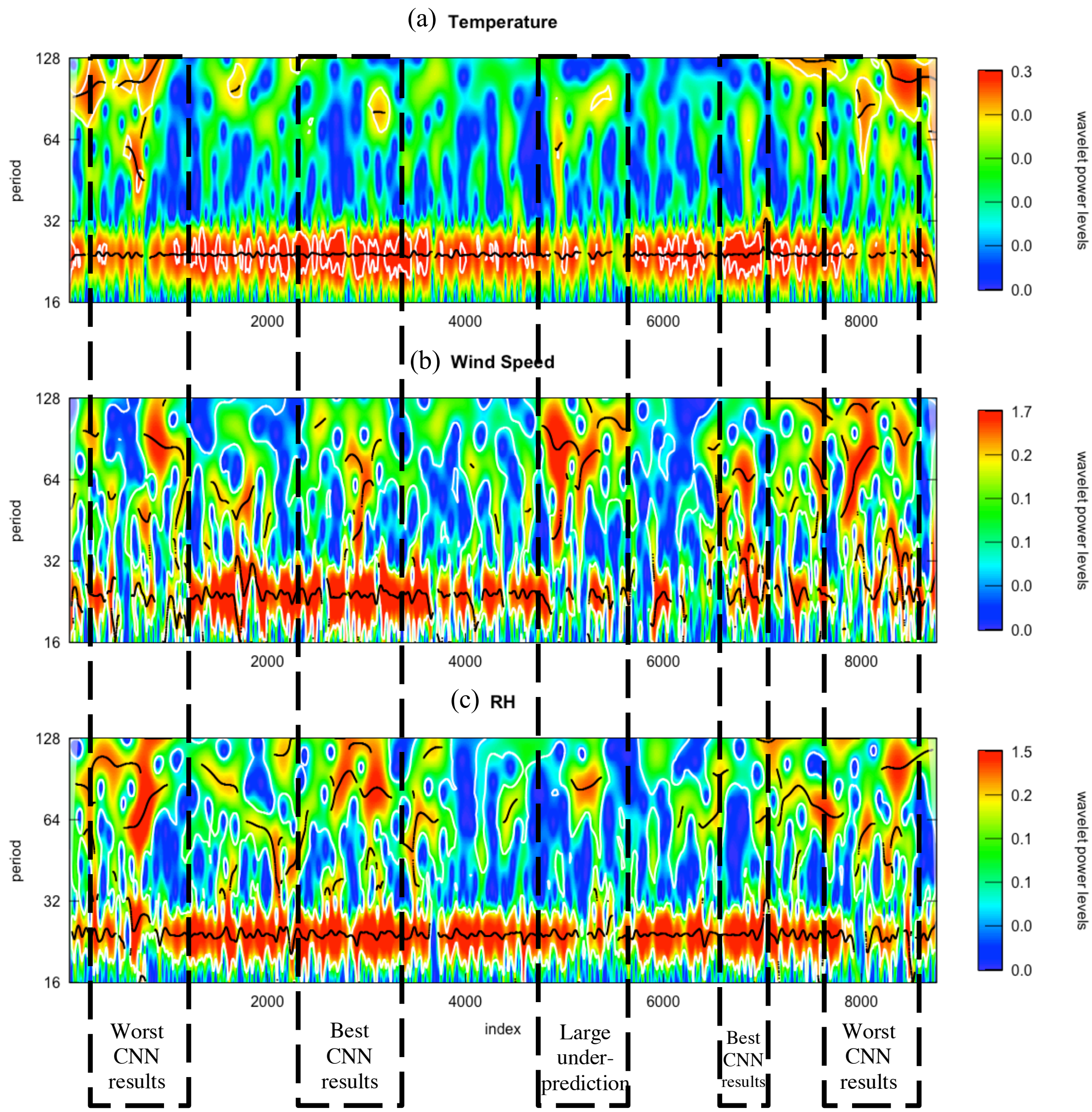


Figure 3.23. Wavelet power transform of (a) temperature, (b) wind speed, and (c) RH% for 2017 in Seoul, South Korea.

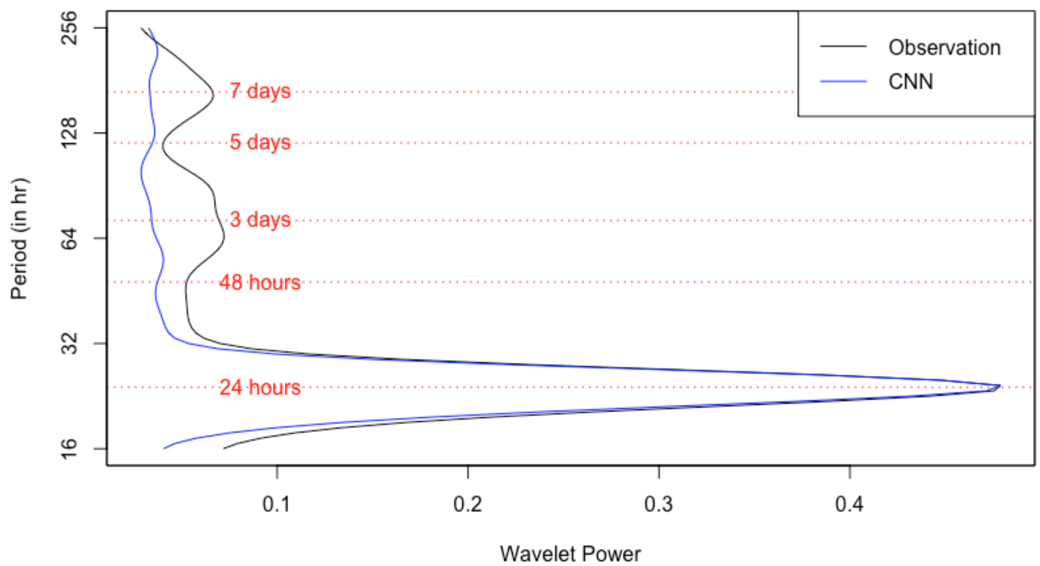


Figure 3.24. Wavelet power for various time periods (modes) for CNN predictions and observations.

3.3.5.2 Low modeling performance during nighttime:

As mentioned before, the CNN ozone forecasting system faced a significant modeling bias in estimating air quality concentrations during the nighttime. This bias resulted more a reduction in prediction accuracy by more than 20% in nighttime compared with that in daytime. A similar issue also can be seen in CTMs even with complex physical and chemical equations in explaining diurnal variation of ozone concentration.

One reason for this bias was likely due to variation of the meteorological inputs during the nighttime. While their absolute values were generally higher during the daytime, the relative frequency of their variations were more pronounced during the nighttime. This caused a discontinuity in learning process in the CNN model. Since both daytime and nighttime hours were presented as inputs, the CNN model minimized a cost function that contained 'normalized' errors in both daytime and nighttime (cost function was mean squared errors or 24-hour ozone predictions at each step). On the other hand, the daytime hours are generally more that nighttime ones. Also, the accumulation of NO_2 concentration for these extreme cases was mainly due to stagnant atmospheric condition with wind speed near its yearly minimum values. Given these facts, the CNN model was prone to a characteristic bias in nighttime ozone estimation. A customized cost function could be a potential solution to this limitation and requires further investigations.

Another reason in reducing the CNN model performance during the nighttime is misinterpretation of extreme conditions of input parameters. Figure 3.25 shows scatter plots comparing CNN predictions and observation by the levels of two important ozone precursors (NO_2 concentrations) and meteorological variable (RH%) separated for daytime and nighttime. NO_2 concentration was generally higher during the nighttime when the ozone concentration was near zero for extreme NO_2 values due to proper condition for ozone depletion with the absence of sunlight. However, the CNN model was unable to capture this relationship and overestimated these cases (See Figure 3.25a).

In contrast to the mentioned overestimated events, Figure 3.25b shows underestimation of nighttime ozone when the level of RH% was generally high, mostly during the warm days. There are two reasons for such underestimated events. First, high (or low) levels of RH% and surface pressure generally occur around the same time in early morning (or late afternoon) when the planetary boundary layer (PBL) is at its lowest (or highest) level during the day. With these extreme conditions, the earlier sunrise (or later sunset) in summer months prepares a proper condition to elevate the ozone concentrations. These events normally occur only during short period of time, resulting in undertraining the CNN model to capture these relations.

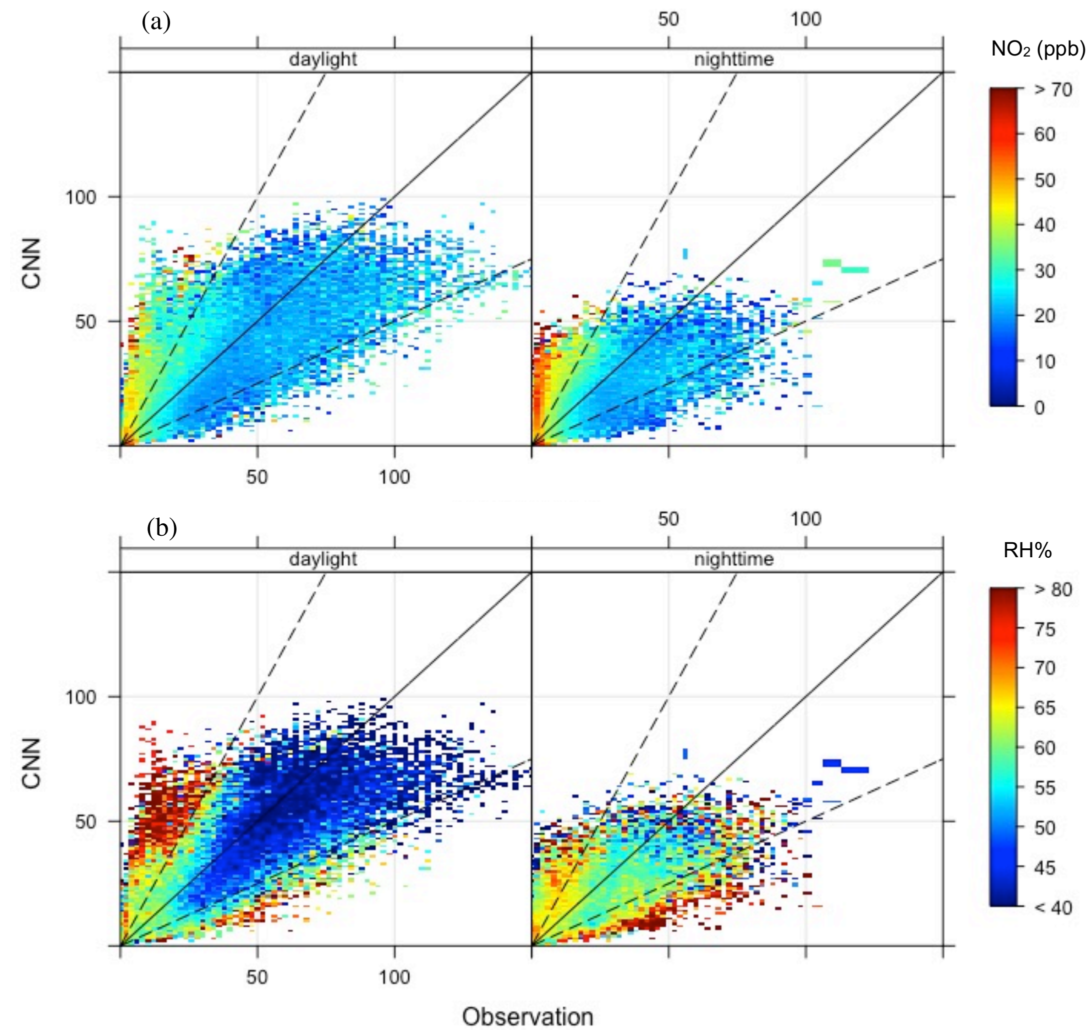


Figure 3.25. Scatter plots comparing CNN predictions and observations with respect to levels of (a) NO_2 concentrations and (b) RH%.

3.3.6 Summary

We discussed the application of deep convolutional neural networks (CNN) for the real-time prediction of ozone concentrations in Seoul, South Korea and trained the model to predict hourly ozone concentrations for the next day using the observations of NO_x, ozone and meteorological variables from the previous day. We evaluated the model for the entire year of 2017. This work has shown that the deep learning approach can predict hourly concentrations with sufficient accuracy (IOA=0.84-0.89, r=0.74-0.81) by modeling the relationship between local meteorological and species concentrations in an urban environment. The CNN model, which showed consistent prediction results across the city of Seoul, reasonably predicted daily and monthly trends of ozone concentrations throughout the year. However, the model generally underpredicted the maximum daily ozone, particularly during the summer. This is due to several important meteorological parameters, such as cloud fraction and solar radiation, were unavailable for the training period in this study. The CNN model was generally under-trained for forecasting the high ozone peaks during the hot summer days of 2017.

The study also demonstrated that the predictions of the CNN model were generally more accurate (higher IOA and r values with a lower mean bias) for the southern region of the Han River since the topography was more consistent and resulted in a more accurate interpretation of the input parameters. Furthermore, the CNN predictions of daytime ozone concentrations were generally more accurate than those of nighttime concentrations with differences in IOA between 0.05 and 0.30. We attribute this discrepancy to the occurrence of most of the daily maximum input variables during the daytime. We compared the performance of the deep CNN model with other common neural network models such as the shallow artificial neural network (ANN), long short-term memory (LSTM), and deep stacked autoencoder (SAE). The CNN model performed better than other neural network models in terms of overall accuracy (around 5% better IOA values than LSTM), smaller bias at daily ozone peaks (1.5 ppb lower that LSTM), and smaller MAE during the nighttime (2.2 ppb lower than LSTM). The CNN model also predicted the hourly ozone concentrations faster that both LSTM and SAE.

The CNN model not only predicts real-time ozone concentration with favorable statistics but also generates the result within less than a minute of initiating the model. Beyond surveying the advances of using a deep CNN, we showed the limitation of such methods for real-time air quality forecasting. For instance, a proper number of input variables (predictors) should be used with a sufficiently large amount of training data. However, if an important predictor of ozone concentration is missing (e.g., cloud fraction and solar radiation), it will influence the sensitivity of the model to the other input variables, which may lead to "misprediction." Regarding underperformed for the cold months as well as during the high ozone episodes, I found that the fine wavelet modes (daily and hourly) were relatively weaker that the rest of the year. Also, when the coarse modes were strong, the CNN model prone to predicting with large errors. I also found that the model's underperformance in nighttime hours was due to undertraining the model, and extreme values of input parameters during the nighttime.

The proposed approach in this paper can be applied to and yield a high prediction accuracy for ozone or other pollutants in other metropolitan areas. In addition, the deep learning approach can potentially be used for a multiple-day forecast of air pollution or air quality index. Fast and accurate air quality prediction using the deep CNN model could be used to reduce the adverse health effects of urban air pollution. Given the computational efficiency of the CNN algorithm, deep learning could supplement deterministic models to more rapidly and accurately forecast air pollution concentrations. I expect that this study will not only provide a more comprehensive understanding of CNNs but also facilitate future research activities and applications within the field of atmospheric sciences.

CHAPTER 4. A DATA ENSEMBLE APPROACH FOR REAL-TIME AIR QUALITY FORECASTING USING EXTREMELY RANDOMIZED TREES AND DEEP NEURAL NETWORKS

4.1 Introduction

Following previous chapter for designing a CNN model for a real-time hourly ozone forecasting system in Korea, this chapter aims to develop the ensembles of various machine learning models for real-time hourly ozone forecasting in Seoul, Korea. Chapter 4 present an ensemble model that integrates two regression models: low- and high-ozone peak models. Since the number of high-ozone peak episodes is significantly lower than that of low-ozone episodes, the ML model is "undertrained" for predicting days with ozone concentrations over 90 ppb. Compared to each of the aforementioned models, the ensemble model, which accounts for the global and local regression characteristics of both low- and high-ozone peak episodes, is a more accurate parametric model. Therefore, its parameters are determined by the dynamic characteristics of the hourly ozone concentrations categorized by their maximum values. We use a combination of two advanced ML models—the deep neural network (DNN) (31) and extremely randomized trees (Extra-Trees method) (69, 70)—as the base models for each step in the ML ensemble model. We also develop two generalized models by i) merging all samples from all sources, and ii) uniformly distributing the samples based on target ozone peaks. In addition, we develop regularized models that establish the training by focusing more on episodes with high-ozone peaks greater than the threshold (90 ppb). Through various comparison metrics and statistical indices, we verify the effectiveness and robustness of our ensemble models with the base ML models.

4.2 Materials and Methods

The goal of the ensemble methods is to combine the predictions of several base predictive models built with a given learning algorithm to improve the generalizability/robustness over a single predictive model. Thus, we combined several independent machine learning models to produce a powerful ensemble model for predicting hourly ozone concentrations, particularly for capturing the relatively higher daily ozone peaks. In this regard, we used two machine learning algorithms, the extra-trees method and DNN models, for our ensemble approach. In our proposed models, we focused on preparing proper training samples with regard to their numbers and distributions. That is, we designed the training dataset for each ensemble approach to either generate more accurate predictions or capture more enhanced ozone peaks.

4.2.1 Extra-Trees method

Decision trees predict the value of a discrete dependent variable with a finite set of values (referred to as a "class"). They use the values of a set of independent variables (called "attributes"), which may be either continuous (for regression problems) or discrete (for classification problems). Decision tree algorithms, also referred to as the top-down induction of decision trees, entail a divide-and-conquer approach (71). Whereas continuous attributes consist of a threshold at which point the (sub)tree splits into two "branches," discrete attributes have "branches" created for each possible value of an attribute. The final subsets, referred to as “leaves,” are labeled with a class. While a limited number of studies have used decision tree methods for air quality forecasting (11, 12), none used advanced decision trees techniques such as extremely randomized trees, that is, the extra-trees method, for such a purpose.

Several studies have attempted to overcome the inadequacies of the conventional decision trees (e.g., their sub-optimal performance and lack of robustness) (72). One popular approach used the extremely randomized trees model to create an ensemble of trees. This method (69, 70), a perturb-and-combine technique designed explicitly for decision trees algorithm, creates a variety of classifiers by introducing randomness in their construction. In extremely randomized trees, randomness goes one step farther in the way “splits" are computed. Like the random forest algorithm, the extremely randomized trees method uses a random subset of candidate features. Instead of searching for the most discriminative thresholds, however, it randomly obtains thresholds for each candidate feature and selects the best randomly-generated threshold as the splitting rule, which usually allows further reduction of the variance of the model at the expense of a slight increase in the bias.

4.2.2 Data

To model the ozone concentration time series, we used several predictors, including the hourly observed values of O3 and NOx concentrations (as recorded by South Korea’s National Institute of Environmental Research, or NIER), surface temperature, relative humidity, wind speed, direction, dew-point temperature, surface pressure, and precipitation (as recorded by the Korean Meteorological Administration, or KMA) as input. Each input parameter had a size of 24, representing the previous day’s hourly values. In addition, to address seasonal and weekly variations of ozone concentration, seasons (spring, summer, autumn, winter) as well as weekday/weekend were one hot-encoded as input for each model (one hot-encoding is a process by which categorical variables are converted into an integer form for a regression prediction by machine learning algorithm). The locations of the air and meteorology stations can be found in Figure 3.1 in Chapter 3. Similar to the CNN modeling approach in Chapter 3, I used several predictive techniques to forecast the hourly surface ozone concentrations for the year 2017 and selected the historical surface measurement data from 2014 to 2016 to train the model. Such a training period provided a broad history for fitting the relationship between the input variables and the ozone concentrations. To treat the missing observation data, I applied SOFT-IMPUTE by Mazumder et al. (59) to the raw observation data. SOFT-IMPUTE is a missing data treatment approach that iteratively replaces missing elements with those obtained from a soft-thresholded singular-value decomposition by taking all available data (spatially and temporally) into account.

4.2.2.1 Data preparation

The ensemble methods presented in this study entail a "cutoff" value determined by the daily maximum ozone concentration, which distinguishes between the data samples for the high-and low-ozone peak models. In Seoul, a maximum hourly ozone concentration above 90 ppb indicates a "poor" air quality level that imposes a severe health hazard to the urban population. Therefore, to prepare our data for analysis, we chose a daily maximum ozone concentration of 90 ppb as the cutoff point.

To construct the data sample for both models, I gathered and merged three years of training data (2014 to 2016) from all of the stations. Then, I counted the number of days with a maximum target ozone value equal to or greater than the 90 ppb cutoff value for the high-ozone peak training dataset and the remaining data for the low-ozone peak training dataset. I used these datasets in the first ensemble sub-model.

4.2.3 Modeling approaches

Figure 4.1 schematically illustrates the models proposed in this study. For the ensemble modeling approach, I used three combinations of the extra-trees method and the deep neural network, the former for constructing two generalized models (station-independent models) following two modeling approaches, “merge” and “uniform.” The characteristics of the extra-trees method are more amenable to the construction of such generalized models than those of the DNN. In addition, compared to DNN, the extra-trees method requires less computational time and little fine-tuning of the hyper-parameters. To capture more accurate daily high ozone peaks, I proposed a regulation process that ensured that the model focused more on the high peaks and distributed the training data in such a way that 50% of the data included high-ozone peaks (above 90 ppb) and the remaining data were uniformly distributed among all other bins of the ozone concentrations (see Section 3-2).

4.2.3.1 Ensemble model:

The proposed ensemble models included two sub-models. The first used the low- and high-peak datasets (as discussed in Section 2-4) separately to predict hourly ozone concentrations. The low-and high-ozone peak models are trained by the corresponding low- and high-ozone peak training sets. The second sub-model used a uniformly distributed training dataset to predict the final hourly ozone concentrations. I used a uniformly distributed training dataset to train the ensemble model in order to ensure that the model accounted for the outputs of both the low- and high-peak models. Such an approach precluded any inclination towards the low-ozone peaks during the training process that results from the relative abundance of low-ozone peak days compared to the lower number of high-ozone peak days. Thus, I used three combinations of the extra-trees regression model and a deep neural network that represent the following sub-models: the DNN-DNN, the ExtraTree-DNN, and the ExtraTree-ExtraTree. Table 4.1 summarizes the modeling approaches in this study, including the proposed ensemble models.

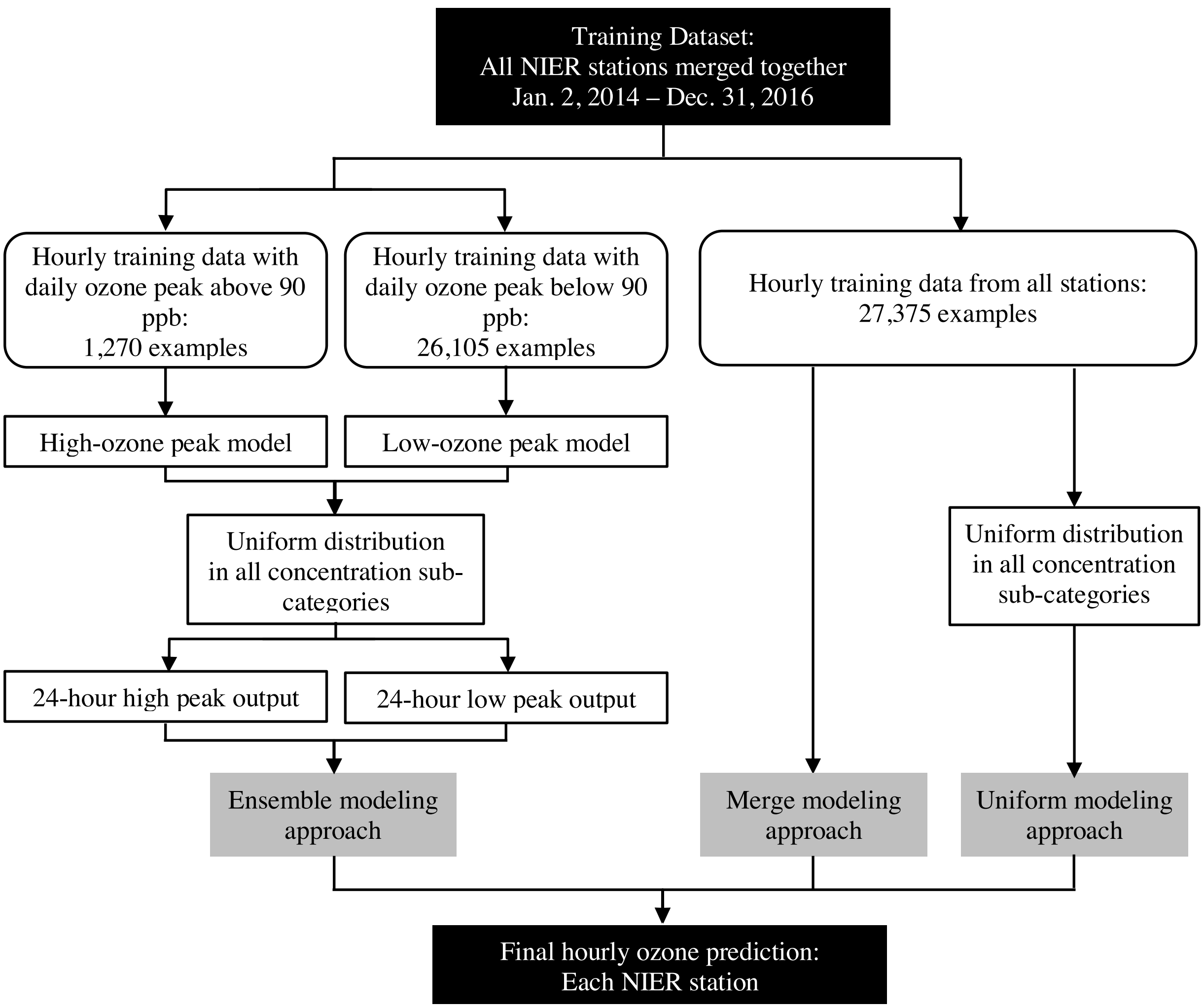


Figure 4.1. Schematic of the proposed modeling approaches (inputs are meteorological parameters and air pollution concentrations). For the ensemble approach, we used three combinations of the extra-trees decision trees and deep neural networks. The extra-trees model was used for the merged and uniform modeling approaches.

Table 4.1. Summary of the models being used in this study.

Model Category	Model Name	ML Algorithm	Model ID
Base models	Extra-Tree	Extra-Trees method	Extra-Tree
	DNN	DNN	DNN
Single models	Merge	Extra-Trees method	Merge
	Uniform	Extra-Trees method	Uniform
	Biased Uniform	Extra-Trees method	Uniform_Regularized
Ensemble models	ExtraTree-ExtraTree	Extra-Trees method	ExtraTree_ExtraTree
	ExtraTree-DNN	Extra-Trees method and DNN	ExtraTree_DNN
	DNN-DNN	DNN	DNN-DNN
Ensemble models with regularization approach	Regularized ExtraTree-ExtraTree	Extra-Trees method	ExtraTree_ExtraTree_ Regularized
	Regularized ExtraTree-DNN	Extra-Trees method and DNN	ExtraTree_DNN_ Regularized
	Regularized DNN-DNN	DNN	DNN_DNN_ Regularized

4.2.4 Uniform model

Using the aforementioned uniformly distributed training dataset, I employed the extra-trees regression model to predict hourly ozone concentrations. The difference between the uniform and ensemble models is that within the uniform model, only one model (the Extra-Trees model) is used for the hourly prediction instead of the two-step modeling process involved in the ensemble models. The uniformly distributed training dataset may improve accuracy levels in each distribution bins, including the bin of higher maximum daily ozone concentrations. To create a uniform distribution of the data for the second ensemble sub-model, I collected all samples from the high-ozone peak datasets with an equal number of samples from the low-ozone peak dataset within a certain type of sub-categories. To ensure uniform coverage of maximum ozone values below the cutoff, we uniformly selected samples from the low-peak dataset from four bins with the following daily maximum ozone values: 0-25, 25-50, 50-70, and 70-90 ppb. Hence, the number of samples with a daily maximum ozone concentration of 90 ppb and above was equal to these four bins.

4.2.5 Merge model

For this model, I used a merged dataset in which all data samples collected from all stations in Seoul were concatenated. Even though using the merged dataset dramatically increased the number of input data, the meteorological input parameters remained the same for all of the stations on any given day. After all, only one meteorological station (KMA station #108) was available during the time period of this study.

4.2.6 Data regularization approach

I adopted this methodology to ensure a more pronounced representation of the high-ozone peaks in the training dataset, and to address the issue of "under-sampling" of high-ozone episodes [3], which resulted in underpredictions of such episodes, we applied a regularization process. Note that, the regularization process in this study is only for constraining training and testing datasets in terms of presenting more samples with higher ozone concentration. Hence, the process is different from the regularization process for optimizing a machine learning algorithm. As a result, I was able to add more samples with the highest daily maximum ozone concentrations to the bin in the uniform model. The training dataset contained an equal number of samples from the low-ozone episodes (samples with daily maximum ozone of less than 90 ppb) and high-ozone episodes (samples with daily maximum ozone of more than 90 ppb). For the samples in the low-ozone episode category, I further distributed the dataset into uniform sub-categories, described in Section 3-2. This model is referred to as the "biased uniform" model (the figures refer to it as the "regularized uniform model"). As this approach, however, could have resulted in an overprediction for low-ozone peak episodes, we implemented a regularization process within the proposed ensemble models to balance the trade-off between the underprediction of high-ozone episodes and the overprediction of low-ozone episodes. To decrease the biased weight of the model towards high peaks, I trained an ensemble of the low-ozone peak model and the biased uniform model. The low-ozone peak model balances the over-trained high-ozone peak predictions rendered by the biased uniform model. This model is called the “data-regularized ensemble” (e.g., the regularized DNN-DNN model) (see Table 4.1).

4.2.7 Model configuration

For the DNN models, I used a deep architecture with two fully-connected hidden layers consisting of 370 hidden units in the first layer and 120 in the second. I employed a rectified linear unit (ReLU) as the activation function in each layer and applied it to the normalized input data (since ReLU passes only values greater than zero). For the deep neural network models, I used 80% of the randomly selected data samples for training the model and the remaining 20% for the cross-validation process; the ratio was the result of a trial/error experience within the hyperparameter tuning. After each epoch, I monitored the performance of the model to ensure that the training process stopped with minimal validation loss to prevent overfitting. For the extra-trees models, we used 355 trees with a depth of 20. It is worth noting that we determined the hyper-parameters for both models after conducting comprehensive trial-and-error experiments. I implemented the deep learning algorithm in the Keras environment with the TensorFlow backend (60) and the extra-trees method in the Scikit-learn environment (73).

For all models, including the ensembles (i.e., the DNN model, the Extra-Trees regression model, the uniform model, and the merged model), I used a three-year training period of 2014-2016 and then used the trained models to generate predictions for all 25 stations across Seoul for the year 2017. For any given station, the ozone concentration of each day (24 hours) is predicted by these models using input observations from the previous day (see section 3-2). Since the training data of the ensemble, uniform, and merged models included samples from all the stations, I trained them only once to build a generalized predictive model. So, each of these generalized models was used to predict all 25 stations at ones. I also trained two station-specific DNN and extra-trees models for each station using only one station's input and output data. I compared the results of these station-specific models to those of generalized models.

One immediate advantage of using the generalized models over station-specific models is the computational time. For predicting a network of monitoring stations similar to Seoul with 25 stations, the generalized model will be trained ones (for a few minutes), and predict for all stations within less than a second. For station-specific, on the other hand, each machine learning modes will be trained individually, and predict the ozone concentration afterward. While the training process of one station-specific (i.e., DNN model) takes less than 10% of one generalized model (DNN-DNN ensemble model), the overall processing time to predict all 25 stations will be around three times faster using a generalized model.

4.3 Results and Discussion

4.3.1 General statistical comparison

For the evaluation of the models, we used the index of agreement (IOA) to compare the performance of different models. Figure 4.2 presents a comparison of the model-measurement statistics averaged over all the NIER stations for all of the proposed models in this study. This figure shows a box plot of the daily index of agreement (IOA) for each month of 2017 for all models. As shown, the performance of the models varies from month to month. Generally, the ensemble models provided more accurate predictions with greater daily IOAs in a majority of the months, particularly during the high ozone months (April-September). Since the number of days with lower ozone peaks is significantly higher than that with higher ozone peaks, the original models were relatively undertrained for predicting high ozone episodes. The ensemble models, by contrast, take both low and high ozone episodes into account by assigning equal importance in the second step. The regularized models also generated more accurate predictions than their original counterparts (including the ensemble and uniform models). The regularization process provided more samples for predicting days with high ozone peaks (over 90 ppb). As a result, the prediction biases of such days decreased, and the IOA increased.

The monthly median of the daily IOA of all models (black dots in Figure 4.2) was close to or greater than 0.8 between March and October, indicating the satisfactory performance of the models during these months. During 2017, the best performance was observed in May and September, with the highest monthly mean (blue dots in Figure 4.2) and monthly median of daily IOA between 0.83-0.88. The worst performance was observed during the winter months (i.e., December, January, and February,) with the lowest mean and median of daily IOA=0.56-0.75. In July, although most models found that the monthly median of the daily IOA exceeded 0.8, the range of the daily IOA was relatively wide, resulting from the "misprediction" of days with relatively high or low ozone peaks that the model significantly underpredicted and overpredicted. Since July is the onset of the monsoon season in the Korean Peninsula, this anomaly can be explained by the scattered rain and thunderstorms during this month. Figure 3.5 from Chapter 3 illustrates the hourly level of precipitation during the different weeks of

2017. The figure shows that the precipitation level increased dramatically during the weeks of July.

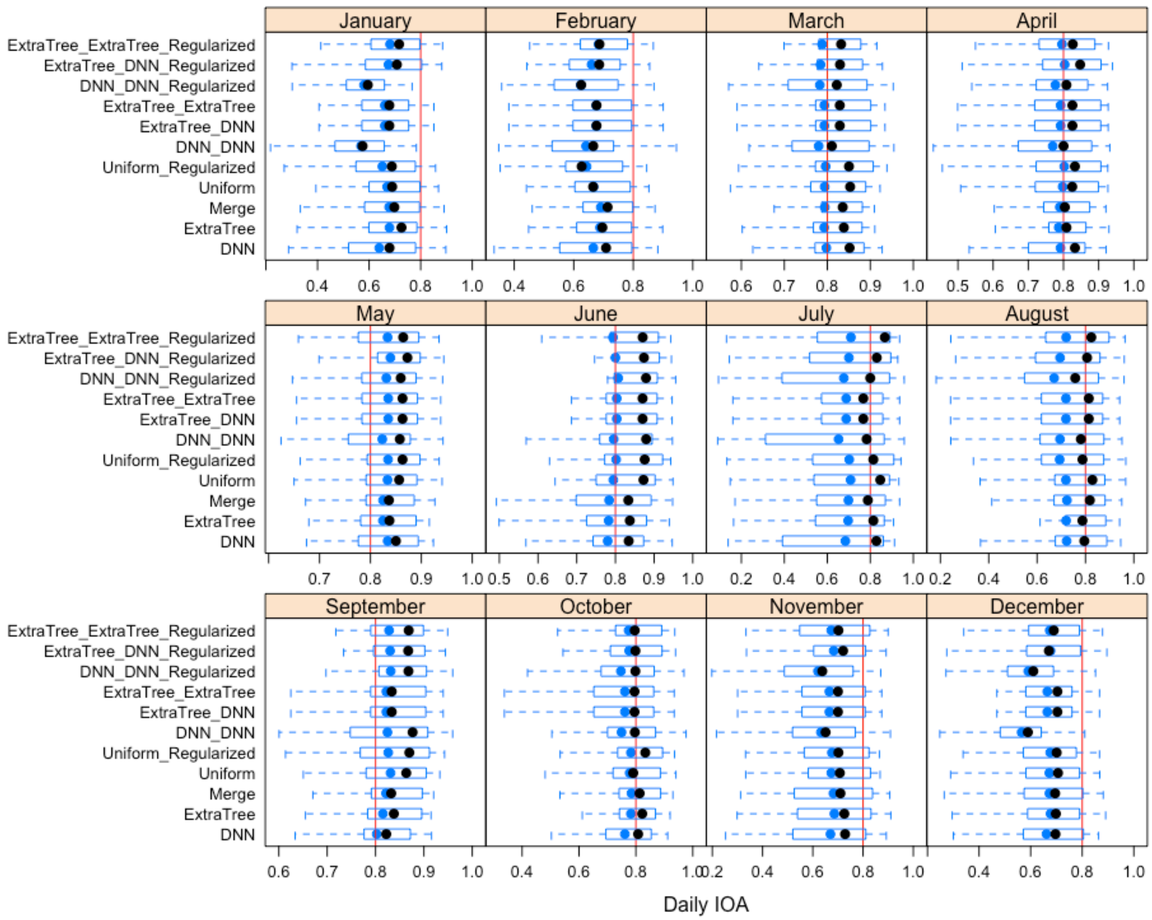


Figure 4.2. Box plot of the daily IOA of various real-time ozone forecasting systems for the months of 2017, averaged over 25 NIER stations in Seoul, South Korea. Blue dots represent the mean and black dots the median. The red vertical lines indicate IOA=0.8 as a reference for comparing the models in different months.

Figures 4.3 and 4.4 illustrate Taylor diagrams showing the performance of all of the models used in this study during different months (Figure 4.3) and under different conditions (Figure 4.4). The diagram shows how the statistics of the performance of the three complementary models vary simultaneously. Generally, the closer the point is to the "observed" value (shown with the point on the horizontal axis), the better the performance of the model is. These statistics are the correlation coefficient (r), the standard deviation (sigma), and the centered root-mean-square error. Figure 4.3 shows that, like the IOA, the correlation coefficient and the RMSE were comparable, except during July, among all models. The standard deviation, however, varied markedly among the models, particularly between April and August. Among all models, the ensemble models, compared to the base models (DNN and Extra-Trees), provided greater IOA, confirming the effectiveness of the ensemble process. Similar to the IOA, the sigma of the regularized models, unlike the original models, was closer to the observed values. During July, both DNN-DNN models (the original and the regularized) provided lower IOA than the other models. One explanation for this finding is that the DNN model may have "overstretched" the results of the high-ozone peak model while capturing the very high ozone episodes in the second step of the ensemble process. This overstretching process could have led to several overestimated predictions during the summer months [3]. Another explanation might be related to the many overpredictions of the DNN ensemble models during the night (for DNN-DNN) and late afternoon (for DNN-DNN regularized).

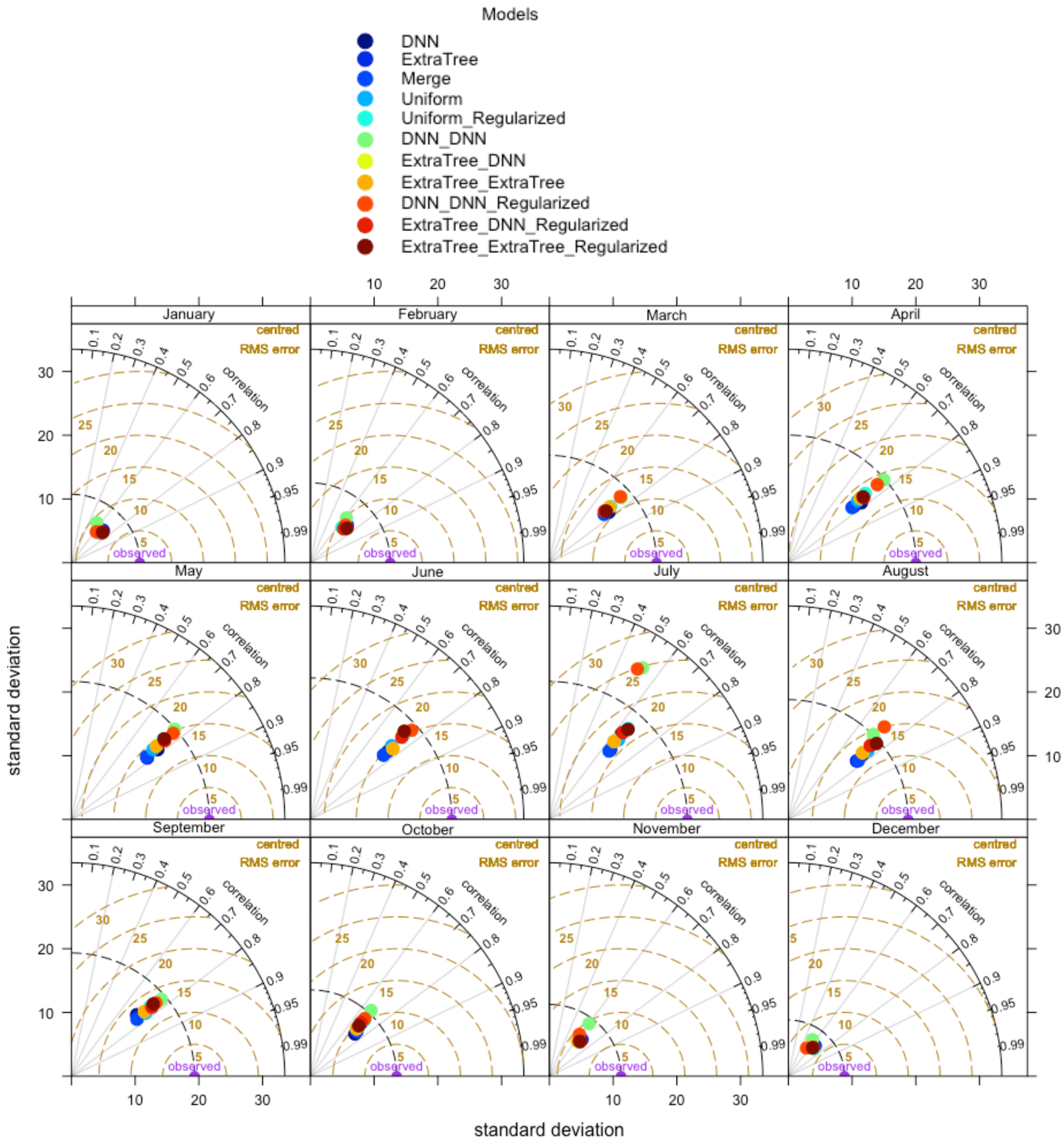


Figure 4.3. Taylor diagram showing the performance of various real-time ozone forecasting systems during different months averaged over 25 NIER stations in Seoul. The model-measurement statistics shown here are the Pearson correlation (r), the standard deviation (sigma), and the centered RMSE.

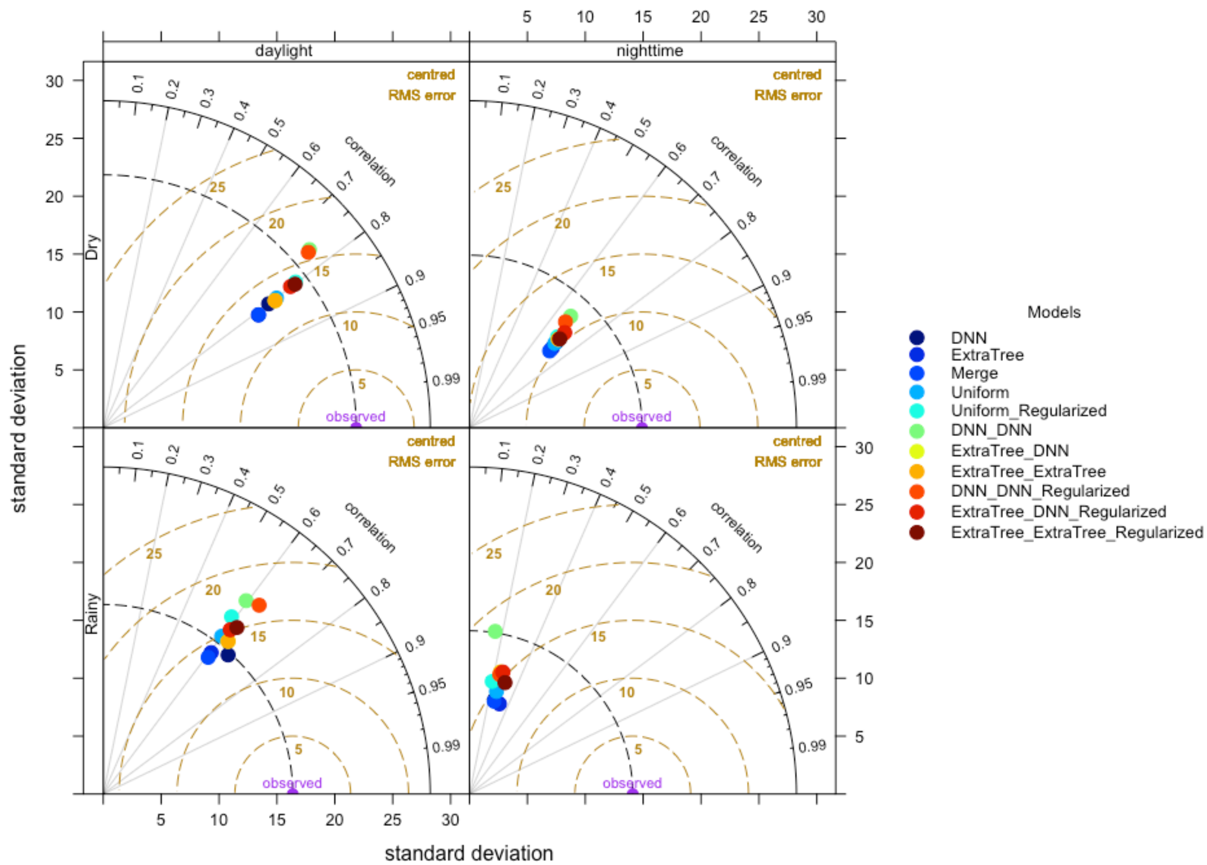


Figure 4.4. Taylor diagram showing the performance of various real-time ozone forecasting systems during daytime/nighttime and rainy/dry days averaged over 25 NIER stations in Seoul. The model-measurement statistics shown here are the Pearson correlation (r), the standard deviation (sigma), and the centered RMSE.

The Taylor diagram in Figure 4.4 compares the performance of all the models during the daytime and the nighttime on dry and rainy days for all of 2017. All models provided greater IOA not only during the daytime than during the nighttime but also during dry days than during rainy days. One explanation for this finding is that during the nighttime (and/or the rainy days), levels of the input predictors (meteorological parameters and air pollutants) were considerably lower than their corresponding values during the daytime (and/or dry days). Figure 4.5, diurnal variation of ozone, shows that the ozone concentrations were significantly lower during rainy days than during dry days. In addition, the number of dry/daytime hours was higher than the number of rainy/nighttime hours. Thus, we conclude that the models were "biased" when predicting higher values; that is, they more accurately predicted higher values than lower values. Another explanation could have been the lack of certain important predictors, such as the cloud fraction, solar radiation, and ozone precursors. These parameters were not measured in Seoul during the time period of this study (2014-2017).

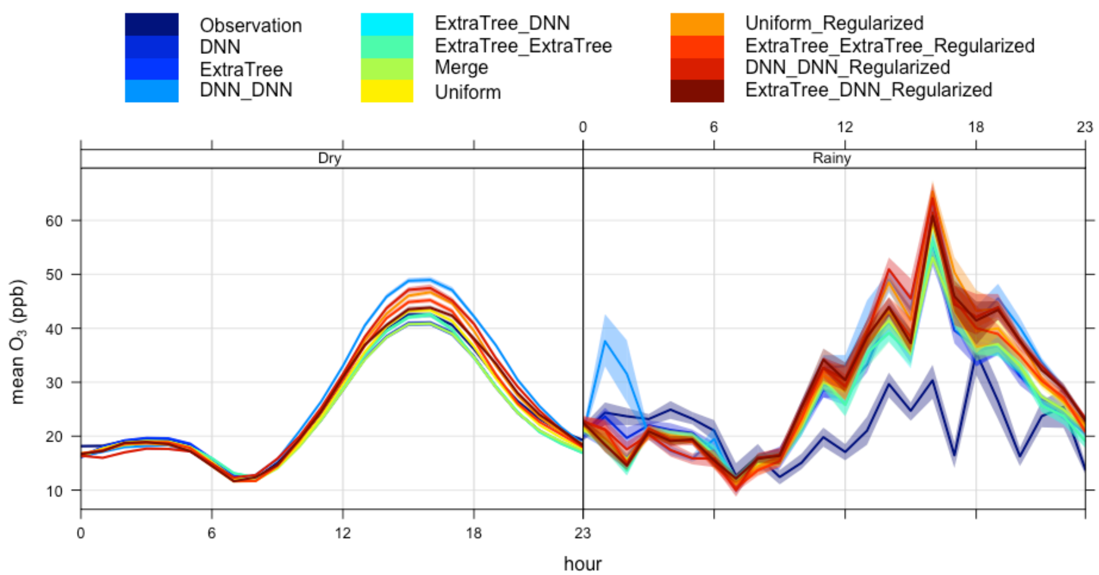


Figure 4.5. Average hourly mean ozone concentrations in 2017 for different models on dry/rainy days.

4.3.2 Model performance for capturing maximum daily values

Figure 4.6 compares the performance of all modeling approaches with respect to the daily maximum ozone for the different months of 2017. The results of nearly all models exhibit a notable underprediction of high ozone peaks during all months. For the winter months (i.e., December, January, and February), this underprediction by the proposed models follows a similar trend. The figure shows that the base models (Extra-Trees and DNN) predicted peak ozone more accurately during these months. For the rest of the year, the DNN-DNN ensemble model was the most accurate at capturing high ozone peaks. This relative success, however, led to several mispredictions of the high peaks. Figure 4.6 illustrates the range of the predicted peaks, which were lower than the observed values, indicating that the predictions of the ozone peaks during the ozone season (April-September) by the ensemble model were less variable than the target values. As a result, although a greater number of days with high ozone peaks were estimated accurately, several “false alarm” predictions actually occurred.

Figure 4.6 also reveals that the merged model performed the worst during most of the months. This finding relates to the redundancy within the inputs of the merged model. Since we measured the meteorological variables in Seoul at only one official station (KMA station #108), seven input variables (out of nine) were the same for all 25 station locations across the Seoul area. This redundancy caused a deficiency in the training process of the machine learning model.

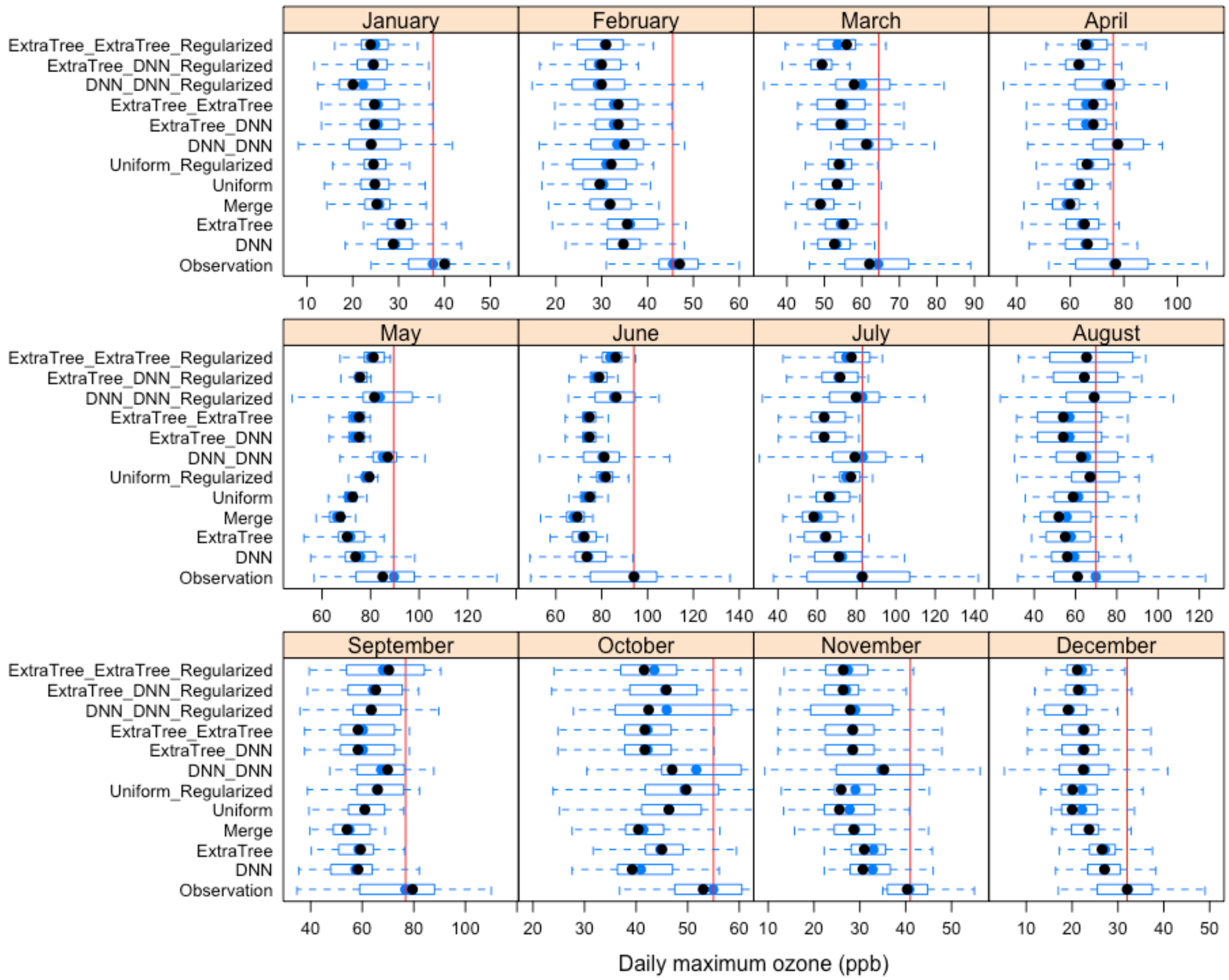


Figure 4.6. Box plot of the daily maximum ozone for various real-time ozone forecasting systems during each month of 2017. The values are averaged over 25 NIER stations in Seoul. Blue dots represent the mean and black dots the median. The red vertical lines indicate the monthly mean for the daily ozone maximums (or maxima) as a reference for comparing the models in different months.

Figure 4.7 shows the performance of the models with respect to the range of daily maximum ozone during the different seasons. Most of the underpredictions occurred on days in which the ozone peak exceeded 70 ppb. The underpredictions were more pronounced during the fall (i.e., September, October, and November). After all, the frequency of high peak episodes was less during the fall than during the spring (i.e., March, April, and May) and summer (i.e., June, July, and August). Hence, the models were undertrained for such high peak episodes. Almost all of the ensemble models, however, showed improvement by minimizing the underprediction of the high peaks, especially during the summer and the fall. Compared to the base models, the ensemble models displayed an improvement of as much as ~30 ppb on some days and ~16 ppb in some months. Among the ensemble models, the DNN-DNN models produced a higher IOA than others when the ozone peaks exceeded 70 ppb. For the low-ozone episodes (i.e., with maximum daily ozone of less than 50 ppb), the base models yielded a higher IOA.

Figure 4.8 compares the statistical performance (i.e., the IOA and the correlation coefficient r) of the various models with regards to the range of daily maximum ozone values. Both the IOA and r in all of the models exhibited higher values for the higher ozone peaks. One explanation for this finding is that for ozone concentration peaks with higher values during the daytime (i.e., higher than 70 ppb), the daily time series usually follows a smooth pattern (at some level with periodic behavior) with predictable locations of daily highs and lows. After all, stable meteorological conditions led to the efficient formation of more ozone during the daytime and ozone dilution during the nighttime. Thus, the ozone concentration could be more effectively explained by current meteorological parameters, which led to a more accurate hourly prediction, and hence, higher IOA and r values for these days. For low ozone episodes, by contrast, the time series varies more frequently without any notable pattern, which results in lower model performance. As a result, the models generally yielded a higher IOA during the spring and the summer than they did during the fall and the winter.

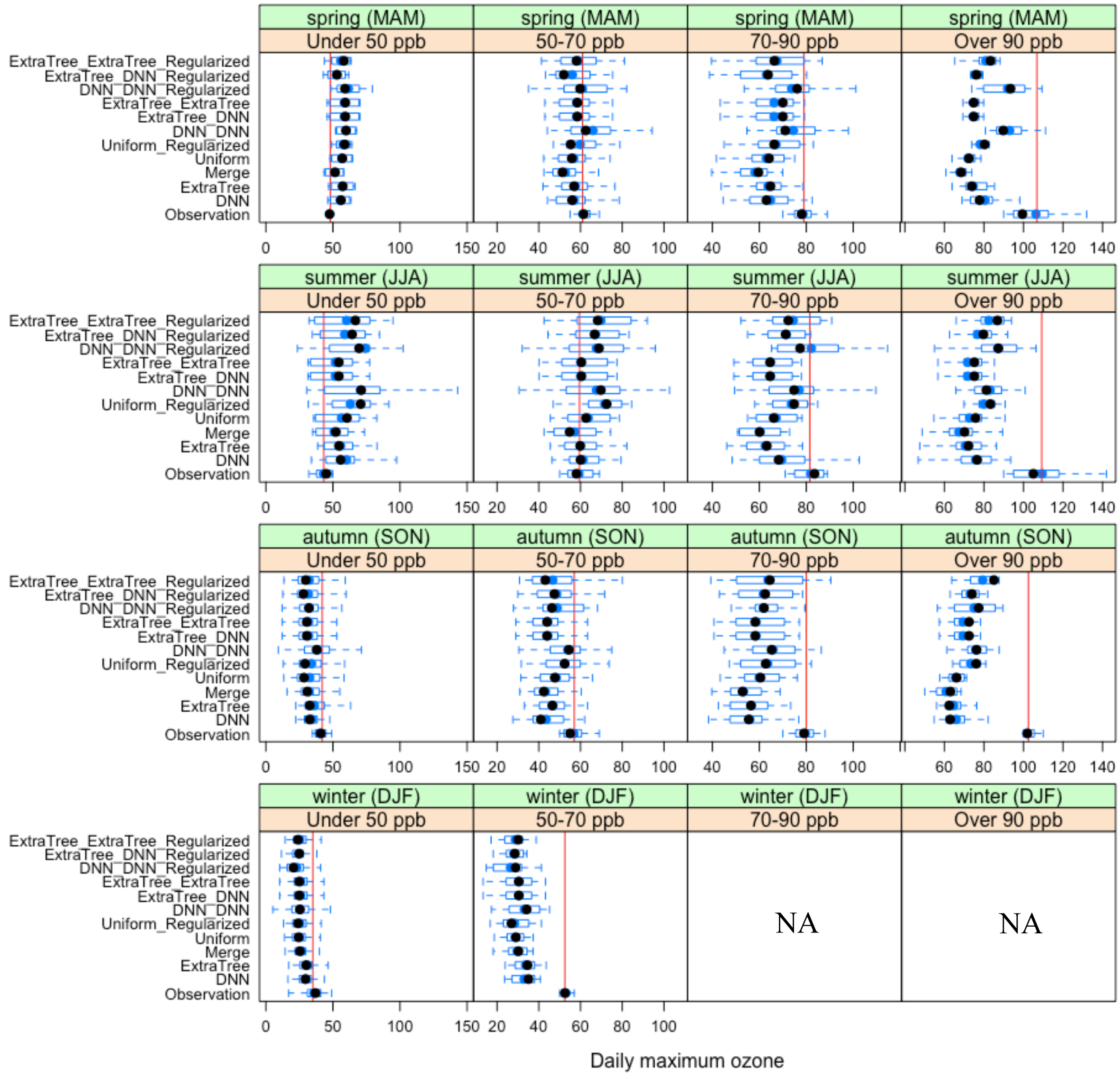


Figure 4.7. Box plot of the daily maximum ozone for various real-time ozone forecasting systems with respect to the ranges of daily maximum ozone for all seasons in 2017. The values are averaged over 25 NIER stations in Seoul. Blue dots represent the mean and black dots the median. The red vertical lines indicate the monthly mean for the daily ozone maximums (or maxima) as a reference for comparing the models in different ranges of the daily maximum ozone.

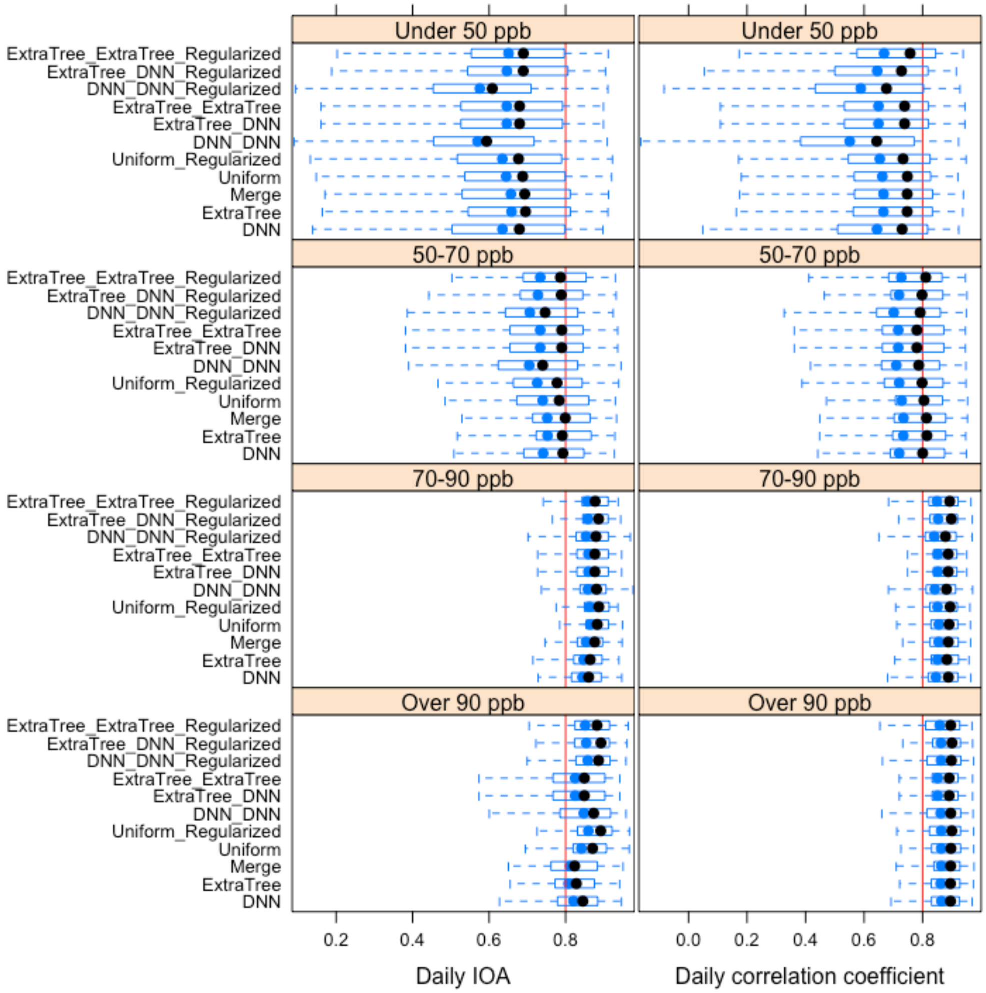


Figure 4.8. Statistical performance of all the modeling approaches with respect to the ranges of daily maximum ozone for 2017. The values are averaged over 25 NIER stations in Seoul. Blue dots represent the mean and black dots the median. The red vertical lines indicate IOA=0.8 and r=0.8 as a reference for comparing the models.

4.3.3 Categorical analysis for selecting the best model

The ensemble forecasting models have clear advantages over the base models. Since I used samples from all stations for the uniformization process, all of the ensemble models are generalized models, suggesting that one model can predict hourly ozone for all of the stations. The immediate advantage of a generalized model over the station-specific model relates to its shorter computation time. As a result, the generalized models are favorable for applications requiring prediction for multiple locations with similar characteristics. Although ensemble approaches have their own advantages and limitations, their suitability depends on the decision time period (e.g., month, season). If used appropriately, these approaches enable a more robust decision-making process, whether they are used individually or in combination. Thus, I developed an algorithm for selecting the most suitable models based on a categorical analysis of the training dataset.

As mentioned in Section 4.3.2, the performance of the models differed during different months. By investigating the training dataset, however, I suggested the use of one model over others for any given month. In order to find the optimal model, I calculated the percentage of the frequency of the daily maximum ozone concentration divided into five category levels for each month. These are the same sub-categories used in creating uniform distribution and regularization processes (See Section 4.2.4). Table 4.2 presents the percentage of the frequency of each category for each month calculated based only on the training dataset (2014-2016). Since the goal was to propose a real-time forecasting system, I drew no samples from the testing dataset (2017) for this analysis. As expected, I found a few dominant categories whose ozone concentrations were lower during the fall and winter than during the spring and summer.

Table 4.2. Percentage of the frequency of the daily maximum ozone concentration for several categories of the ozone magnitude based on the training dataset. The numbers shown as the reference in each bin are ozone concentrations in ppb. With a threshold of 10%, the dominant bins are shown in each month in gray-shaded cells.

Categories (in ppb)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Train < 25	61	22	5	1	0	0	8	2	4	11	44	70
Train 25-50	39	75	69	44	16	16	35	38	52	74	53	30
Train 50-70	0	3	24	41	44	38	27	29	30	14	3	0
Train 70-90	0	0	3	12	27	30	20	21	12	1	0	0
Train > 90	0	0	0	2	14	16	10	10	2	0	0	0

Table 4.3 presents the monthly IOA of the models in different months and those that exhibited the best performance each month. During the cold months (winter and fall), the base models (particularly the Extra-Tree model) exhibited a higher IOA, but during the warm months (spring and summer), the regularized models were comparably better. For the training data with fewer dominant categories, the Extra-Tree model was capable of extracting information from the input features. By contrast, the ensemble models, which used the high-ozone peak model in their second step, overstretched the results. No high-ozone peak episodes, however, occurred during the cold months. As a result, this extra process degraded the results of the base models. During the warm months (March to September), the smallest bin (an ozone concentration of less than 25 ppb) was a non-dominant category. In addition, the relative hourly values of the input variables were usually higher during these months, which accounts for the nonlinearity of the relationships between the inputs and outputs. I, therefore, conclude that one can obtain more accurate predictions using a more complicated method (i.e., the ensemble model).

Table 4.3. Monthly IOAs of the modeling approaches compared to that of the model selected by the categorical analysis of the training dataset. Here, "ET" represents the Extra-Trees model and 'R' the regularized model. For each month, the models with the highest monthly IOAs are indicated by bold font, and the best models found by the categorical analysis appear individually in a box cell.

Models	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ET	**0.80**	**0.79**	0.83	0.82	0.83	0.82	0.78	0.84	0.83	**0.82**	**0.78**	**0.79**
DNN	0.75	0.77	**0.84**	**0.85**	0.85	0.82	0.78	**0.85**	0.82	0.81	0.77	0.76
Merge	0.79	**0.79**	0.83	0.82	0.83	0.82	0.77	**0.85**	0.84	**0.82**	0.76	**0.79**
Uniform	0.78	0.76	0.83	0.83	0.84	0.84	0.79	0.84	0.85	**0.82**	0.77	0.78
Uniform R	0.75	0.73	0.83	0.84	0.85	0.85	**0.80**	0.82	0.85	**0.82**	0.77	**0.79**
ET-ET R	**0.80**	**0.79**	0.83	0.84	0.85	0.85	**0.80**	**0.85**	0.85	0.81	0.76	0.77
ET_DNN R	0.78	0.76	0.82	**0.85**	**0.86**	0.85	0.79	0.83	**0.86**	0.81	0.77	0.78
DNN_DNN R	0.72	0.76	**0.84**	**0.85**	**0.86**	**0.86**	0.68	0.81	**0.86**	**0.82**	0.74	0.68
ET-ET	0.79	0.77	0.83	0.83	0.85	0.84	0.78	0.84	0.84	**0.82**	0.75	0.78
ET_DNN	0.78	0.78	0.83	0.84	0.85	0.85	0.78	**0.85**	0.85	0.81	0.74	0.75
DNN_DNN	0.70	0.76	0.83	0.84	0.85	0.84	0.68	0.82	**0.86**	0.81	0.77	0.72

In light of the above findings, we proposed an algorithm to select the best model for each month using categorical analysis only, independent of the station location. Figure 4.9 presents a flowchart of the proposed algorithm. I chose the Extra-Trees model to represent the base models and the regularized ensembles of ExtraTrees-ExtraTrees and DNN-DNN because their performance was more robust in different months than that of the other models (see Table 4.3). Following the algorithm proposed in Figure 4.9, we created a time-series of the "best model" for all of 2017. Figure 4.10, which shows the weekly mean and max time series of the "best model" prediction for all stations in Seoul, illustrates the reasonable performance of the best model in most of the stations. Compared to the selected models, the optimal model yielded higher values of yearly IOA (with an increase in the IOA value of 0.3% to 2.4%), and compared to the base models (Extra-Trees and DNN), the optimal model showed an increase in the yearly average IOA at almost all stations.

4.4 Summary

The worldwide deterioration of air quality poses a unique challenge for decision-makers in their efforts to reduce the effects of air pollution on human health. Effectively addressing this situation requires a fast, comprehensive, and reasonably accurate forecasting system that will support robust decision-making processes. To serve this purpose, I have developed a total of eleven machine-learning, real-time hourly ozone forecasting models using multiple techniques, including the ensemble approach, uniformization, and regularization. I trained these models to predict hourly ozone concentrations for the following day using observations of NOx, ozone, and meteorological variables from the previous day.

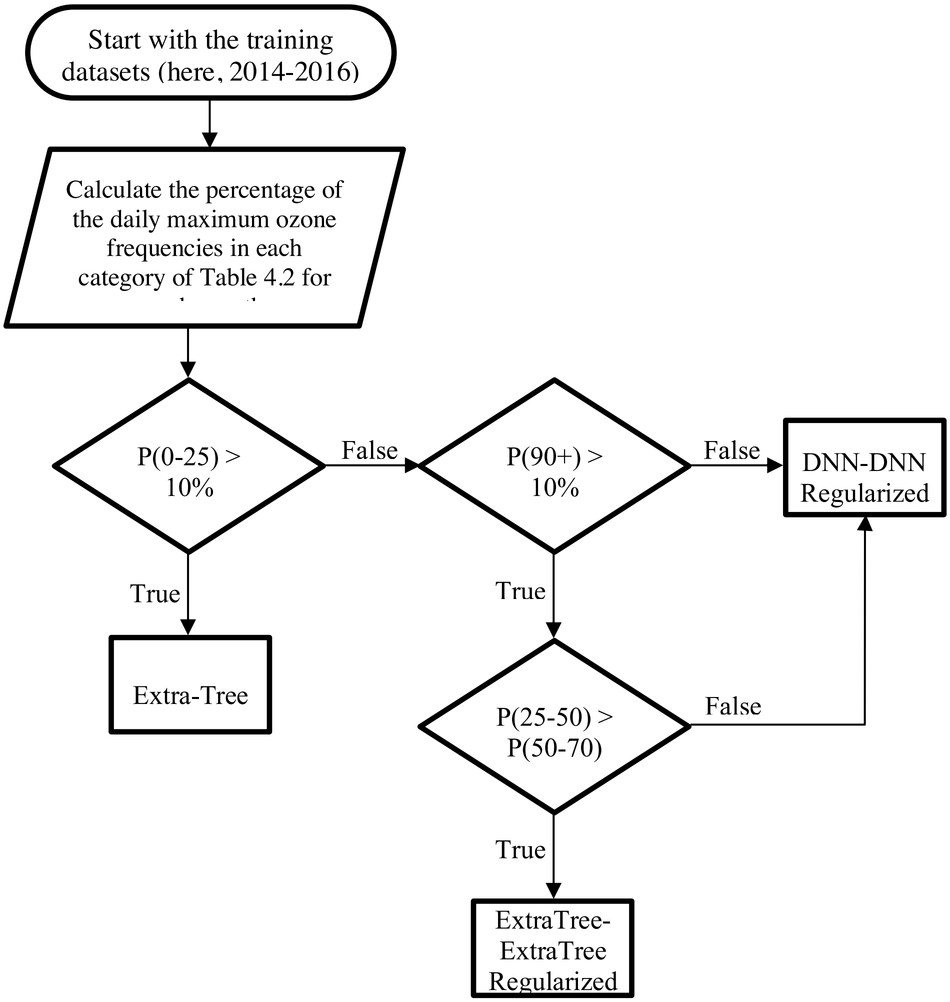


Figure 4.9. The proposed flowchart for selecting the best model based on the percentage of the frequency of the daily maximum ozone concentration for each month obtained from the training dataset. P(A-B) represents the percentage of the frequency of daily maximum ozone equal to or greater than A ppb and less than B ppb.

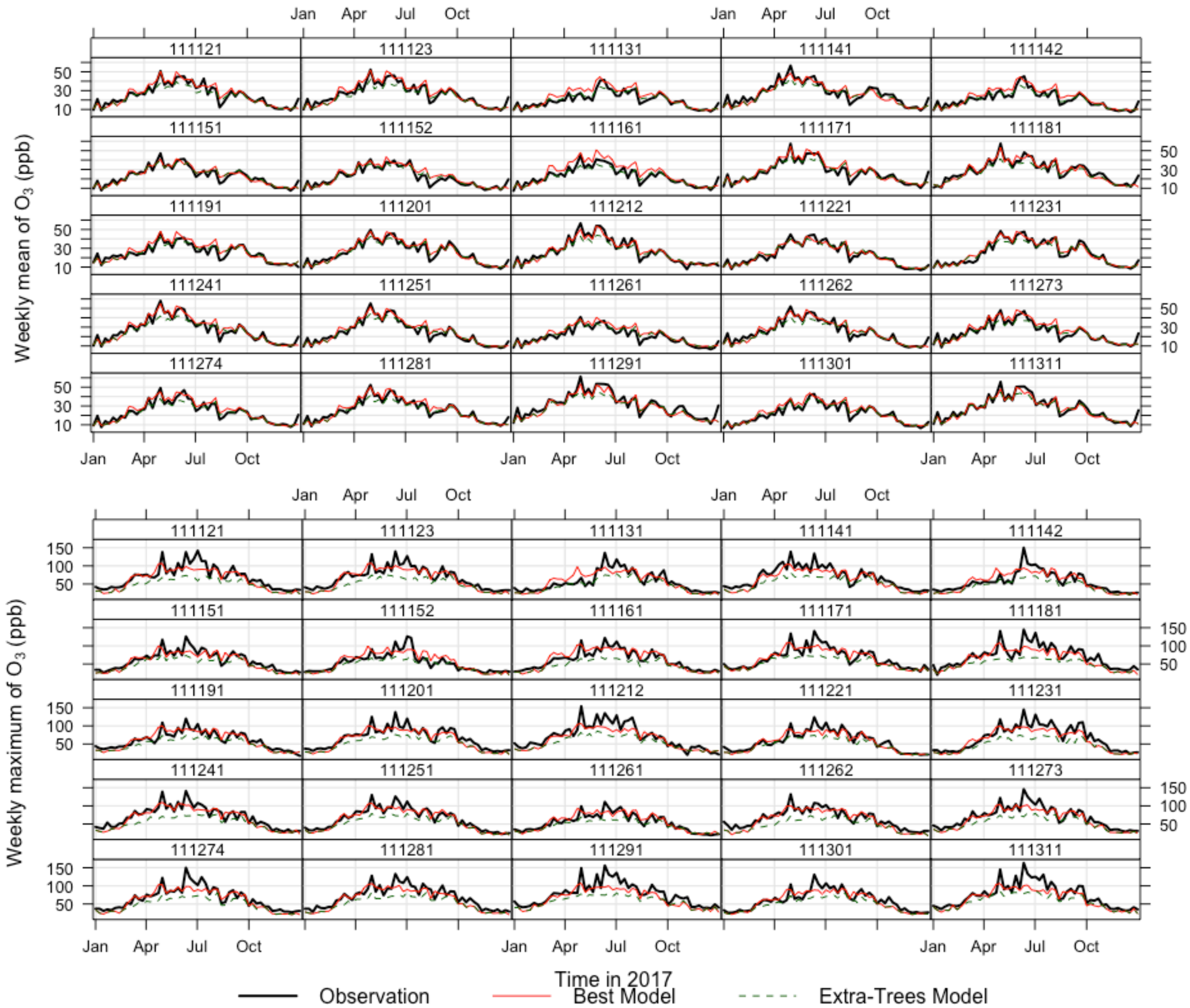


Figure 4.10. Weekly mean (top) and max (bottom) time series of the best model selected using the algorithm proposed in Figure 4.9 for all 25 NIER stations in Seoul.

To develop the base models, I used two powerful machine-learning models: the extremely randomized decision tree (or Extra-Trees) and a deep neural network (DNN). I designed, trained, and tested the base models to predict the hourly ozone concentration at each station based on its input and output variables. Both models showed reasonable performance, with yearly IOAs in the range of 0.83 to 0.89, and yearly correlation coefficients in the range of 0.72 to 0.80. High ozone episodes, however, were significantly underpredicted, especially during the high ozone season (April-September). One explanation for this underprediction was the considerably fewer occurrences of high ozone episodes than low ozone episodes. Therefore, the base models were relatively undertrained to predict days with high ozone peaks (more than 90 ppb).

To address the underprediction issue of the single-model forecasts, I combined these two models in the form of the non-regularized and regularized ensemble models to conduct six two-step data ensemble models. In the first step of the ensemble modeling approach, I trained two models (low-ozone and high-ozone peak models) using the prepared sample inputs according to the daily maximum values of the target ozone (i.e., observed ozone) and then organized the sample inputs in a uniform distribution of ozone peaks by assigning an equal number of samples to each of the five ranges of ozone concentrations. In the second step, I combined the results of the two models from the first step and predicted the final hourly ozone concentrations for the following day. For the regularization process, I added more training samples targeting daily maximum ozone over 90 ppb. The results of the ensemble models showed sensible improvements in the IOA of as much as 0.05 during the high ozone season and reduced the underprediction bias by as much as 31 ppb. In July, however, the DNN-based ensemble models performed poorly for some stations. Generally, the deployment of ensemble models showed clear advantages over the base models. The advantages were more pronounced when higher values of ozone concentrations were expected. In addition to their statistical advantages, the ensemble models performed faster than the base models for predictions in multiple locations. Because of their robust performance, the ensemble models are more suitable for applications to urban areas with multiple observation stations.

I also developed three more generalized models in which we employed only the Extra-Trees method to predict outputs using three different training datasets. One training dataset consisted of a pool of input samples that we merged from all of the samples collected across all of the stations. Compared with the station-specific Extra-Trees model, this model underperformed both in terms of the yearly IOA (with an average 0.02 IOA reduction) and bias for high-ozone peaks (with an average 5 ppb absolute bias increase). For the other two models, I used regularized and a non-regularized uniformly distributed training dataset. The results showed that the performance of the uniform models was similar to that of the station-specific model, with marginal improvement in the IOA in most of the stations during the high ozone season (April-September). I also found that the model performance significantly decreased when the target day was rainy, the result of undertraining the machine learning model because of fewer rainy days than dry days in the training dataset. The decrease in the correlation coefficient was even more noticeable during the nighttime (from ~0.7 to ~0.3) than it was during the daytime (from 0.8 to 0.65).

By performing a categorical analysis based on the training dataset, I ultimately proposed an algorithm for selecting the most suitable model for each month. Utilization of the proposed algorithm resulted in greater accuracy (up to 0.045 in station-wise yearly IOA) by enabling the use of a customized, more sophisticated modeling approach.

I attribute the superiority of the proposed ensemble approaches in this study to the machine-learning architecture and the data-ensemble technique. While the former can effectively extract the nonlinear, stochastic nature of the atmosphere, the latter can modify the notable prediction bias caused by an improper training dataset for a target objective (i.e., capturing higher ozone peaks). Machine-learning ensemble models delivered reasonable accuracy for real-time hourly prediction with less (around three times) computation time compared to the station-specific machine-learning models. The days with high ozone peaks, however, were still significantly underpredicted on some days (see Figure 4.7) for several reasons: (1) The number of samples (collected from the NIER monitoring stations) was limited to a four-year collection. Usually, a larger training set will result in a more accurate model. (2) In a rapidly changing city such as Seoul, the pollutant levels and trends change over time, generally the result of changes in the traffic volume, regulations, and related policies. Therefore, the target year in this study (i.e., 2017) might have differed from that of the training dataset (i.e., 2014-2016). (3) Because of the insufficient availability of predictors (from NIER monitoring stations), high ozone peaks were not adequately addressed. For instance, the cloud fraction, solar radiation, and the planetary boundary layer (PBL) height along with ozone precursors such as OH are among the essential predictors of ozone levels, particularly high-ozone episodes [7-8]. To address this issue, one could use the results of physical models such as CMAQ as input variables, especially during the ozone season. This will be potentially discussed in the next Chapter.

CHAPTER 5. CMAQ-CNN: A NEW-GENERATION OF POST-PROCESSING TECHNIQUES FOR CHEMICAL TRANSPORT MODELS USING DEEP NEURAL NETWORKS

5.1 Introduction

The last several decades have witnessed a number of advances in chemistry transport models (CTM) for estimating regional air quality forecasts. These advancements have improved the accuracy of air quality advisory plans and promoted research in the physical understanding of the ambient atmosphere (11). Through complex computation schemes and specific initial and boundary conditions, CTM models simulate ambient air pollution concentrations by considering emission, transport and deposition mechanism, and other physical processes (17). CTMs provide temporally and spatially customized forecasts of regional air quality episodes of pollutants that are not monitored and use a physically-based knowledge of atmospheric processes that cannot be conducted by other modeling approaches (e.g., statistical approaches). As a result, they do not require a large quantity of measured data to attain reasonable forecasting accuracy.

Regarding the operational real-time air quality forecasts, however, CTMs have several drawbacks. For one, they demand sound knowledge of pollution sources and processes governing their evolution in the atmosphere (11). Thus, understanding the sources of bias and developing strategies for mitigating them is both complex and costly. In addition, because of imperfect simulation settings and simplified physical processes, CTMs exhibit significant model-measurement errors, especially during high concentration episodes (10). Finally, the accuracy of CTMs depends on the accuracy of meteorological predictions, emission estimates, initial and boundary conditions, and other model inputs (55). Thus, biases in such parameters and uncertainties in CTM inputs can propagate into the final predictions. Particularly in predictions of ozone concentrations as secondary air pollution, which are sensitive to meteorological parameters and chemical precursors, such biases can result in considerable inaccuracy (13). To reduce model errors, therefore, most operational air quality forecasting systems require some level of post-processing before unveiling forecasts as a complementary reference to health advisory decision making (6, 15, 74).

The only representation of real-world atmospheric conditions is observed measurements. Thus, any post-processing approach should ensure that CTM forecasts are verified against observed measurements. When a long history of observations is used, the use of a data-driven bias correction model may produce more accurate forecasting by CTM (6, 15, 74). Common forms of such bias correction models are multi-variable linear regression models, usually referred to as model output statistics (MOS) corrections (74). To bias-correct a certain air pollutant, these models use long histories of CTM outputs and observations, preferably from a range of important factors such as meteorological parameters (75). Real-world applications of these models, however, are limited in several ways. Results from the linear structure of these models; that is, they are reliable for only the post-processing of regional forecasts; they cannot be generalized to an entire domain (75) (i.e., the continental United States, as studied here). In addition, to generate stable corrections, these models require no change in modeling processes (75).

In the past few years, however, CTM modeling processes have significantly improved, owing to the enhancement of our understanding of secondary processes in the atmosphere (18) (thanks to several comprehensive measurement campaigns) and advancements in the modeling of the feedback between weather and chemical processes (19) (i.e., two-way modeling). Another limitation of these models is that they bias correct CTM outputs using variables that are independent from the modeling biases. Bias correction (i.e., the process that attempt to address bias) and the modeling process (i.e., the process that generates the bias) are not physically connected, which is particularly important when CTMs perform accurately; thus, the post-processing level should be adjusted accordingly to avoid adverse effects (75) (i.e., producing less reliable forecasts than the original ones). To produce more accurate bias-corrected forecast, these models also rely on high-quality observations, which are limited in time, sparsely distributed in space, costly to collect, and contain a notable portion of missing values. Hence, bias correction is limited to modeling grids in which a long-running monitoring station is available (76, 77). Finally, even if historical observations take the form of “big data” for developing bias-correction models, the level of improvement in final forecasts has peaked (mostly due to the above-mentioned limitations), as several significant biases in high concentration episodes have not yet been addressed (See Chapters 3 and 4).

Precisely focused on addressing the limitations and the challenges that methods of forecasting ambient air quality in the real-time entail, I introduce a new generation of post-processing tools for real-time air quality forecasts that employ the physical process of a numerical model in conjunction with an advanced deep learning algorithm. Deep learning algorithms can be trained to approximate smooth, highly nonlinear functions, rendering them capable of analyzing nonlinear processes in the atmosphere. Here, I explore the use of a deep convolutional neural network (CNN) model (44) to develop a fast, robust, and accurate real-time bias-correction system. Such a system will enhance the predictions of state-of-the-art air quality forecasting models—the community multi-scale air quality model (CMAQ) (17) and the weather research and forecasting (WRF) (78)—that will benefit public users, scientific communities, and the federal government[24]. The proposed model, the CMAQ-CNN model, overcomes the limitations of conventional methods, supported by the results of this study.

5.2 Materials and Methods

Figure 5.1 displays the conceptual framework of the CMAQ-CNN model. To model the atmosphere as a physical system, we generally use numerical models that estimate the state of the atmosphere by taking multiple parameters into account. These parameters are estimated by the model itself at a previous step (e.g., vertical wind and trace gas precursors), other input parameters (e.g., the planetary boundary layer, a.k.a. PBL), or empirical or semiprimal schemes (e.g., aerosol chemistry). A computer simulation generates results for a 3D modeling domain in continuous time steps while it verifies them against observational data. Although this is a complex, knowledge-intensive process, the numerical model relies on simplified physics and thus produces inaccurate estimates of air pollution. With artificial intelligence (AI), we simplify the process by training the AI model with parameters from the numerical model (outputs of WRF and CMAQ) as input to map actual air pollution concentrations (here, ozone) from the observational network. In this way, we combine the physical intelligence of the WRF and CMAQ models while adding continuous feedback to the modeling process. Thus, the AI model (here, deep CNN) learns the dynamic of the model errors through an accelerated training-testing process. The simultaneous use of meteorology and chemistry models in the WRF-CMAQ model represents a significant advancement in the routine operational real-time air quality forecasting system, considerably enhancing our understanding of the underlying complex interplay of meteorology, emissions, and chemistry. Thus, this approach increases the likelihood of producing accurate, reliable estimates of critical variables in the atmosphere.

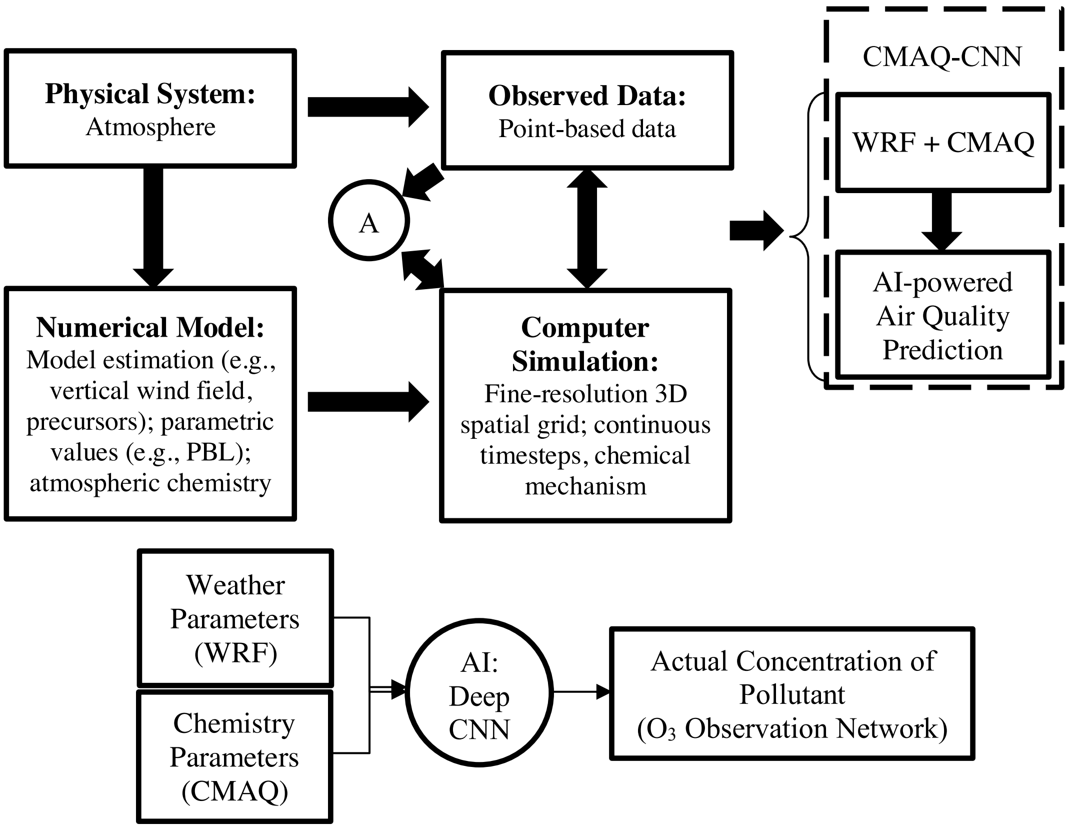


Figure 5.1. Conceptual framework of this study.

5.2.1 Numerical modeling module: WRF-CMAQ

Because of the increasing maturity of the physical processes and available data infrastructure (e.g., emission models, initial and boundary conditions) of CMAQ, we selected it to generate robust regional air quality predictions. Developed by the U.S. Environmental Protection Agency (EPA), CMAQ is a sophisticated atmospheric dispersion model with an active open-source development platform that combines current knowledge of the atmospheric sciences[25]. Representing several fields of physical and chemical sciences, the CMAQ-CNN model will likely become a prevalent forecasting system. To predict the meteorological parameters in the CMAQ-CNN model, this paper uses the National Center for Atmospheric Research (NCAR) Weather Research and Forecasting model (WRF) as the numerical weather prediction model and to develop our bias-correction model; we chose WRF because of its well-known history in both research and operational applications. We used the modeling data prepared by Choi et al. (2016) (79), Souri et al. (2017) (80), and Jeon et al. (2018) (81), who explained the configurations of both models in detail. This study focused on the months between April and October (215 days), considered the "ozone season," because of the more frequent higher ozone concentration events across the United States. We validated the modeling results against several observational sources, including the EPA AirNow network and the Texas Commission on Environmental Quality (TCEQ) Continuous Ambient Monitoring Station (CAMS) network.

5.2.2 Artificial intelligence module: Deep CNN

We used a five-layer CNN model with a number of filters and kernel sizes of 64 and 2, respectively, selected via a comprehensive validation process. We implemented the deep learning algorithm in the Keras environment with the TensorFlow backend (60) with Adam (63) as the optimizer function and adopted a cross-validation process with 20% randomly selected testing data to monitor the training process of the CNN model. Assuming the CNN was fully trained by the provided sample data, we stopped the training process when the validation loss was minimized. To create the output data, we selected 1,217 monitoring stations across the United States operated by the EPA, which monitored hourly ozone concentrations between 2011 and 2014. For the training/testing/validation dataset, we prepared each sample based on the information collected on each day, that is, outputs from the WRF and CMAQ models and observations. Table 5.1 lists the 33 input variables from the WRF and 13 input variables from the CMAQ used in the CNN model. Therefore, more than 780,000 (215 days in three years for 1,217 stations) samples were available to train the CNN model with 1,104 (33+13 variables for 24 hours per day) input features. Since the target was to predict (or bias correct) the entire day (24 continuous hours), we applied SOFT-IMPUTE (59) to the raw measured data to create a complete hourly dataset and then selected three years of modeling and observational history (2011-2013) to train and test the CNN model. To validate the bias-correction capability of CMAQ-CNN modeling system, we select modeling and observational history from 2014.

5.3 Results and Discussion

5.3.1 CMAQ-CNN Portability and generalizability

Owing to the inherent properties of neural networks such as low sensitivity to noisy and high variational data, they are a promising candidate for developing a portable real-time air quality forecasting model. Portability of the model, however, can be problematic when a single pre-trained model is applied to the entire U.S. domain; that is, the accuracy of CMAQ-CNN might be inconsistent temporally and geographically. To test the portability and generalizability of the CMAQ-CNN model, we developed two CNN models. To train the first (noted by CNN_Standalone), we used data samples available for each station. Thus, more than 1,000 station-specific models were developed to bias correct the entire domain. To train the second (noted by CNN_Generalized), we polled all available samples to apply a single trained model to all stations across the domain; that is, the second model was a portable, generalized model (the term "generalized" refer to the capability of the model to generate stable results within a spatial domain, which differs from the term used in the machine learning algorithm.)

Table 5.1. List of input parameters from the outputs of the WRF and CMAQ models.

Abb.	Variable Name (WRF)	Units	Abb.	Variable Name (CMAQ)	Units
PRSFC	surface pressure	Pascal	**NO_2**	Nitrogen dioxide	ppbv
USTAR	cell averaged friction velocity	m/s	**NO**	Nitrous oxide	ppbv
WSTAR	convective velocity scale	m/s	**O_3**	Ozone	ppbv
PBL	PBL height	m	**NO_3**	Nitrates	ppbv
MOLI	inverse of Monin-Obukhov length	1/m	**OH**	Hydroxide	ppbv
HFX	sensible heat flux	Watts/m^2	**HO_2**	Hydroperoxyl	ppbv
QFX	latent heat flux	Watts/m^2	**N_2O_5**	Dinitrogen pentoxide	ppbv
RADYNI	inverse of aerodynamic resistance	m/s	**HNO_3**	Nitric acid	ppbv
RSTOMI	inverse of bulk stomatal resistance	m/s	**FORM**	Formaldehyde	ppbv
TEMPG	skin temperature at ground level	k	**ALD2**	Aldehyde	ppbv
TEMP2	temperature at 2 m	k	**ISOP**	Isoprene	ppbv
Q2	mixing ratio at 2 m	kg/kg	**XYL**	Xylene	ppbv
WSPD10	wind speed at 10 m	m/s	**TOL**	Toluene	ppbv
WDIR10	wind direction at 10 m	deg			
GLW	longwave radiation at ground level	Watts/m^2			
GSW	solar radiation absorbed at ground level	Watts/m^2			
RGRND	solar rad reaching sfc	Watts/m^2			
RN	nonconvec. pcpn per met TSTEP	cm			
RC	convective pcpn per met TSTEP	cm			
CFRAC	total cloud fraction	-			
CLDT	cloud top layer height (m)	m			
CLDB	cloud bottom layer height (m)	m			
WBAR	avg. liquid water content of cloud	g/m^3			
SNOCOV	snow cover	decimal			
VEG	vegetation coverage (decimal)	decimal			
LAI	leaf-area index	-			
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^3			
SOIT1	soil temperature in top cm	K			
SOIT2	soil temperature in top m	K			
SLTYP	soil texture type by USDA category	-			

Figures 5.2 and 5.3 compare the accuracy of the two CNN models as well as their ability to post-process CMAQ results. As a result of the high percentage of missing data in

observations reported in Oregon and Iowa, the standalone models were severely undertrained, so their results were not included in the comparison. Figure 5.2 indicates that while the improved accuracy of both CNN models for almost all states was similar, the generalized model provided more stable post-processing enhancement in most states. This finding suggests that after the generalized model was applied, the greater range of accuracy indicated greater improvement than that of the standalone model. The statistics in the Taylor diagrams of Figure 5.3, which are the correlation coefficient (r), the standard deviation (sigma), and the centered root-mean-square error (RMSE), vary simultaneously, revealing the similarity between the two CNN models in ozone forecasting in different months. Generally, the closer the point is to the "observed" value (shown with a point on the horizontal axis), the stronger the model performance is. While the standalone model is slightly closer to the observed values than the generalized one (e.g., August), their difference is negligible.

Once a portable, generalized air quality forecasting model is developed, it will be beneficial in various forms: (i) It will reduce the computational time on the order of hundred times, a critical period in which high concentration episodes should be predicted; (ii) when new observational data (e.g., new locations, new months) are ready, it can be used as a pre-trained model, which will reduce the time for its update; (iii) it can be employed for constructing and testing other numerical models (e.g., WRF, Lagrangian models); (iv) it analyze nonlinear processes and exceptional phenomena (e.g., high ozone episodes) occurring in the atmosphere in idealized experiments; (v) it can incorporate customized data assimilation techniques using the results of the generalized post-processing model for the modeling grids for which observations are not available; and (vi) it is trained by a numerical model variable and contains a valuable source of pre-trained knowledge that can be used in inverse problems. Therefore, we chose the generalized CNN model to represent our proposed CMAQ-CNN model.

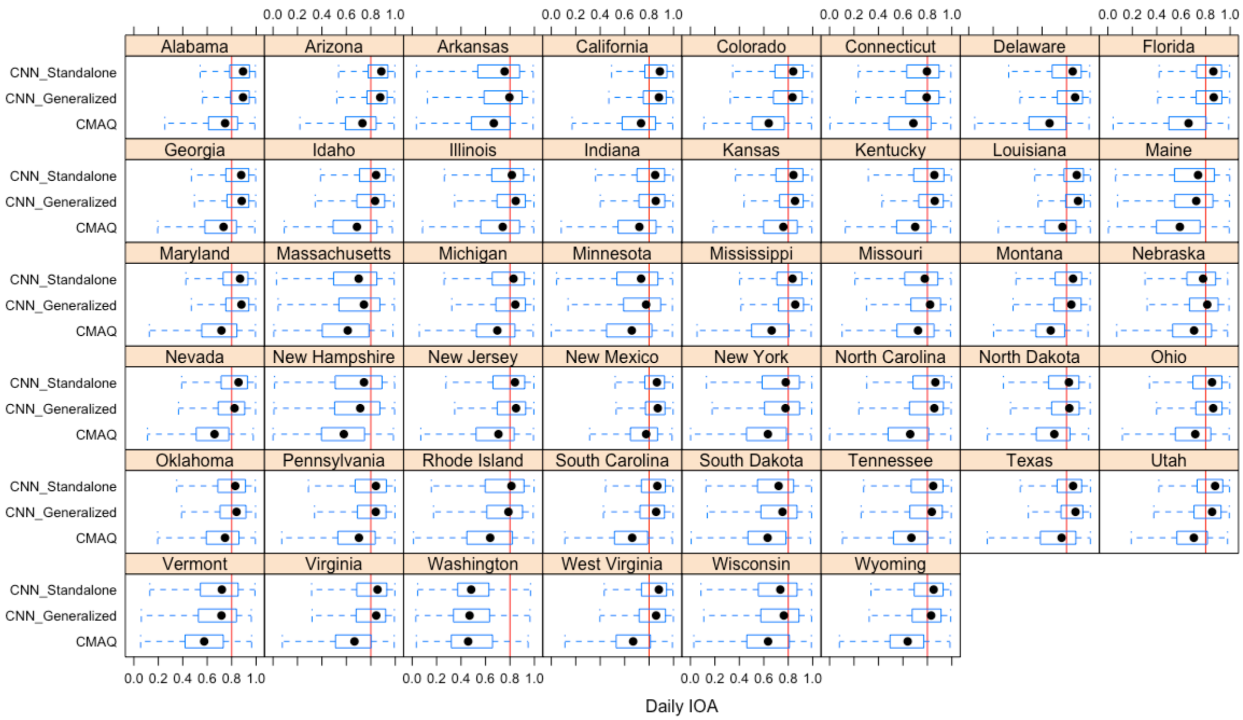


Figure 5.2. Box plot comparing the performance of the two CNN models (generalized and standalone) in enhancing CMAQ daily IAOs in 46 states for the 2014 ozone season.

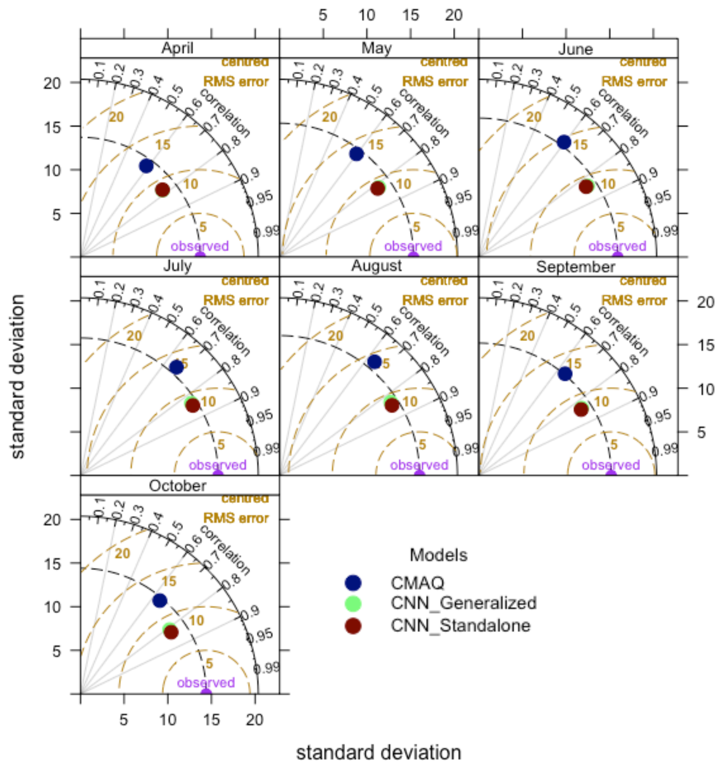


Figure 5.3. Taylor diagram showing the performance of all models in different months of 2014, averaged over all monitoring stations in the continental United States.

5.3.2 CMAQ-CNN model performance

Figure 5.4 compares the box plot of the daily IOAs of the CMAQ and CMAQ-CNN models in different states. This figure shows that, generally, the daily IOA improves for all states with a model enhancement of more than 0.20 in the absolute daily IOA. The average annual IOAs over the US for the CMAQ and CMAQ-CNN were 0.74 and 0.86, respectively, representing more than 15% enhancement in the annual relative IOA. The IOA enhancement and the accuracy of the CMAQ-CNN model differed in the various states and regions of the United States. The CMAQ-CNN was more accurate in the southern and southeastern United States but less accurate in the north and northeastern states. Part of the reason is due to the relationship between the accuracy of the CMAQ and CMAQ-CNN models.

Figure 5.5 shows the range of IOAs of the CMAQ-CNN model at different levels of the CMAQ IOA. The results denote a direct relationship between the performance of the two models. The more accurate the CMAQ model used as input is, the greater the change in accuracy of the CMAQ-CNN model. For a CMAQ IOA ranging between 0.45-0.7, this relationship is generally linear. As the influence of the accuracy of CMAQ, however, decays for stations with both high and low IOAs, other factors are involved. One important factor is the uncertainty of the WRF model in the northern and northeastern United States, which is significantly higher for wind fields, solar radiation, and temperatures, among the most important predictors of ozone (82). In particular, estimates of night temperatures and daytime wind fields in these states, unlike those in the southern and southeastern states, are subject to significant model-measurement bias (83). Since surface ozone is more sensitive to the meteorological parameters in the south and southeastern regions, this finding is particularly important. Thus, a more accurate estimation of meteorological parameters leads to a greater understanding of atmospheric conditions for ozone formation or destruction.

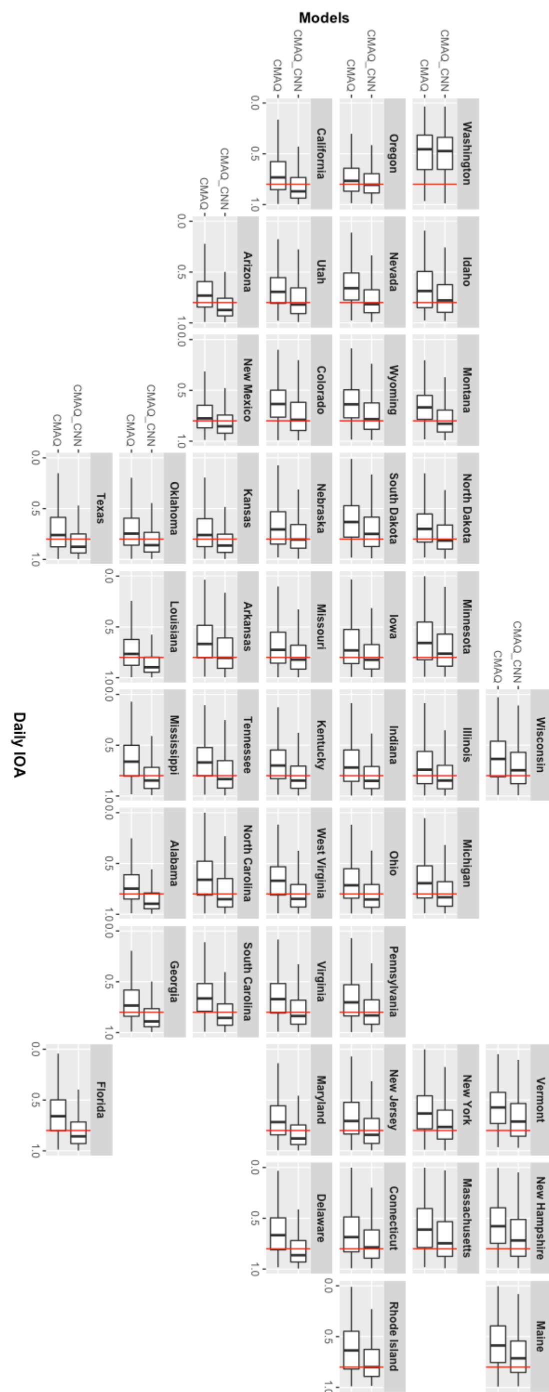


Figure 5.4. Box plot showing the daily index of agreement (IOA) of the CMAQ and CMAQ-CNN models for the 2014 ozone season (April-October). The vertical red lines indicate that IOA=0.8 as a reference.

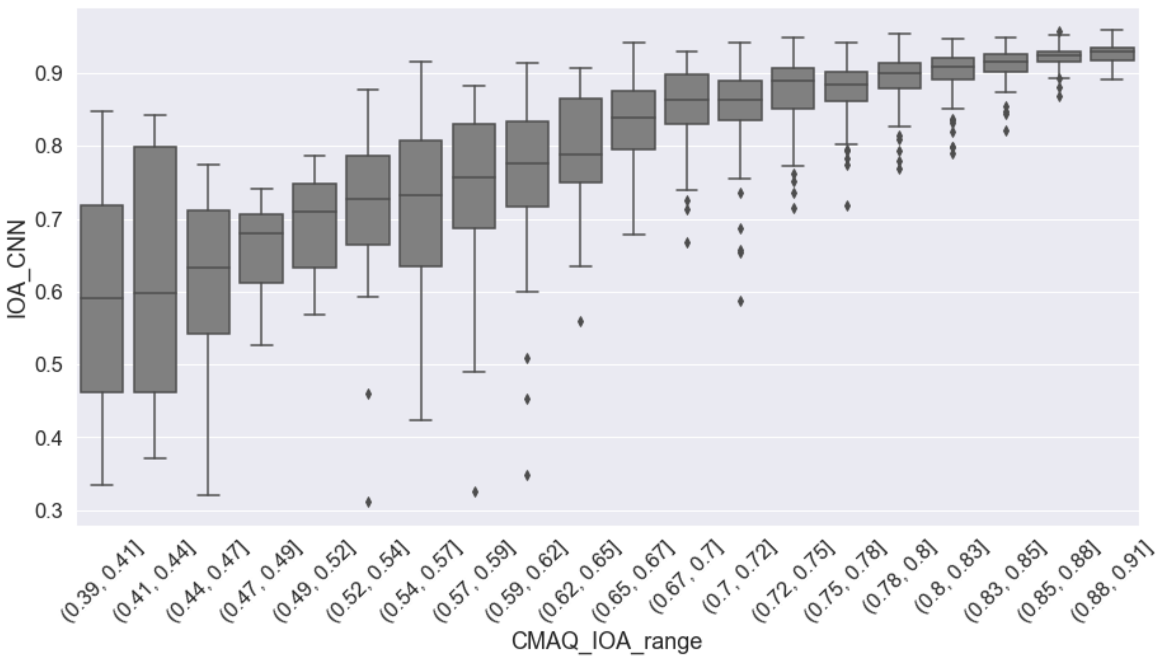


Figure 5.5. Box plot showing the range of CMAQ-CNN performance in the range of CMAQ performance.

In Figure 5.6, the performance of the CMAQ model is compared to that of the CMAQ-CNN model for all states in other statistical measurements using the Taylor diagram. Predictions by the CMAQ-CNN model were significantly more accurate for most states with relatively stable enhancements. With regard to r and RMSE, Washington, Oregon, and Idaho exhibited the worst enhancement levels likely due to the high variability of the WRF and CMAQ models resulting from special geographical conditions (e.g., elevation, leaf area index, diurnal temperature). Therefore, ascertaining the reason behind CMAQ forecasting bias based on a complex environment is challenging. In addition, because of their unique geography of the three northwestern states, the CMAQ-CNN model, which was trained for the entire United States, was severely “undertrained” to represent these states.

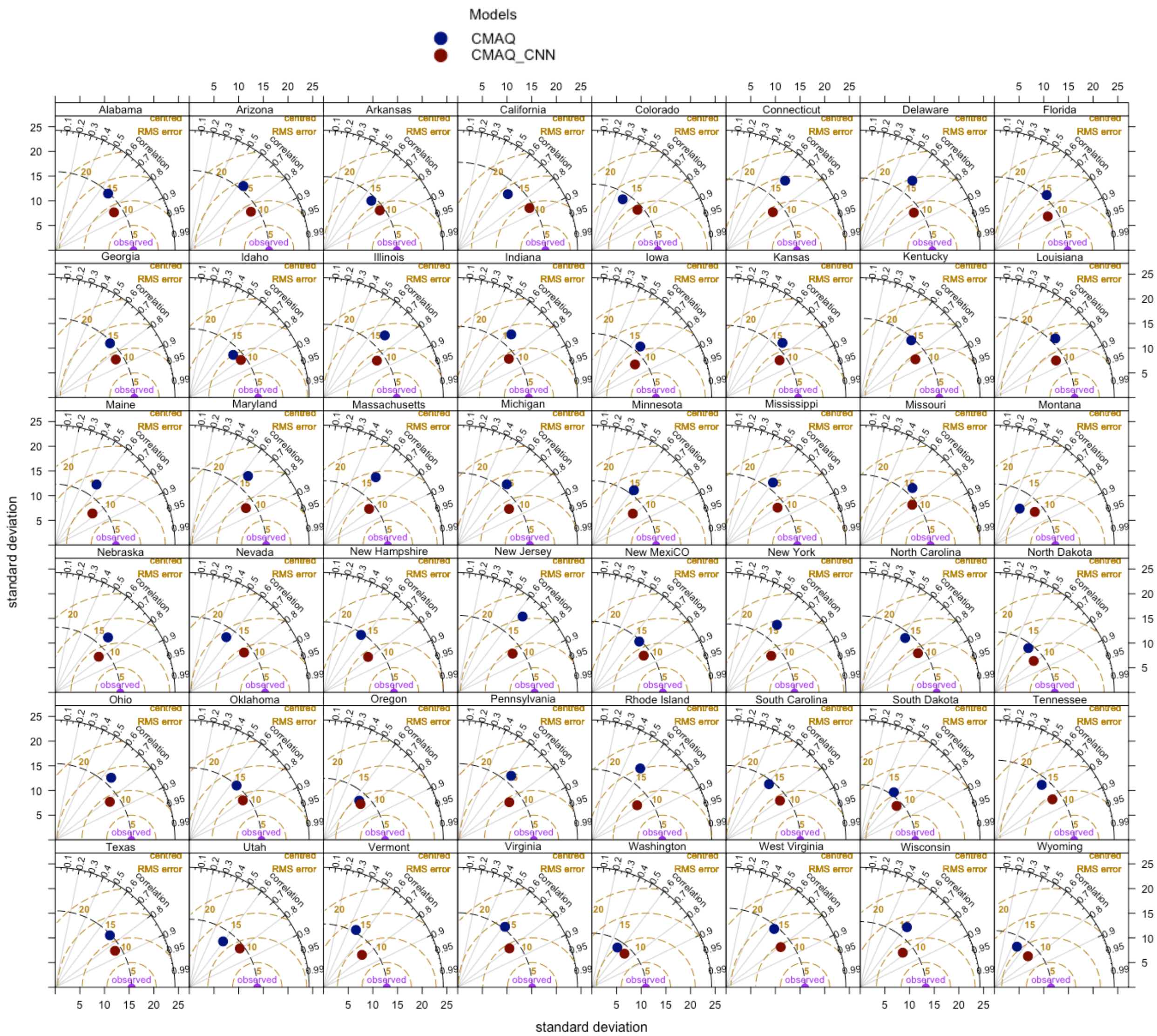


Figure 5.6. Taylor diagram showing the performance of the CMAQ and CMAQ-CNN models in different months of 2014 averaged over all AQS stations in the continental US.

5.3.3 CMAQ-CNN dynamical bias-correction

Figure 5.7 compares the estimates of high ozone episodes of the CMAQ model to those of the CMAQ-CNN model during the ozone season of 2014 across the United States. Results of the CMAQ model exhibit a notable overprediction of high ozone peaks in nearly all states during all months. This trend is more pronounced in the eastern half of the United States than in the western half. The CMAQ-CNN model, by contrast, accurately addressed such overprediction biases in the western states with a sustainable prediction of high ozone peaks. This finding was independent of the relatively large monthly variations of CMAQ estimations in the south and southeastern states (e.g., Carolinas). Figure 5.8 shows a variation in daily IOA changes after the CMAQ-CNN model was applied to all states in different months of the 2014 ozone season. This figure shows that the CMAQ-CNN model used the CMAQ low biased prediction to adjust its bias-correction scheme for the western states (e.g., California, Utah, and Nevada), boosting its accuracy sustainably. Figure 5.8 depicts a trend similar to that in Figure 5.7; in other words, variable performance of the model from month to month in different states.

Both Figures 5.7 and 5.8 display some variation in model performance from month to month. In general, the CMAQ-CNN model provided more accurate predictions with higher daily IOAs in the majority of the months, particularly high ozone months (June-September). This finding was not unexpected. For one, the uncertainty of CMAQ and WRF output is generally higher during cold months than warm months (84), known as the "cold bias." Since the CMAQ-CNN model used CMAQ outputs as input parameters, their uncertainty was not explicitly included in the CMAQ-CNN modeling system, resulting in no footprint of the variation in the input uncertainty in the CMAQ-CNN model. In addition, cloud cover is seasonal and impacts the quality of estimated ozone precursors (by CMAQ) as well as meteorological parameters (by WRF) (85). While the cloud fraction is an input parameter, its impact varies seasonally and geographically. This study, however, explored neither seasonality nor spatial relationships between stations. The relatively significant uncertainty propagated through the bias-correction process showed some degree of inconsistency in the bias-correction level. For instance, the CMAQ-CNN model showed a more than 10% difference when applied in two neighboring stations in the Houston area with the same CMAQ IOA performance, necessitating further studies that develop seasonally- and spatially-specific models to identify such uncertain input parameters.

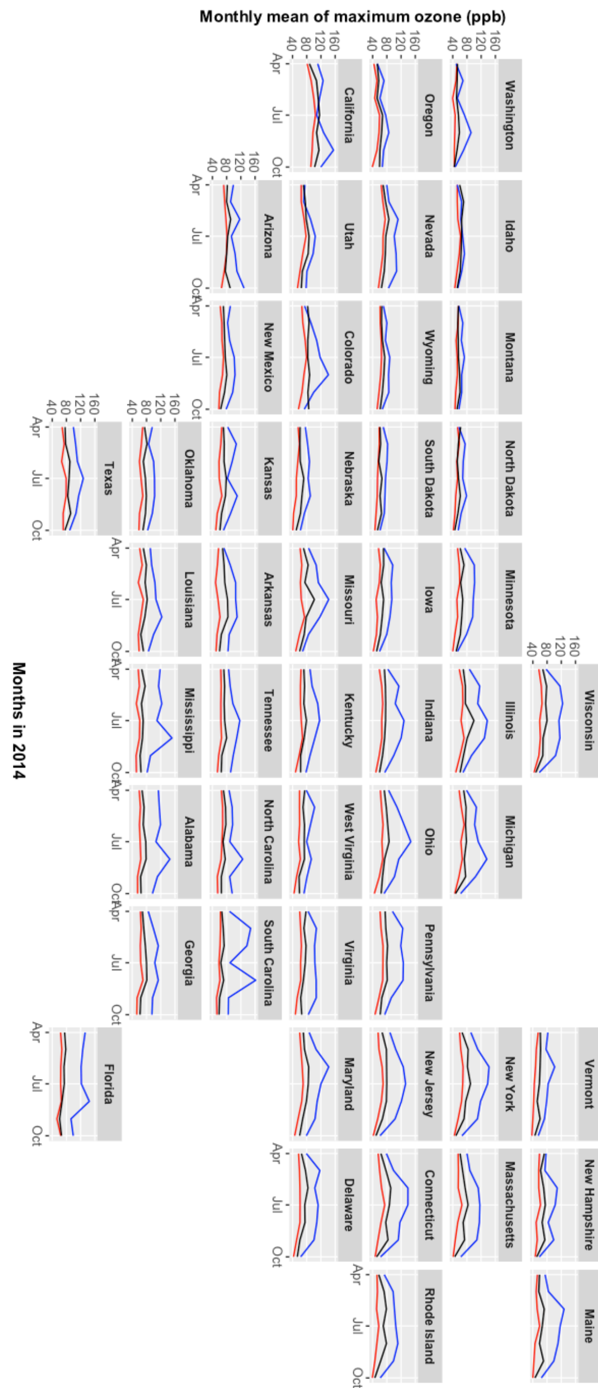


Figure 5.7. Monthly mean of maximum ozone for the 7 months of the ozone season (April-October) in 2014.

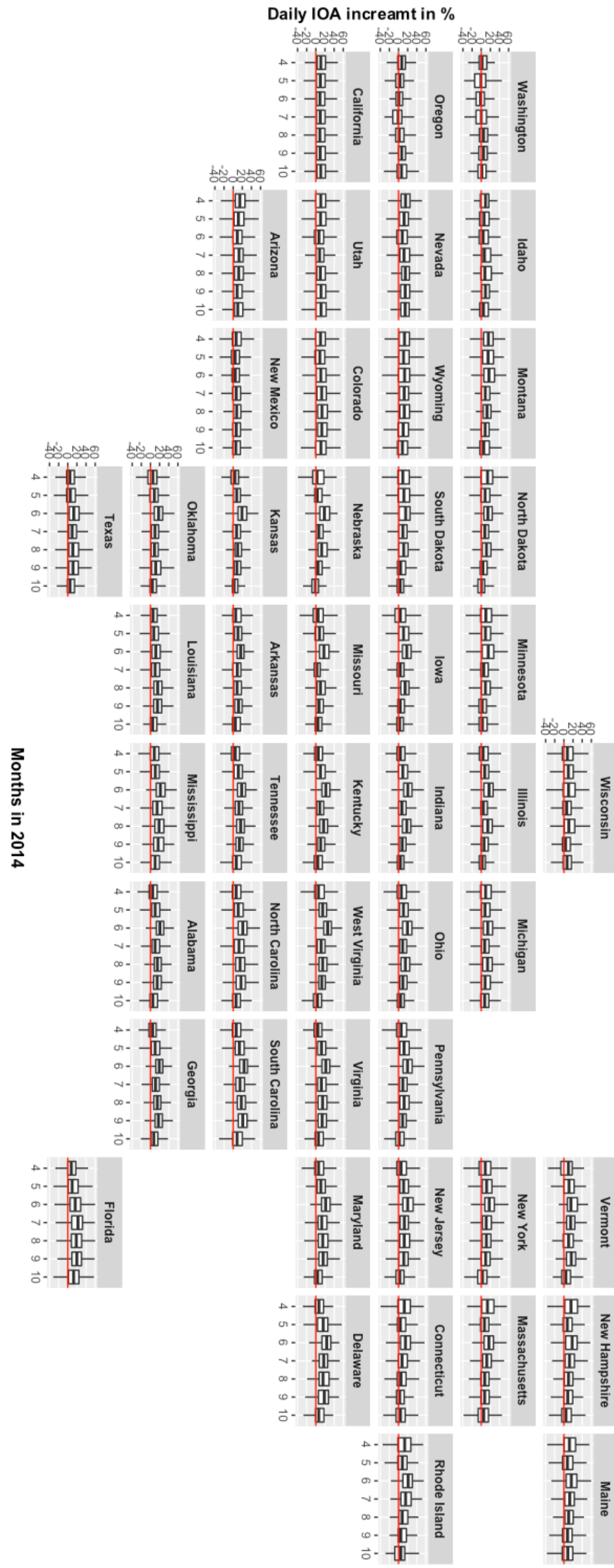


Figure 5.8. Changes in the daily index of agreement (IOA) of the CMAQ-CNN model for the 2014 ozone season (April-October). The vertical red lines represent no changes in the IOA as a reference.

Another reason why we expected the CMAQ-CNN model to provide more accurate predictions with higher daily IOAs during high ozone months was that the sample data we reconstructed from queuing hourly diurnal values (24 hours for each input parameter for each day for each station). This arrangement directed a learning process based on a repetitive, smooth variation of hourly ozone concentration. Depending on meteorological conditions, emissions patterns, and formation/destruction processes, ozone concentrations typically follow a diurnal pattern with minimum values before sunset and peaks in the afternoon. When we compared the warm and cold months, we found that the shape of diurnal ozone concentrations differed significantly. The ozone formation (and destruction) process performed better during the warm months than during the cold months, leading to less variability in the diurnal pattern. Lastly, the hourly variation of ozone concentrations was closely associated with the mesoscale circulation (e.g., sea/land breeze) and the availability of NOx (86). During the warm months, the sea/land breezes are stronger than in cold months because of the greater temperature differences between the land and the ocean or between two landmasses. Stronger breezes lead to stronger winds with less directional variation. Higher wind speed values can boost the accuracy of a machine learning-based forecasting model (as mentioned previously In Chapter 3 and 4), which carries out a more robust transport process of ozone precursors and thus a more homogeneous ozone formation/destruction process.

5.3.4 CMAQ-CNN performance dependencies

The performance of the CMAQ-CNN model widely depends on the performance of the CMAQ base model. Influenced by the significant overprediction of the CMAQ model during the warm months, the CMAQ-CNN model overpredicted ozone peaks in most cases during the warm months (June-September), shown in Figure 5.9, which displays a scatterplot of both the CMAQ and CMAQ-CNN models. During April and October, both models slightly underpredicted ozone peaks, the results of unbalanced data incorporated into the training process of the CNN model, which led to inconsistent training targeting accurate forecasts for a range of daily ozone peaks. Figure 5.10 shows changes in the IOA after we applied the CMAQ-CNN model to different levels of daily ozone peaks. Most positive changes were prevalent on days with daily ozone peaks of more than 70ppb and the absolute daily IOA increased more than 10%. For days with lower ozone peaks, such positive changes were significantly less pronounced. Although predictions of more accurate high ozone episodes were positive in the context of air pollution modeling, "mispredictions" of low ozone days reduce the reliability of the model, requiring further investigation.

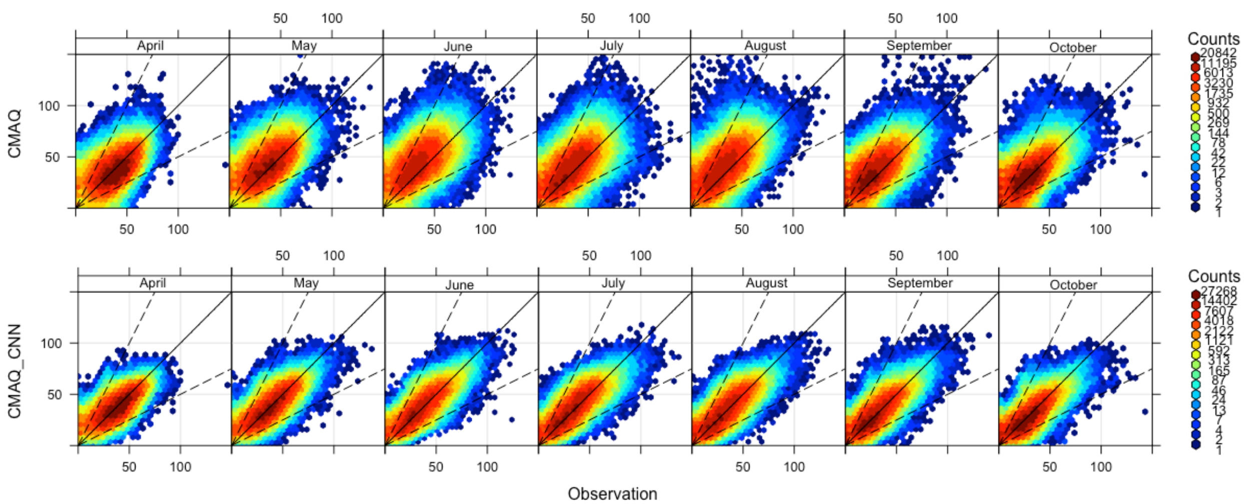


Figure 5.9. Scatter plots of the CMAQ and CMAQ-CNN models in different months in 2014.

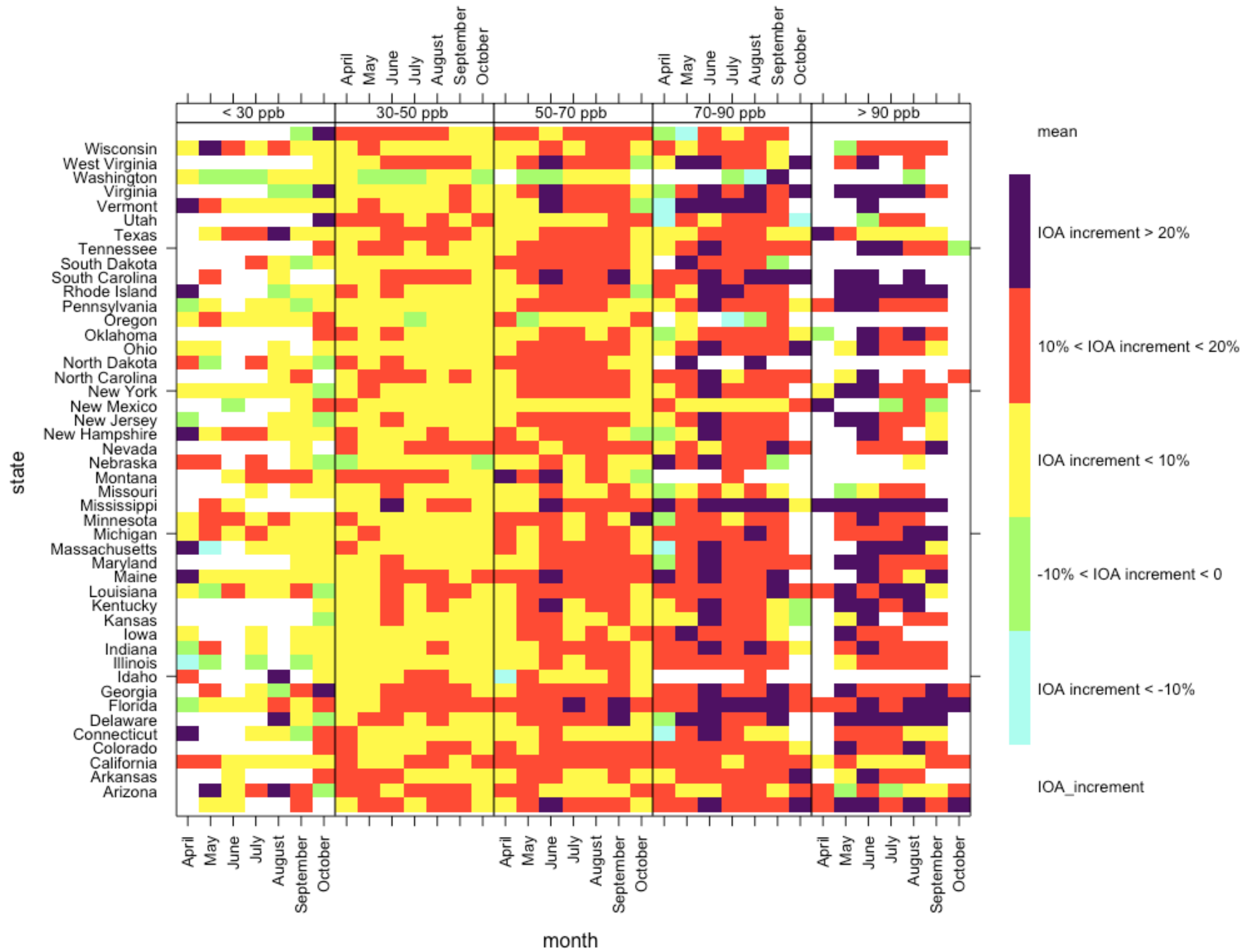


Figure 5.10. Heatmap showing changes in the IOA of the CMAQ-CNN model for different levels of daily observed ozone peaks.

Figure 5.11 the direct relationship between changes in the IOA after application of the CMAQ-CNN model and those of the CMAQ model. Although this relationship follows a mostly linear trend, the nonlinearity of CMAQ IOA was extreme for those days with IOA less than those with 0.7 (similar to that in Figure 5.5). Also, in this figure, $y + x = 1$ (where y represents changes in the IOA and x represents CMAQ IOA) represents the results of an ideal post-processing model. Comparably, $y + x = 0.91$ can be drawn for the CMAQ-CNN model as a representation of the relatively strong likelihood of IAO enhancement.

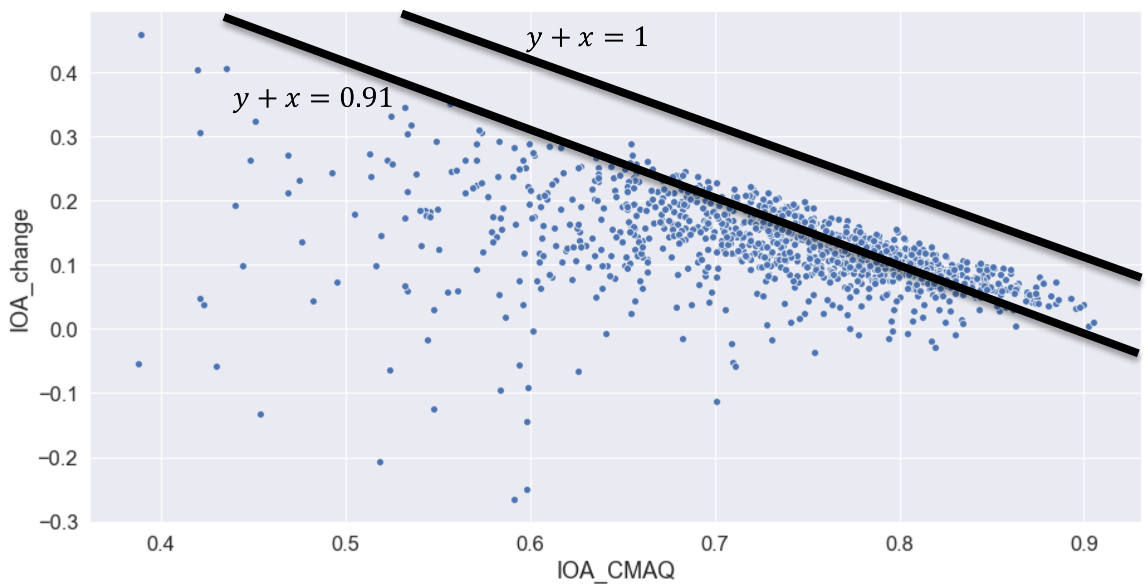


Figure 5.11. Relationship between changes in the IOA of the CMAQ-CNN model and those of the CMAQ model.

Figures 5.12 and 5.13 show the geographical distribution of yearly IOA and changes in the correlation coefficient, the percentage of missing data in observations, and the density of the population (provided by the US Census Bureau) by county across the United States. As expected, the distribution of changes in both the IOA and r was similar to geographically (see Figure 5.12). Generally, when observations had a lower percentage of missing data, their accuracy was greater even though they exhibited no clear correlation (see Figures 5.12a and 5.13a). After all, the spatial generalization of the CMAQ-CNN model was trained by a constant level of sample quality. Figure 5.14 illustrates this finding more clearly. The figure compares the normalized values of the CMAQ and CMAQ-CNN models, represented by the percentage of missing data in the various states. The CMAQ-CNN model shows consistent performance for all states independent of the quality of observed data. Changes in the IOAs and r show closer agreement when compared with the population density (see Figures 5.12a and 5.13b) because the more NO_x emissions there are in highly populated areas, the more accurate the estimates of ozone precursors of the CMAQ model are. Even though the CMAQ model might encounter significant bias in these regions, the higher quality of ozone precursors—available as input in the CMAQ-CNN model—ensures a better understanding of the ozone chemistry, which particularly important when a pool of meteorological and chemical variables is incorporated into the CNN model to address CMAQ biases.

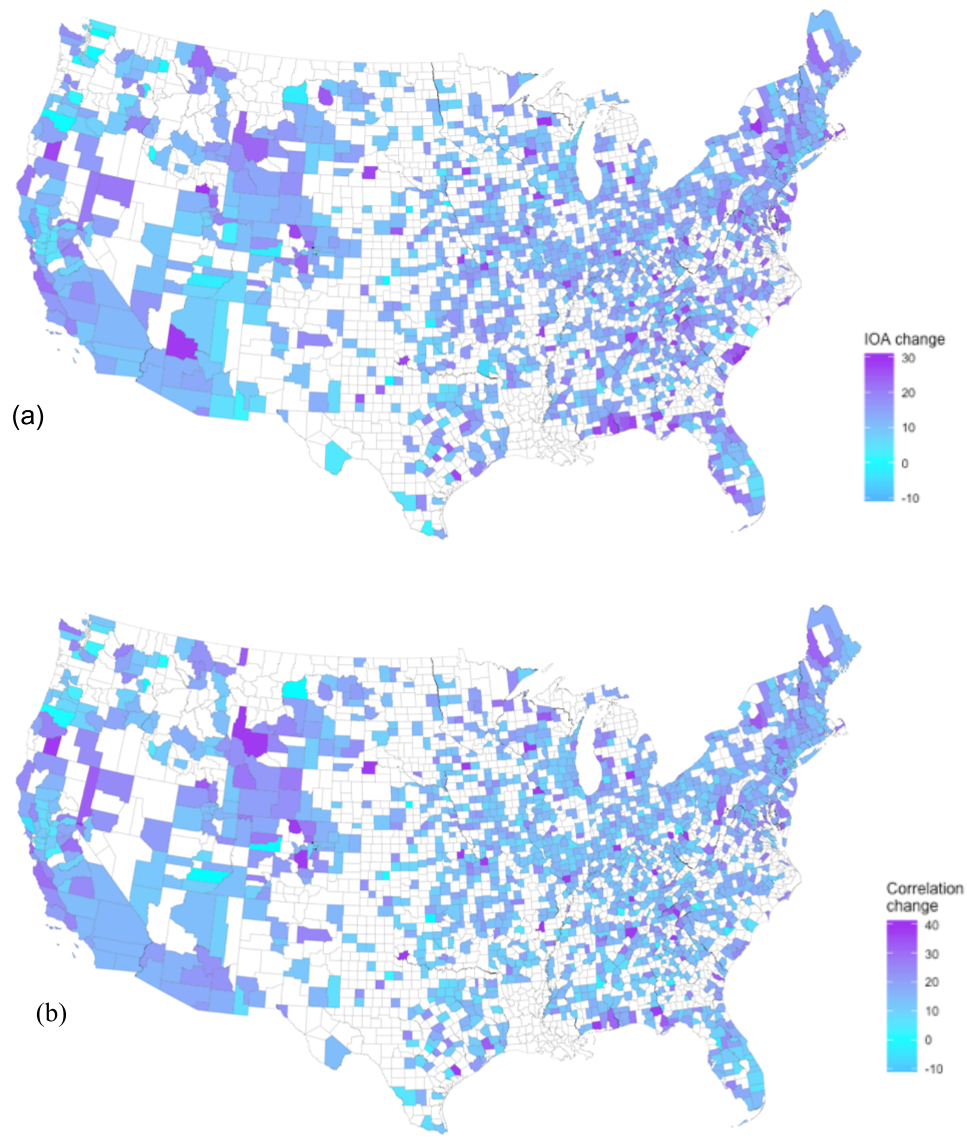
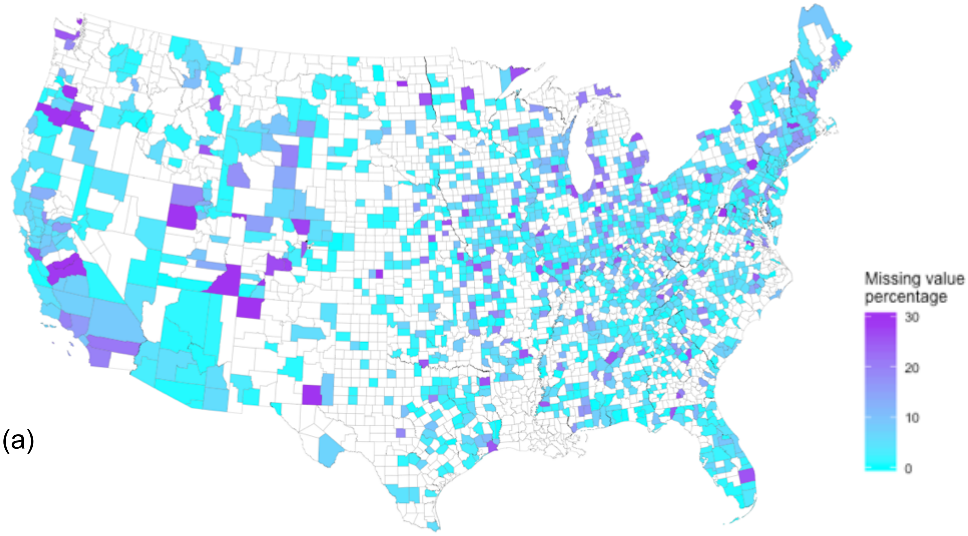
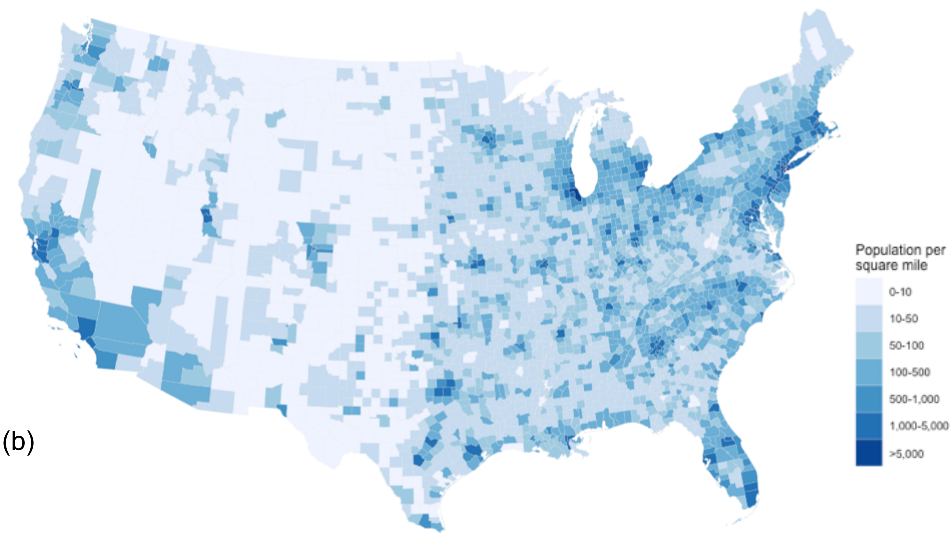


Figure 5.12. County-level changes in (a) the yearly IOA by CMAQ-CNN model; (b) the yearly correlation by the CMAQ-CNN model.



(a)



(b)

Figure 5.13 County-level changes in (a) the percentage of total missing data points in the observational data; and (b) the population density.

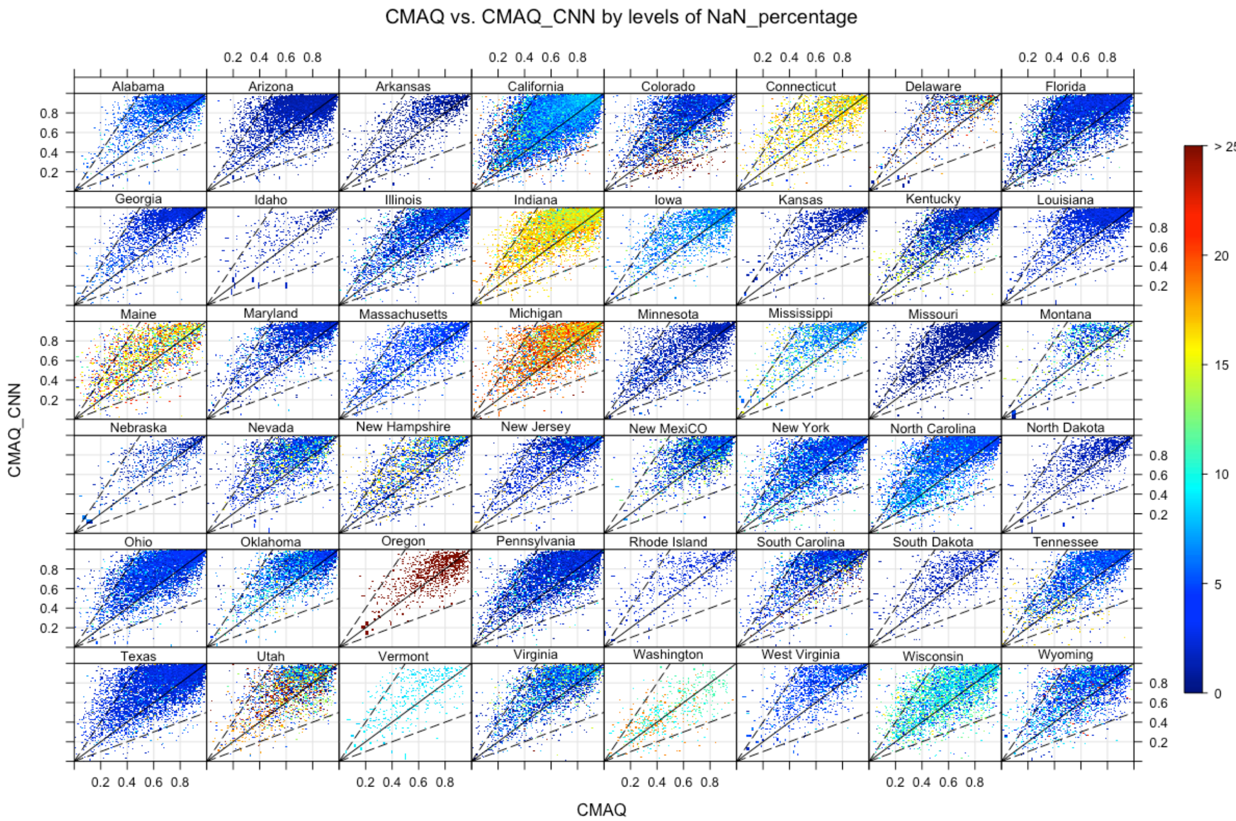


Figure 5.14. Scatterplot comparing the normalized values of the CMAQ and CMAQ-CNN models for different percentages of missing values.

5.4 CMAQ-CNN model diagnoses

It was clear that the success chance of the CMAQ-CNN model to produce accurate results was vastly related to the quality of CMAQ forecasts; when CMAQ forecasted ozone with a yearly IOA more than 0.5, the IOA of CMAQ-CNN model was more than 0.8 for most cases. However, the level of such success was generally unrelated to the level of CMAQ accuracy. For instance, the CMAQ-CNN model was unable to produce a yearly IOA=0.8 even though the CMAQ IOA was more than 0.7 (e.g. EPA #101 Tennessee: CMAQ IOA=0.7; CMAQ-CNN IOA=0.78). Also, in some cases, the yearly IOA after post-processing approach was less than 0.7 (e.g. EPA #1011 California: CMAQ-CNN IOA=0.63). Here, we used the distance analysis from dynamic time warping (DTW) to explain (i) why CMAQ-CNN works satisfactory for some stations than others, and (ii) why it performed poorly in some stations.

To assess similarity between two time-series, DTW works by expanding or contracting a given time-series to minimize the difference between two time-series (87). Its advantage over Euclidean distance, a conventional distance analysis method, highlights when there is a shift (e.g. time lag) between two time-steps in two time-series (see Figure 5.15 comparing DTW with Euclidian distance). The Euclidean distance takes any pairs of data within the time-series and compares them to each other. DTW, on the other hand, calculates the smallest distance between all points, hence, matching one time-step to many counterpart steps on the linked time-series (see Figure 5.15). Due to its non-linear mapping capability, it is widely used in various domains, from time-series classification, to bioinformatic, engineering, health signal processing, and speech recognition (87).

DTW has the benefit that two time-series of the same shape will be classed as similar even if each time-series has different absolute values or if one time-series contains large variability. Figure 5.16 compares the DTW distance between the observation time-series and two prediction models for an ozone monitoring station in Texas. DTW detects the differences between CMAQ estimation and observation with highest difference in the middle of 2014.

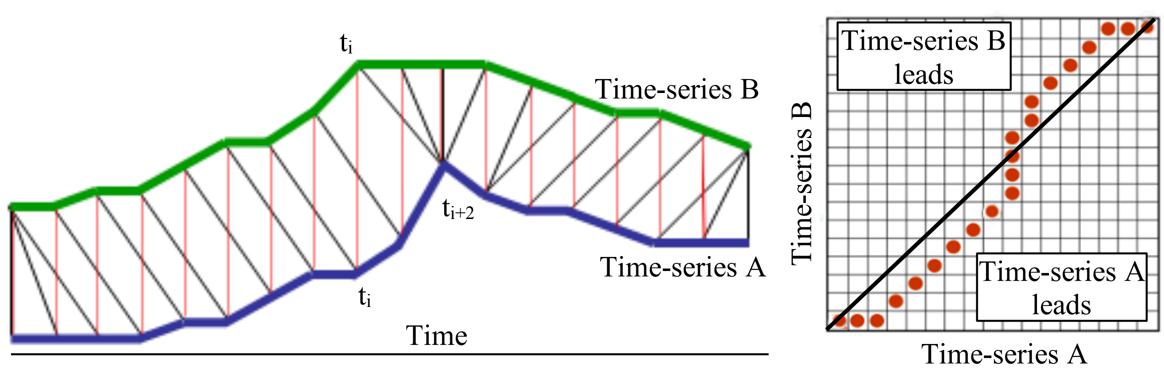


Figure 5.15. Comparison of time steps based on the DTW (black) and Euclidean distance (red) analysis.

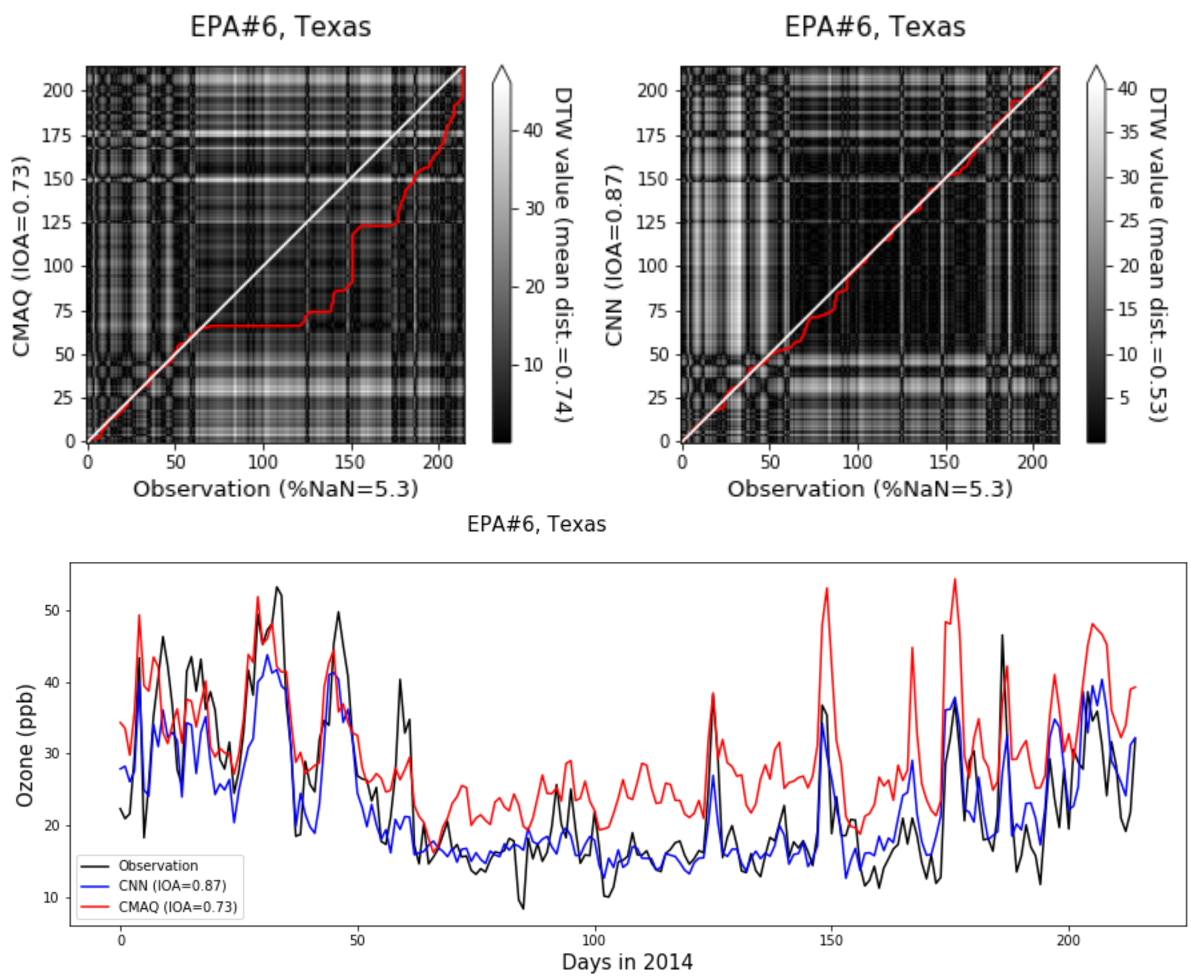


Figure 5.16. Distance plot calculated by DTW comparing similarities (red line) between observations and two prediction models (CNN in the top right and CMAQ in the top left panels).

5.4.1 Satisfactory post-processing scenarios

Figure 5.17 shows the time-series of CMAQ, CMAQ-CNN, and observed daily ozone concentrations in three EPA stations. These stations were selected since the IOA accuracy of CMAQ-CNN model was wither more than 0.9 (Figures 5.17a and 5.17b) or 20% more than CMAQ (Figure 5.17c). Figure 5.18 compares the DTW distance analysis of CMAQ and CMAQ-CNN for the same stations. These are three typical cases of satisfactory improvement by CMAQ-CNN post-processing approach:

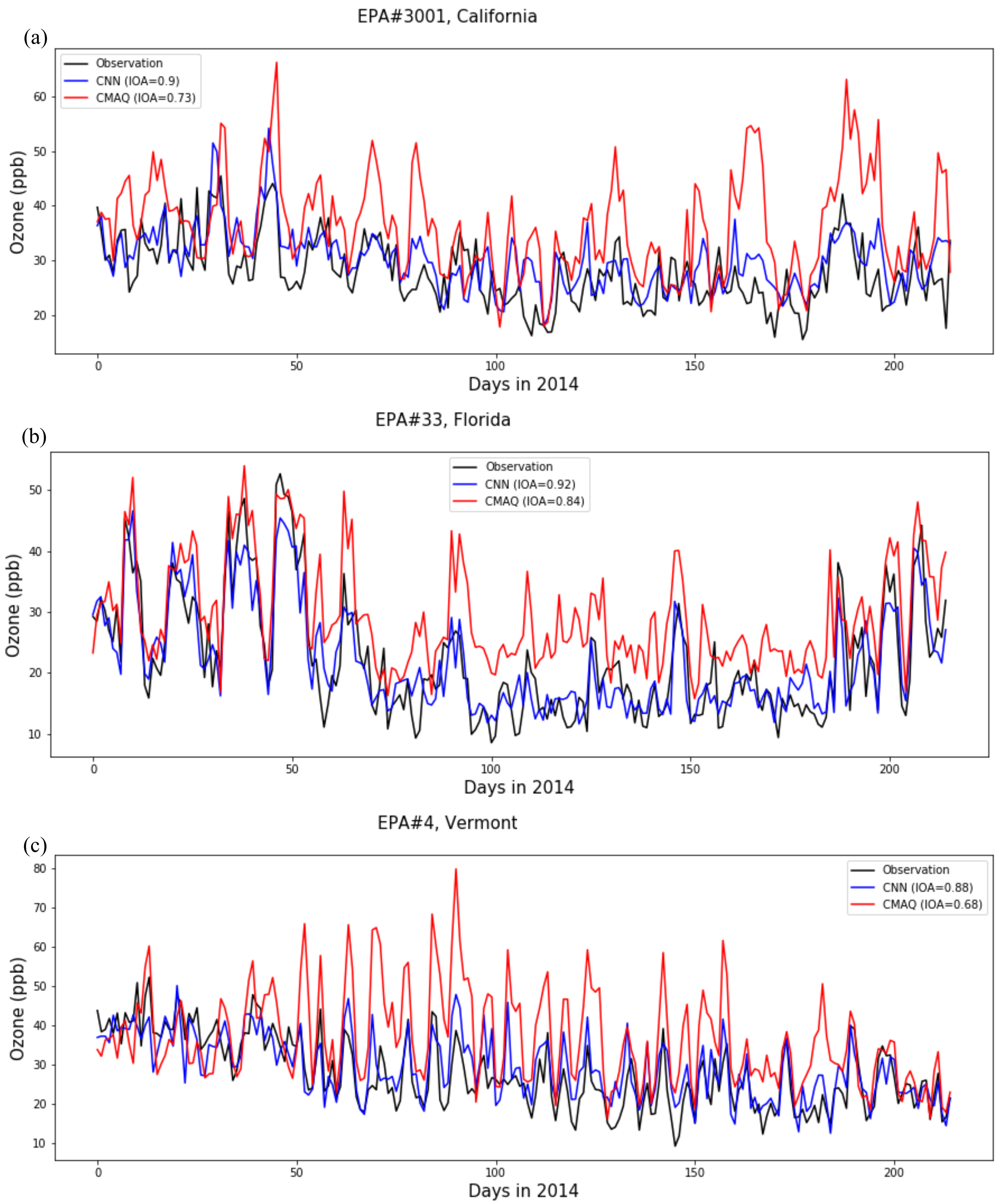


Figure 5.17. Comparison of the time series of CMAQ and CMAQ-CNN predictions for EPA stations (a) #3001 (California), (b) #33 (Florida), and (c) #4 (Vermont).

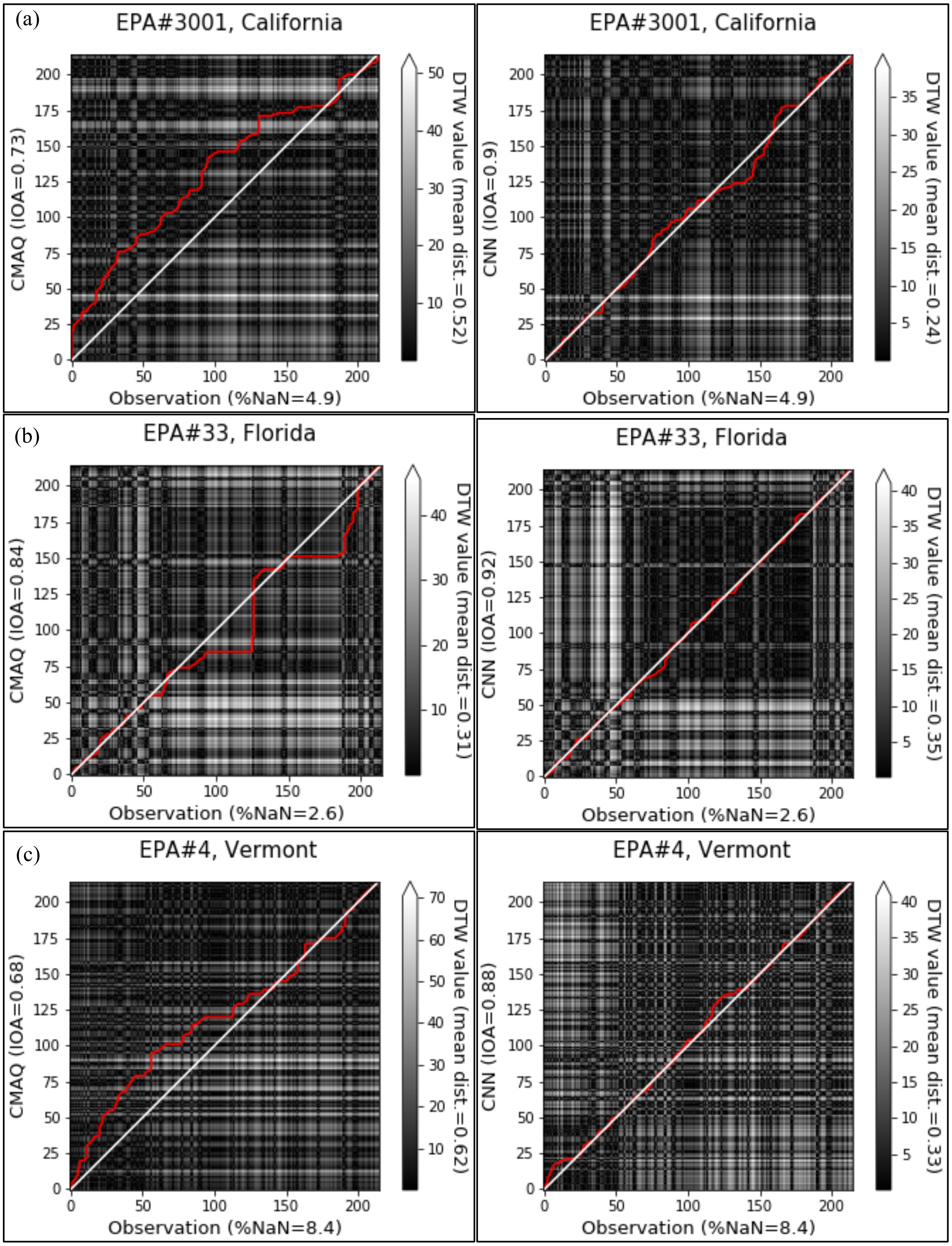


Figure 5.18. Comparison of the distance analysis of CMAQ and CMAQ-CNN predictions for EPA stations (a) #3001 (California), (b) #33 (Florida), and (c) #4 (Vermont).

Figures 5.17-18(a): The ozone observation in this California location was higher during the beginning of the ozone season followed by a relatively steady values ranges between 20-40ppb. CMAQ, however, significantly overestimated daily ozone concentrations after May. The overestimation was more pronounced for the end of the ozone season, resulted in an overall IOA accuracy of 0.73. The DTW distance analysis shows a consistent distance between CMAQ predictions and observed values. This consistency helped the CMAQ-CNN model understands the bias trends in CMAQ and boost the prediction accuracy by 0.17, even though the large distance from CMAQ predictions (mean distance=0.52) mirrored as a relatively significant overestimation in CMAQ-CNN post-processed results.

Figures 5.17-18(b): Here, the trend in ozone concentration followed a U-shaped curve in ozone season due to strong summer winds coming from the large bodies of water near Florida (North Atlantic Ocean and the Gulf of Mexico). For this station, CMAQ accurately predicted this trend thought the ozone season with a relatively constant bias in July-September. As a result, the overall IOA accuracy was 0.84 for CMAQ prediction. The CMAQ, also, showed a consistency in the DTW analysis with two distance gaps in July and September (beginning and the end of the CMAQ overestimation period). The CMAQ-CNN model used the decent performance of the base model in its post-processing algorithm and further improved the CMAQ's IOA accuracy by around 10%.

Figures 5.17-18(c): The trend of observed ozone followed a steadily decreasing trend in this northeast state, due to significantly cooler summer and fall months. That along with the smaller availability of ozone emission sources around of this station resulted in lower level in the ozone formation during the ozone season. However, the CMAQ model overestimated the ozone by more than 50% for the most the season with a relatively large mean DTW distance (0.62). However, the CMAQ-CNN model was able to address this issue due to the consistency of the bias trend in CMAQ predictions (see left panel for DTW distance). Thus, the overall IOA accuracy improved by 0.2.

The key ingredient in a satisfactory post-processing result using the CMAQ-CNN model was the regularity of the bias trend in CMAQ as the base model for training the CNN model. As shown by the DTW distance analysis, when the DTW distance of CMAQ predictions from observed values was consistent throughout the ozone season, the CNN model was able to improve the CMAQ results to a reliable level (IOA>0.8). We have tested this hypothesis by considering typical unsatisfactory scenarios using the CMAQ-CNN post processing approach.

5.4.2 Unsatisfactory post-processing scenarios

Figure 5.19 compares the time-series of ozone observation with two models, CMAQ and CMAQ-CNN, for three selected EPA stations. For all of these stations, the CMAQ-CNN model failed to reach a reliable IOA accuracy level of 0.8, even though it improved the accuracy of the CMAQ model in all cases. Figure 5.20 represents the DTW distance analysis between two models and the ozone observation for the same stations. These are three typical cases of unsatisfactory improvements by the CMAQ-CNN model:

Figures 5.19-20(a): The ozone trend in this station fluctuated through the ozone season with frequent spikes in May, July, and October mostly due to biomass burning (Choi et al., 2016). While the CMAQ model predicted the ozone with a relatively small bias (IOA=0.7), the bias trend varied time to time—meaning that the under- and over-prediction trends were changed frequently. A footprint of this can be seen in the DTW analysis as changing in the path of distance trend. This inconsistency was mirrored in the equivalent DTW analysis for the CMAQ-CNN model as a consistent distance trend, resulted in unsatisfactory IOA accuracy level (IOA=0.78) with an increased mean DTW distance (0.89 compared with 0.74 for CMAQ time-series).

Figures 5.19-20(b): The ozone trend in this California location was relatively constant concentration generally ranged between 10-30ppb. The CMAQ model significantly overpredicted the ozone concentration for the entire time period. This was mostly due to proximity of this station to the Pacific Ocean (San Diego county) that controls the variation of daily ozone concentration (Pan et al, 2017). The DTW distance analysis shows a significant, yet steady, spike in distance between CMAQ and the observation. Thus, even though the CMAQ-CNN significantly improved the accuracy of CMAQ model (IOA=0.63 compared to CMAQ IOA=0.44), such a severe distance was the reason for underperforming the post-processing approach. That also was mirrored a consistent distance in the CMAQ-CNN's distance trend (see the right panel).

Figures 5.19-20(c): In this station, the ozone concentration followed an infrequent trend with lows and highs spread indiscriminately across the ozone season. That is because several factors were affecting the air pollution in this region, including biomass burning, strong frontal system, etc. As a result, the CMAQ model underperformed with large overestimation for the most of time-period (IOA=0.55). The bias of CMAQ model also did not follow a clear trend as shown in the DTW distance analysis. The CMAQ-CNN model improved the prediction results by more than 10% with a reduced DTW distance (0.27 vs 0.35 for CMAQ time-series). However, the varying ozone trend in accompany with inconsistency in prediction bias trend resulted in low overall IOA accuracy of the CMAQ-CNN for this station (IOA=0.67).

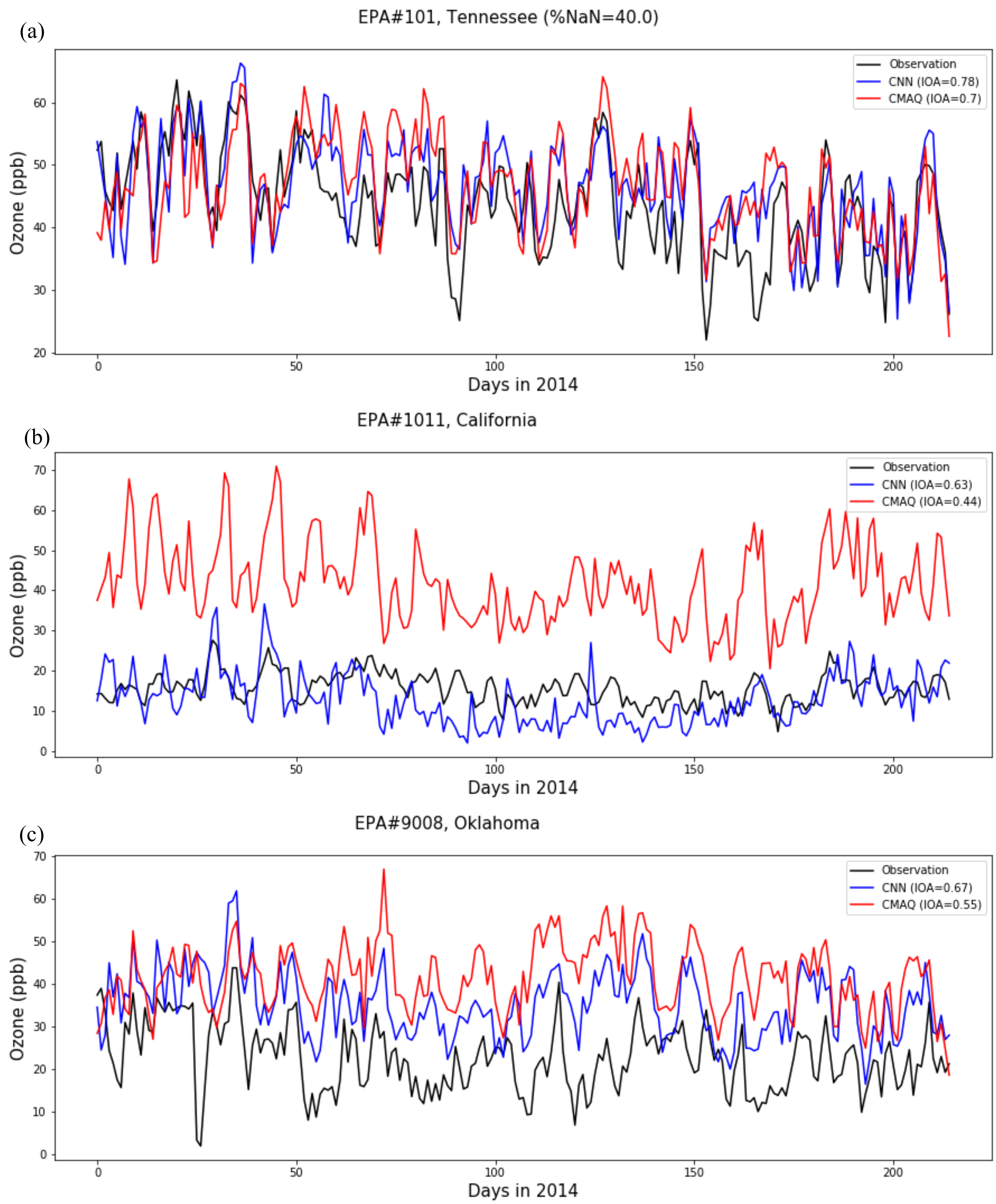


Figure 5.19. Comparison of the distance analysis of CMAQ and CMAQ-CNN predictions for EPA stations (a) #101 (Tennessee), (b) #1011 (California), and (c) #9008 (Oklahoma).

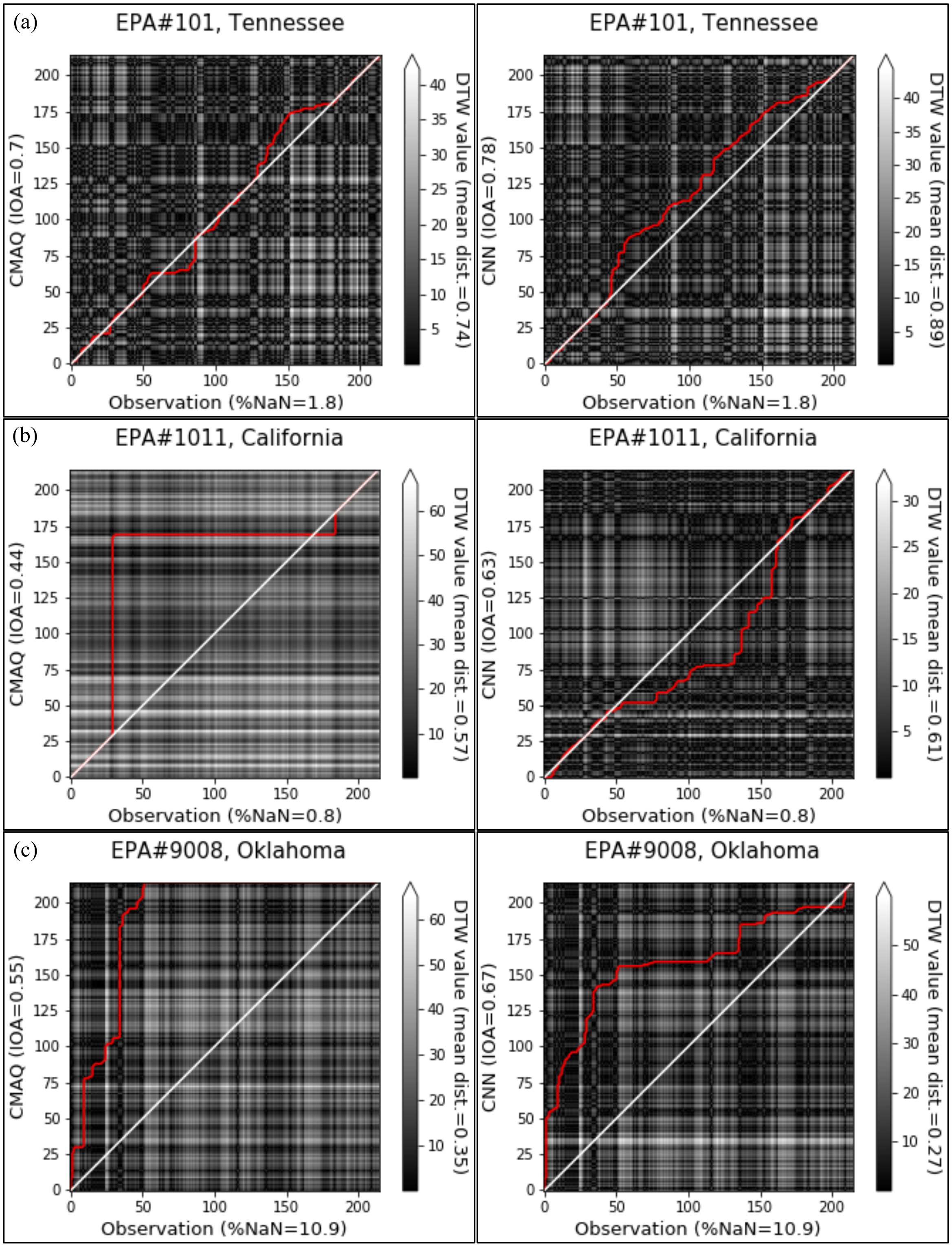


Figure 5.20. Comparison of the distance analysis of CMAQ and CMAQ-CNN predictions for EPA stations (a) #101 (Tennessee), (b) #1011 (California), and (c) #9008 (Oklahoma).

As oppose to the satisfactory cases, the main reason behind the unsatisfactory post-processing results using the CMAQ-CNN model was the inconsistency in bias trend shown by the DTW distance analysis. Other influential factor was the variability of observed ozone concentration. The frequent variation in observation data make it more complicated for the CMAQ-CNN model to be trained to address the bias in the CMAQ model. The geographical location of a station was also an important factor in the improvement level of the post-processing approach. The proximity to the large body of water, or the sources from biomass burning in the ozone season is among the influential geographical features. Also, as can be seen in Figures 5.19-5.20, the DTW distances of CMAQ-CNN predictions from the observed ones followed a consistent trend, hence, according to the statements of Figures 5.17-5.18, a secondary post-processing model might be a possible solution to boost the prediction accuracy.

5.5 Summary

To tackle air quality forecasting problems, a number of studies have recently proposed deep learning models, particularly CNNs. In this chapter, I extend of the capability of CNNs to handle the biases of a particular type of physics-based numerical model, chemical transport models (CTMs). Current methods for overcoming such biases have several limitations: They are not portable; they rely on a long history of quality observations; they cannot be spatially generalized, nor do they incorporate any knowledge from the model in the bias-correction process; and they underestimate high concentration episodes. By introducing a new generation of post-processing techniques, I have addressed these limitations in our proposed model, namely, CMAQ-CNN.

I found that the CMAQ-CNN is capable of being a portable, generalized model and that a single trained model can be applied to an entire modeling domain (here, hourly ozone forecasting across the continental United States). Results, from different points of view, showed that the CMAQ-CNN model enhanced the performance of CMAQ in most cases. It improved accuracy (i.e., the daily IOA) by 15% and mitigated the significant overprediction problem of CMAQ, especially during the warm months (June-September). The model also dynamically handled CMAQ biases. The level of bias correction of CMAQ-CNN was consistent with and independent of the level and consistency of the CMAQ overprediction. After post-processing, we found a direct relationship between the quality of WRF and CMAQ forecasting and changes in the IOAs. Nevertheless, we found no clear relationship between the quality of observations (the percentage of missing data) and changes in the IOA, resulting from the generalization of the model. With regard to the geographical distribution, the forecasts of the CMAQ-CNN model were more accurate for regions with denser populations because of its richer description of ozone precursors in the presence of more NO_x emissions.

The results showed that the improvement level was dependent to the DTW distance of the CMAQ model to the observations. When the calculated distance followed a consistent trend, the post-processing model was able to address the CMAQ's bias independent from its accuracy level or error range. However, when such consistency was absent, or observed ozone varied frequently, the errors in the CMAQ model were mirrored in the results of the prost-processing model.

This study primarily focused on developing a post-processing tool for a CTM model. I combined the advantages of CTM to the understanding of complex physics and chemistry in the atmosphere with the portability of CNN in the construction of a generalized post-processing model. I found that consistent error patterns are critical to the development of an effective and reliable post-processing model bias reduction tool for future air quality forecasts. Thus, if modeling processes are consistent, the same approach can be used for other physics- or chemistry-based numerical models. It is expected that the approach proposed in this study will assist the end-users of numerical models with identifying forecasting errors.

CHAPTER 6. A HYBRID HURRICANE FORECASTING SYSTEM: DEEP LEARNING ENSEMBLE APPROACH AND KALMAN FILTER

6.1 Introduction

As hurricanes are complex phenomena, so is crucial the accurate forecasting of their trajectories and intensities. Certain regions depend on accurate forecasts to execute their disaster preparedness plans and reduce the impact of hurricanes on both people and property. The Gulf Coast, for instance, is prone to tropical storms and hurricanes, as evidenced by Katrina in Louisiana in 2005, Harvey in Texas, and Irma in Florida, both in 2017. As such hurricanes result in significant loss of life and property, it is imperative that communities living in the path of a hurricane receive forewarning to evacuate and mitigate potential damage to property. Timely and accurate forewarning relies on accurate forecasting of the path of a hurricane. As the evolution of a hurricane, however, depends on many factors at different scales, altitudes and time, modeling them can pose extreme challenges (88).

Current dynamical hurricane models use mathematical equations that govern the behavior of the atmosphere at every point on the globe. Common hurricane models, including the Hurricane Weather Research Model (HWRF) (89), are deterministic and solve energy and momentum balance equations to predict the spatiotemporal evolution of a hurricane. Although the accuracy of the model predictions of hurricanes has improved in the last several decades, models still produce a significant number of track and intensity errors (90). The challenge of forecasting hurricanes stems from many complex factors and interactions such as those among ocean temperatures, wind shear, pressure systems, and topography (91). The lack of accuracy was evident when it failed to predict the path of Hurricane Harvey when it stalled near the coast of Texas and made two landfalls in August-September 2017 (92). Figure 6.1 illustrates the prediction of the track of Hurricane Harvey by several models less than two days before it made landfall, indicating significant modeling bias. Hurricane models have a history of unreliable predictions of the maximum sustained wind speed of a cyclone, or intensity (93). One explanation for their unreliability is their inability to observe intensity as a cyclone expands over a wide geographical area with an irregular intensity profile (94). Alternatively, statistical models, which run mathematical equations on a grid with a shortlist of physical predictors, generate more reliable predictions in a relatively short time frame (95). These models, however, are unable to adequately interpret the spatiotemporal relationships in a grid system, rendering them unreliable for accurate track forecasting (96).

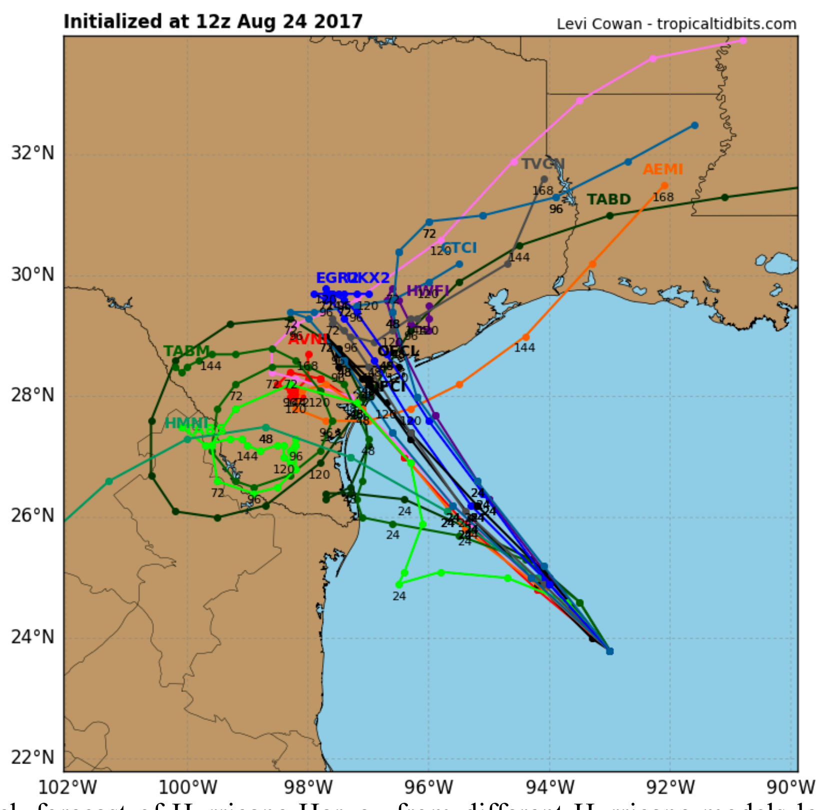


Figure 6.1. Track forecast of Hurricane Harvey from different Hurricane models less than two days before landfall. Figure produced by Levi Cowan and is available at tropicaltidbits.com.

Acknowledging these limitations, I have investigated the potential use of deep learning (DL) algorithm that ensembles the results of different hurricane modeling techniques. Despite successful applications to geospatial and medical image analysis, only a few studies that applied DL for hurricane tracks have been published. In one study, Racah et al. (2017) (97) presented a spatiotemporal convolutional encoder-decoder model to detect extreme climate events. Alemany et al. (2018) (98) used a fully connected recurrent neural networks (RNN) model to predict the trajectory of the hurricane, and Kim et al. (2019) (99) proposed a prediction method for route trajectories using an incremental neural network model. These studies, however, involved a number of limitations. For one, the regional maps (for physical input images) of their models were spatially fixed, posing several problems. One is that a tracked storm must remain in a region, forcing the selection of a vast region with a coarse spatial resolution, which may nonetheless account for the influence of regional physical phenomena such as a nearby pressure system on a hurricane.

To overcome the above shortcomings, our proposed hybrid model adopts two advanced statistical approaches, deep learning and Kalman filters, which use the outputs of several regional and global dynamical models for 24-hour advance prediction. My proposed model involves multiple independent steps for predicting the track and the intensity of a cyclone. With each individual sub-model, the model is capable of identifying important factors that significantly influence the modeling bias and estimating a bias-corrected forecast. One limitation of DL models used as hurricane models is that they do not accurately address sub-grid phenomena inherent in the models. This issue, in some cases, significantly reduces the accuracy of the track forecast when two sub-models are combined. Therefore, I used an ensemble Kalman filter (EnKF) (100) to combine tracking sub-models and their results, further improving the track forecasts EnKF uses a recursive mathematical framework that provides a computationally efficient solution that can easily adapt to any adjustment using given observations (101). This study focuses on the capability of the proposed model to improve the outputs of the direct hurricane model, even in cases of uncertain forecasts. Thus, I chose 2017 to validate our model and Hurricane Harvey as a case study.

6.2 Materials and Methods

The conceptual framework of this paper appears in Figure 6.2. To model the atmosphere as a physical system, we generally use numerical models that estimate the state of the atmosphere by taking several factors into account. These factors are estimated by each model with its specific modeling configuration (e.g. ocean-atmospheric interaction scheme). A computer simulation generates results for a modeling domain in continuous time steps while it verifies them against observational data. However, every dynamical model is subjected to several limitations caused by either restriction in numerical modeling or computer simulation. For example, a simplified physical interpretation of a modeling configuration based on limited knowledge of an ocean-atmospheric interaction causes severe uncertainty in the modeling of the formation of a cyclone. Girded time-space modeling schemes in a computer simulation also cause unavoidable biases that may propagate bias at each step (102).

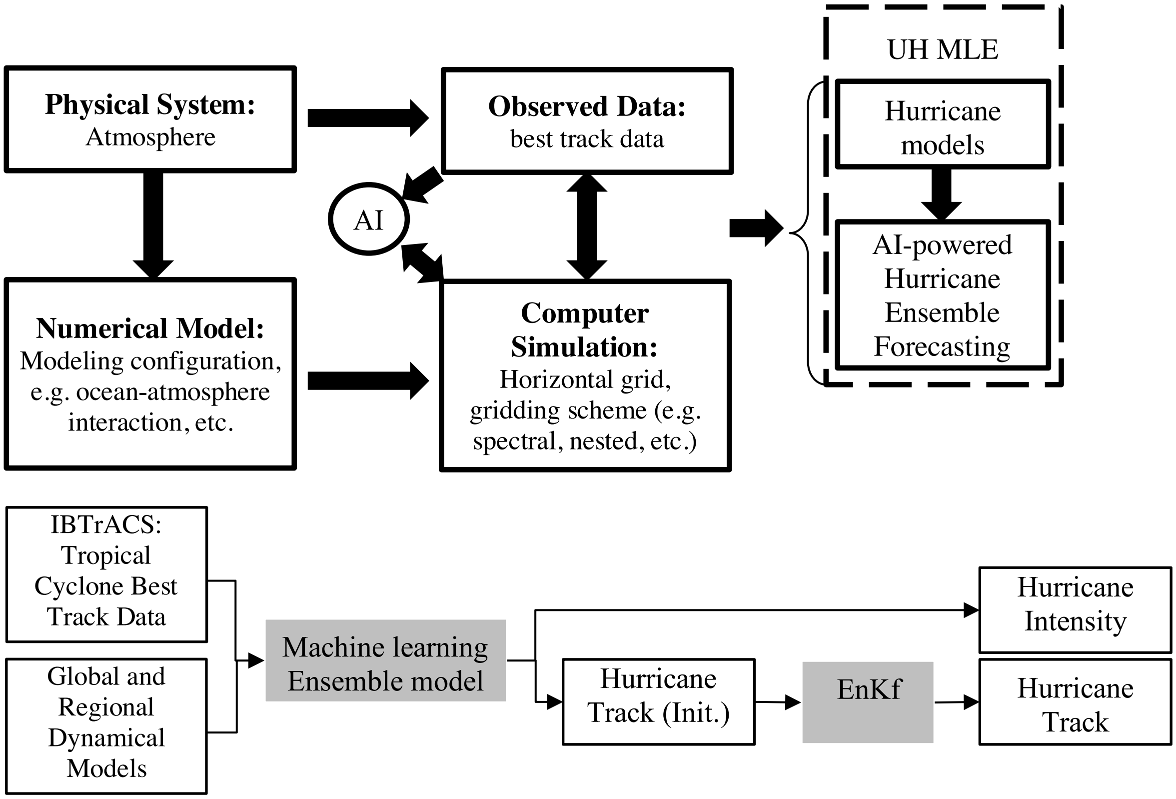


Figure 6.2. Conceptual framework of this study.

One solution to this problem is to use the results of multiple models by implementing an ensemble approach. An approach such as CNN, the machine learning ensemble approach, namely the University of Houston (UH) MLE model, used in this study, combines the strengths of hurricane models and advanced statistical techniques, so, it is more likely to produce more accurate results than either method alone. Unfortunately, ensemble models may not be effective in addressing random errors in the base models (ensemble inputs) (103). In physics-based numerical modeling systems, errors are caused by incomplete physical implementation, so they are not randomly generated (104). Ensemble models may be unable to fully address such errors occurring in all models. However, as they are equipped with the likelihood of an occurrence of an event, however, they are valuable, and their reliability increases when they are provided sufficient data (103).

Figure 6.3 displays the schematics of the proposed ensemble hurricane forecasting model. We propose using the results of eight different global and regional models as inputs of the ensemble model. These models referred to as the Automated Tropical Cyclone Forecasting System (ATCF) by the National Hurricane Center (NHC), which frequently updates their results. Table 6.1 provides a summary of these models. We acquired our hurricane data, comprised of latitude and longitude coordinates, wind speed, and pressure from International Best Track Archive for Climate Stewardship (IBTrACS), and the sea-surface temperature (SST) from the Global Forecast System (GFC). The IBTrACS provides the most complete tracking data of global tropical cyclones that have occurred since 1851, and the NHC houses the largest collection of observational data of hurricanes and tropical storms within the North Atlantic and Eastern Pacific that have occurred since the 1950s. The data were gathered from both hemispheres, and the number of records per storm varies from 2 to around 120 time-steps for different tropical cyclones. Figure 6.4 shows the historical tropical cyclones archived by IBTrACS in the North Atlantic and Eastern Pacific. Operated by the National Centers for Environmental Prediction (NCEP), GFS provides SST data at a resolution of 0.5° for operational proposes. I collected a 14-year history of hurricane tracking model data from 2003 to 2016, which contained a complete set of modeling results.

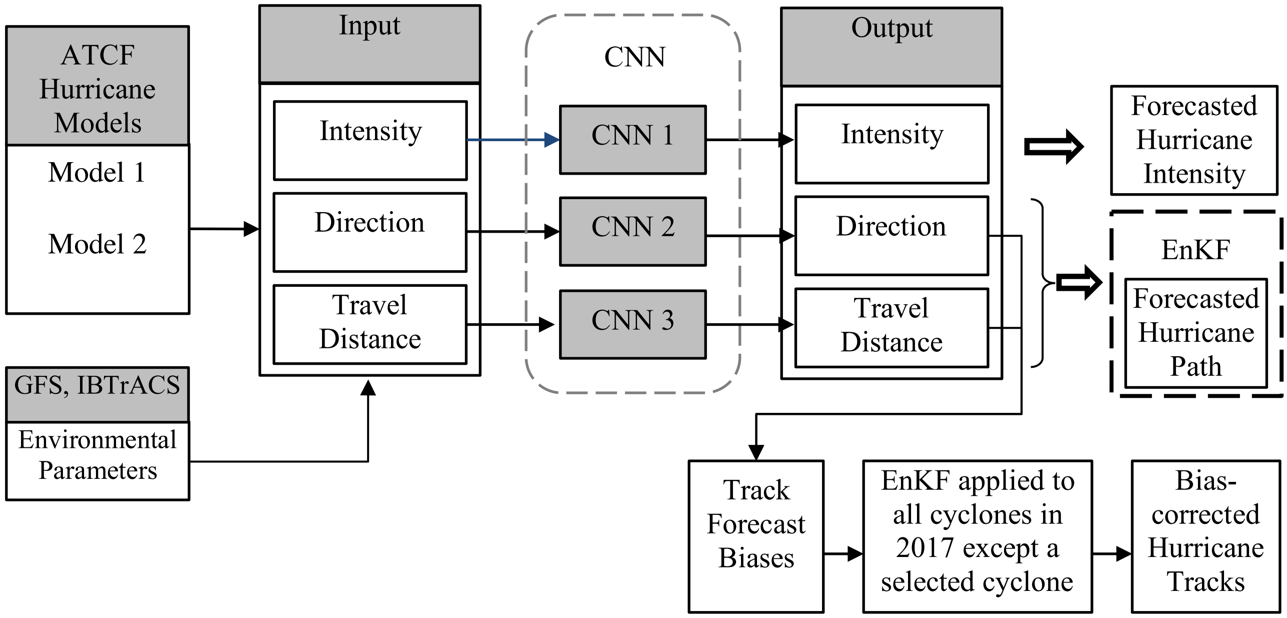


Figure 6.3. Schematic of the proposed hurricane forecasting system.

Table 6.1. Summary of global and regional dynamical models for track, intensity, and wind radii.

ATCF* ID	**Model Name**	**Horizontal Resolution**	**Cycle/Run Period**	**NHC Forecast Parameters**
NVGM/NVGI	Navy Global Environmental Model	Spectral (~31 km)	6 hr (144 hr)	Track and intensity
AVNO/AVNI GFSO/GFSI	Global Forecast System	Spectral (~13 km)	6 hr (180 hr)	Track and intensity
EMX/EMXI/EMX2	European Centre for Medium-Range Weather Forecasts	Spectral (~9 km)	12 hr (240 hr)	Track and intensity
EGRR/EGRI/EGR2	U.K. Met Office Global Model	Gridpoint (~10 km)	12 hr (144 hr)	Track and intensity
CMC/CMCI	Canadian Deterministic Prediction System	Gridpoint (~25 km)	12 hr (240 hr)	Track and intensity
HWRF/HWFI	Hurricane Weather Research and Forecast system	Nested Gridpoint (18-6-2 km)	6 hr (126 hr)	Track and intensity
CTCX/CTCI	NRL COAMPS-TC w/ GFS initial and boundary conditions	Nested Gridpoint (45-15-5 km)	6 hr (126 hr)	Track and intensity
HMON/HMNI	Hurricane Multi-scale Ocean-coupled Non-hydrostatic model	Nested Gridpoint (18-6-2 km)	6 hr (126 hr)	Track and intensity

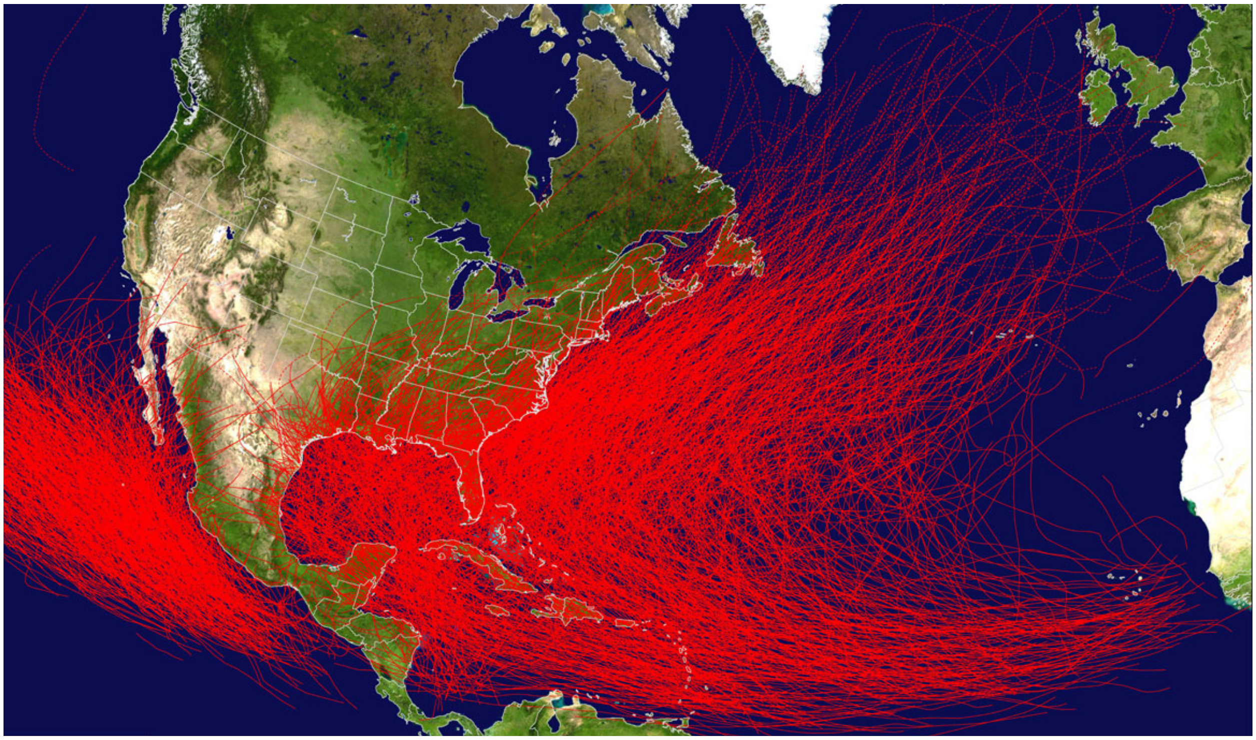


Figure 6.4. Historical North Atlantic and Pacific tropical cyclone archive collected by International Best Track Archive for Climate Stewardship (IBTrACS) since the 1950s.

6.2.1 CNN ensemble modeling

I implemented CNN in the proposed hurricane forecasting model. Inspired by biological processes, In CNN (see Figure 6.4), the convolutional layer applies a convolution operation to the input and passes the results to the next layer. In the fully connected layer, every neuron in the last convolutional layer is linked to every neuron in the output layer. Implementing CNN for accurate hurricane track forecasting, however, requires a large number of track images. As Figure 6.4 shows, the available history of hurricanes might be insufficient to fully train a CNN ensemble model. Thus, I proposed a multi-step modeling approach that individually predicts several hurricane parameters. CNN1 predicts hurricane wind intensity. Using two CNN models, I broke down the track forecast as a two-dimensional problem (image forecasting) into two one-dimensional problems (time-series forecasting) by forecasting the direction of a hurricane (0-

360 degrees) (CNN2) and travel distance (number of miles that a hurricane travels in each step) (CNN3).

6.2.2 Ensemble Kalman Filter post-processing

I explored the advantages of using EnKF to post-process our deep-learning hurricane forecasting system. In particular, by reducing the bias of the deep learning model, I was able to track the path of a hurricane over time. While Kalman filters and their extensions are commonly used for prediction, data fusion, and bias correction (105-107), they have primarily been applied to objects with known or fixed dynamics. The EnKF, however, is a Monte Carlo-based implementation of the Kalman filter for extremely high-dimensional, possibly nonlinear, and non-Gaussian state estimation problems. Its ability to handle state dimensions in high order has made the EnKF a popular algorithm in several geoscientific disciplines (107).

The use of the EnKF as a scalable algorithm was vital for the proposed deep-learning hurricane tracking model for several reasons. (i) The deep-leaning hurricane forecast contains two sub-models that predict the path of a hurricane. Such a sub-modeling prediction approach requires a proper, independent compilation algorithm such as the EnKF to fuse produced results with various scales (degree and miles) of bias and values. (ii) Despite the use of a large amount of data in our deep learning models, unavailable critical factors affect the precise prediction of a hurricane. Thus, dynamically observing the characteristic bias of the forecasting model and correcting it at each prediction step requires a post-processing technique. And (iii) the EnKF will reduce the processing time of the hurricane forecasting system to make sense of the ever-increasing amount of both measured and modeled input data. Here, instead of running the multi-step deep learning procedure after the occurrence of each hurricane, the EnKF systematically modified the trained deep learning models after each measurement update.

In this study, the EnKF estimated the track prediction error and then subtracted by the forecast from CNN2 (direction) and CNN3 (travel distance). The linear updating scheme used in the EnKF is useful for small displacements of forecasted hurricanes from their actual positions; this situation is operationally relevant since the position of a hurricane is often frequently estimated in near real-time by different sources of modeling data. The EnKF used all tropical cyclones in the North Atlantic in 2019 except targeted cyclones, indicating that for each cyclone, the EnKF updated the positions based on the model-measurement errors of all other 2019 cyclones.

6.3 Results and Discussion

6.3.1 Model performance for 2017 Atlantic hurricane season

I compared the performance of our models to that of the observation benchmark, IBTrACS. We also directly compared model errors and modeling skills using root mean squared errors (RMSE) with the CLImatology and PERsistence (CLIPER) model[3] as the NHC official forecast for 2017. Figure 6.5 compares the results of our hurricane forecasting model, namely, the University of Houston Machine Learning Ensemble model or UH MLE, with the NHC official forecast and selected hurricane forecasting models. The results showed relatively better results than the NHC official forecast; intensity forecast errors (Figure 6.5a) decreased by 34% and the track forecast errors (Figure 6.5b) by 13%. Also, Figure 6.5c indicates the relative advantage of UH MLE over other hurricane models for 2017 North Atlantic tropical cyclones.

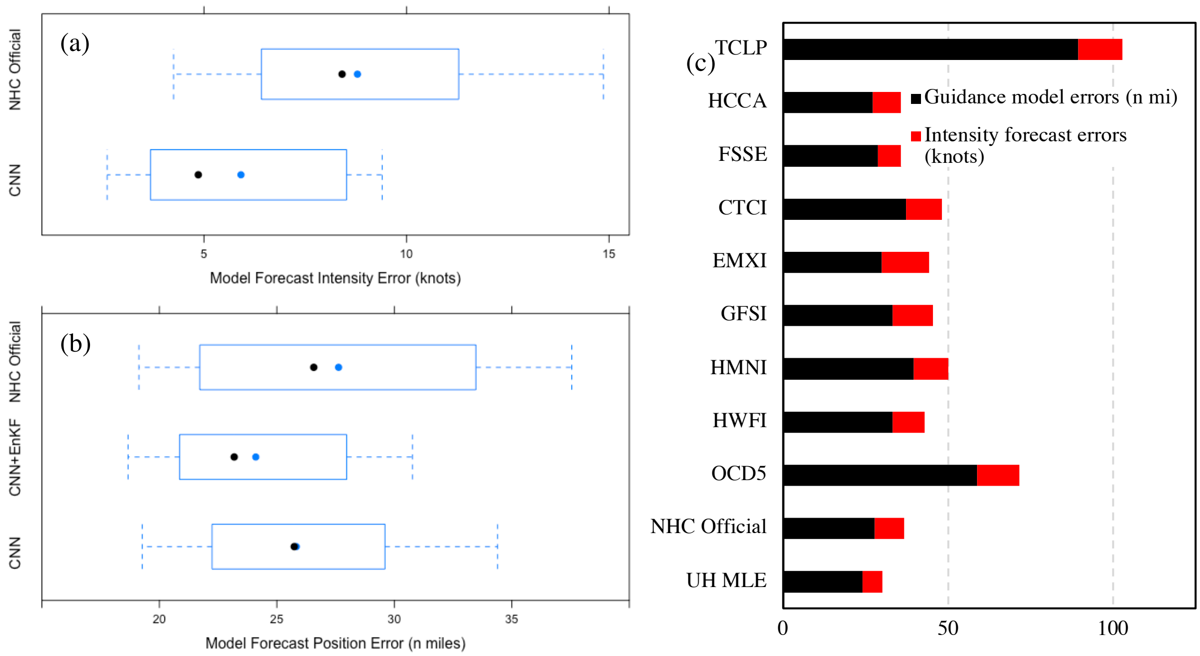


Figure 6.5. Comparison of (a) the intensity forecasting errors of UH MLE (CNN1) and the NHC official forecast, (b) the track forecasting errors of UH MLE (CNN2+CNN3+ENKF), CNN (CNN2+CNN3), and the NHC official forecast, and (c) the path and intensity forecasting errors of UH MLE, the NHC official forecast, and selected model forecasting errors for the North Atlantic tropical cyclones in 2017.

The UH MLE predicts the parameters of a hurricane with a more stable range of errors. This is particularly beneficial when a forecasting system is used for operational purposes. The reason might be that environmental factors (pressure and SST) have been incorporated into the ensemble approach, so the model is able to relate the biases of dynamical models to these footprints of weather conditions. This finding is more apparent in intensity forecasts, which contain more intense relationships between these parameters and the wind field.

The post-processing by EnKF further improved the track forecast by an average of more than 6% and as much as 19%. This proves the capability of the filter at adjusting the noisy predictions (CNN models) by a given set of similar observations (i.e. cyclones that formed during the same hurricane season in a similar geographical location). Tables 6.2 shows the detailed statistics of forecast results from several cyclones. For all cyclones except for Tropical Storm Emily, the EnKF showed a scope of improvement. Emily was a rare rapidly-forming, a short-lived tropical storm that originated in the Gulf of Mexico relatively near the western coast of Florida. The location of tropical cyclones of the 2017 Atlantic hurricane season can be seen in Figure 6.6. Consequently, the dynamical models estimated its track with a highly biased forecast (+37 nocturnal miles of the official track error forecast). The CNN models already reduced the bias by more than 15 miles, more than 40%. The remaining errors were likely randomly generated (i.e., they were the result of unknown sources); hence, use of the post-processing approach left no room for improvement.

Table 6.2. Comparing the forecast biases of track and intensity of the UH MLE model and NHC official forecast for 2017 North Atlantic tropical cyclones.

	Name	UH MLE track (init.) (n miles)	UH MLE track (n miles)	EnKF impact (n miles)	EnKF improvement (%)	NHC Official track (n miles)	UH MLE Intensity (knots)	NHC Official Intensity (knots)
Tropical Storm	Arlene	20.67	20.03	0.64	<3*	28.22	7.91	9.23
Tropical Storm	Bret	20.43	19.49	0.94	4.6	19.13	8.52	4.25
Tropical Storm	Cindy	29.60	27.97	1.63	5.5	36.59	3.68	4.61
Tropical Storm	Don	**32.37**	**30.35**	2.02	6.2	34.11	3.68	4.56
Tropical Storm	Emily	22.24	22.57	**-0.33**	**-1.5**	37.55	9.20	8.43
Hurricane	Franklin	27.79	23.87	**3.92**	**14.1**	26.42	4.74	7.36
Hurricane	Gert	23.82	23.18	0.64	<3*	32.09	5.91	6.42
Major Hurricane	Harvey	20.10	19.47	0.63	<3*	26.57	8.71	11.77
Major Hurricane	Irma	30.91	29.95	0.96	<3*	30.88	3.54	11.29
Major Hurricane	Jose	19.27	18.67	0.60	<3*	21.72	9.17	13.90
Hurricane	Katia	25.74	20.86	**4.88**	**19.0**	21.50	3.37	8.07
Major Hurricane	Lee	**34.39**	**30.77**	3.62	10.5	36.08	3.21	10.64
Major Hurricane	Maria	25.95	25.14	0.81	<3*	22.06	**9.40**	**14.86**
Hurricane	Nate	**30.97**	**30.01**	0.96	<3*	33.47	4.86	6.97
Major Hurricane	Ophelia	24.61	21.81	2.80	11.4	19.47	7.81	8.41
Tropical Storm	Philippe	25.84	24.02	1.82	7.0	22.88	4.21	5.75
Tropical Storm	Rina	24.37	21.57	2.80	11.5	20.90	2.61	12.90
	Average	25.83	24.10	1.73	6.4	27.63	5.91	8.79

* EnKF exhibited a notable instability in bias-correction, Thus, the average values 5 runs are selected as represented results.

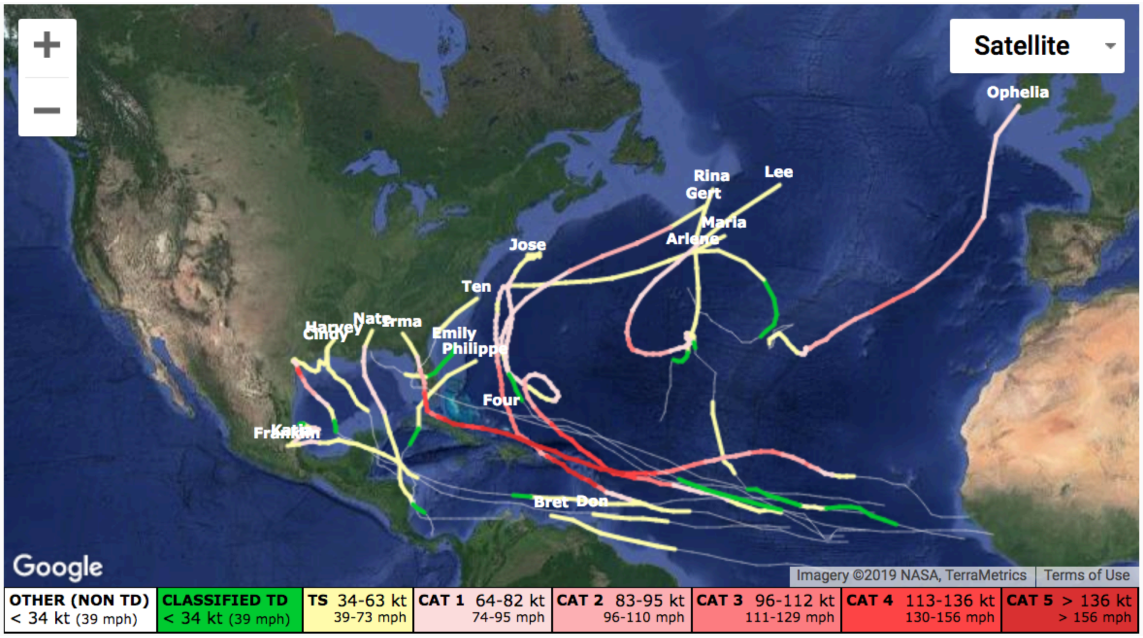


Figure 6.6. 2017 North Atlantic tropical cyclones, obtained from ropicalatlantic.com.

As seen in Table 6.2, the UH MLE failed to address several cases with large track forecasting errors, Tropical Storm Don, Hurricane Lee, and Hurricane Nate, all highlighted in bold font. Although the forecasts for all of these cases were significantly more accurate than the NHC official forecasts, the forecast error was over 30 nocturnal miles. This finding could have several explanations. For one, the hurricane models (the inputs of our ensemble model) estimated the track with larger than average uncertainties caused by either the short lifetime of a cyclone (Don), a specific location of the cyclone formation and path (Nate), or complex interactions with neighboring pressure systems (Lee). These uncertainties were mirrored in the CNN ensemble model. Another explanation is that a limited number of environmental variables were used to regulate the dynamical modeling biases, so the CNN models failed to understand the cause(s) of errors. Finally, although hurricanes are among the most extreme weather events, the training history of CNN models was limited to 14 years of modeling outputs. If more data and more environmental variables were used, the CNN models might have been able to contend with some of their inherent biases by incorporating a more comprehensive understanding of weather patterns.

6.3.2 Model performance for Hurricane Harvey

I studied Hurricane Harvey as a case study of improvement in forecasting by our proposed model. Figure 6.7 compares the results of our modeling approach with observations (IBTrACS), and the NHC official forecast at various time steps before and after the first and second landfall of Harvey, which intensified on August 24 and made its first landfall on August 26 in South Texas. The second landfall was on August 30 in the northern part of the Gulf of Mexico as a weak tropical storm that caused flooding damage across a wide geographical area in the southern regions of the United States (108). The results show significant improvement in both the track forecast (by 27%) and the intensity forecast (by 26%) compared to the NHC official forecasts. Both before and after landfall, the forecasting results reflected these improvements, indicating the robustness of the forecasting product. Improvements were the most noticeable after the first landfall and before the second one when Harvey stalled for several days (see Figure 6.7). Although, in this case, the dynamical hurricane models were significantly uncertain in the forecast 24 hours in advance, our model captured the errors, resulting in more accurate forecasts (Figure 6.7c). Also, when Harvey weakened after its first landfall, simplification of the hurricane track in the "direction" and "travel distance" improved the forecasting results (Figure 6.7b).

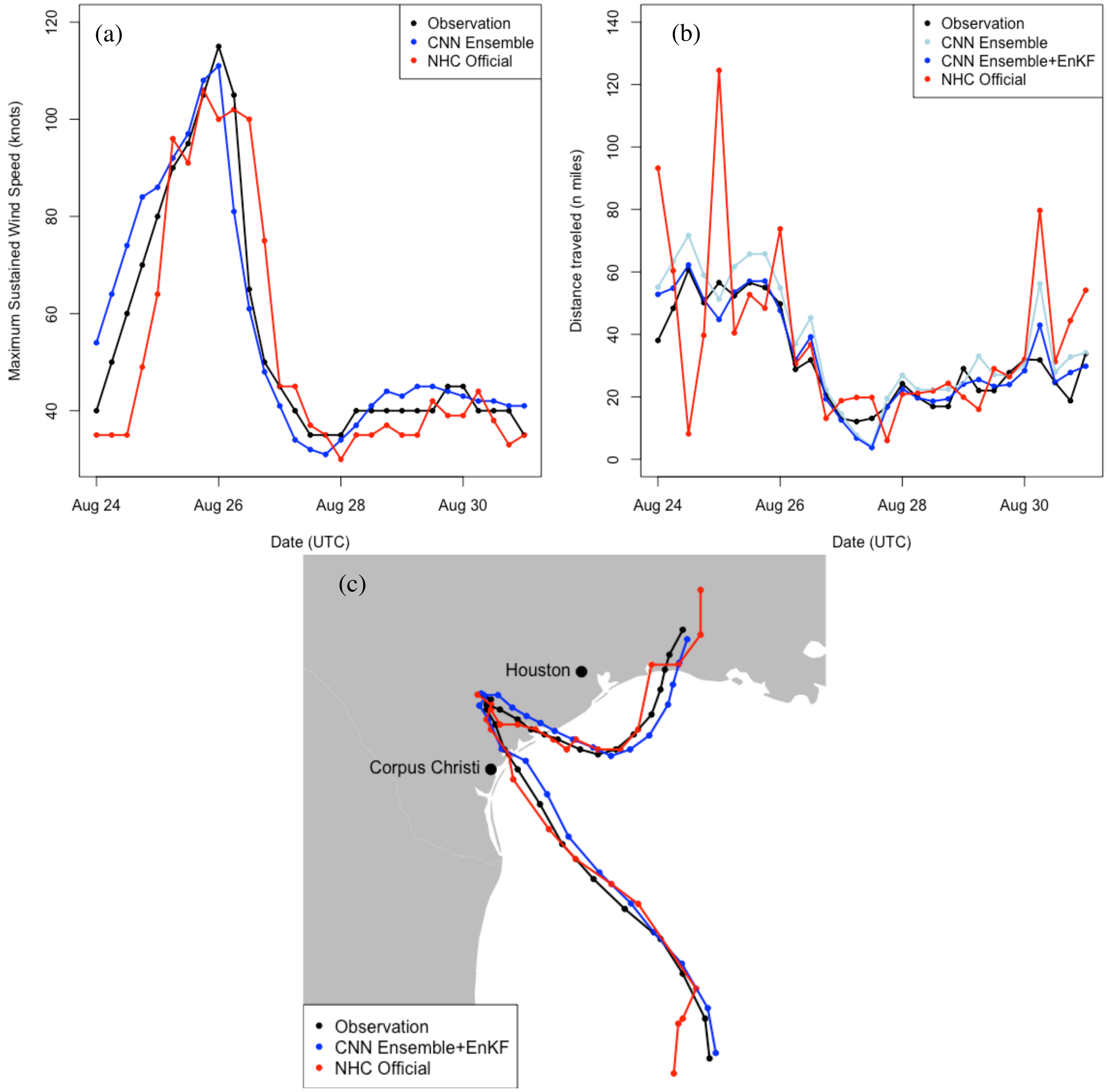


Figure 6.7. Preliminary results of the UH Machine Learning Hurricane Ensemble Forecasting system for the forecasts of (a) the intensity, (b) the distance traveled, and (c) the intensity of Hurricane Harvey in 2017.

6.4 Summary

A hybrid multi-step hurricane forecasting system was developed. First, three CNN ensemble models were developed for predicting a hurricane's wind intensity, distance traveled, and direction using the results of eight dynamical hurricane models. These models were trained using all tropical cyclones in North Atlantic and Pacific before 2017 and tested for those in 2017. Then, an ensemble Kalman filter was applied to post-process the track forecast of the CNN ensemble model. At each step, the hybrid model forecasted all hurricane characteristics 24 hours in advance.

The results of the hybrid model showed the statistical advantage over the NHC official forecast with approximately 13% better track and 34% wind intensity forecasts. The ensemble Kalman filter, as a post-processing step, further improved the accuracy of the CNN ensemble model by more than 6%. Hurricane Harvey was considered a case study to explain the advancement of the proposed hurricane forecasting model.

CHAPTER 7. CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this study, various applications of deep learning algorithms, particularly convolutional neural networks, were successfully executed in the field of atmospheric sciences, including in air quality forecasting, model post-processing, and hurricane tracking. These applications showed prediction results with promising accuracies with an easy-to-use, computationally-efficient framework and flexible capabilities.

The first task used a CNN model to forecast hourly ozone concentrations over Seoul, South Korea for 2017. Model-measurement comparisons for the 25 monitoring sites for the year 2017 yielded average indices of agreement (IOA) of 0.84-0.89 and a Pearson correlation coefficient of 0.74-0.81, indicating reasonable performance for the CNN forecasting model. The forecasting results were found to be generally more accurate for the stations located in the southern regions of the Han River, the result of more stable topographical and meteorological conditions. Furthermore, through two separate daytime and nighttime forecasts, I found that the monthly IOA of the CNN model is 0.05-0.30 higher during the daytime, resulting from the unavailability of some of the input parameters during the nighttime. Although the CNN model successfully captured daily trends as well as yearly high and low variations of the ozone concentrations, it notably underpredicted high ozone peaks during the summer. The high prediction bias was addressed by proposing a data ensemble approach in the next task.

In the second task, six generalized machine learning (ML) ensemble models were developed to predict the real-time hourly ozone concentration of the following day. The training, testing, and input variables were similar to those in first task. The ensemble models fused two regression models: a low-ozone peak model and a high-ozone model. For both, extremely randomized trees and deep neural networks were used. A regularization approach was also adopted that adjusts the model toward capturing higher ozone peaks by resampling the training dataset based on the peaks. My results indicated that dopting the proposed ML ensemble forecasting method over single-model ML techniques as a part of mainstream practice for air quality forecasting will be beneficial for several reasons. For one, the proposed method, which captures daily maximum ozone concentrations during the high ozone season (April-September), reduces the ozone peak prediction error by 5 to 30 ppb. In addition, compared to station-specific (independent) ML models with more frequent low-ozone values, models are trained with a uniformly distributed dataset, so they are more generalizable in nature. As a result, unlike station-specific models, they retain their accuracy (yearly IOA=0.84-0.89) in all stations with an IOA increment. The proposed models also make predictions several times faster, requiring only one-time training for predicting an entire station network. Based on a categorical analysis of the training dataset, an algorithm was proposed for selecting the most suitable model for each month. The "best" model further improved the accuracy of both the ML ensemble and individual models by up to 2.4 %. This study shows that the ML ensemble modeling approach is a fast, reliable, and robust technique that can benefit environmental decision makers in urban regions.

For the third task, a deep CNN was used to map ozone precursors from CMAQ model and meteorological parameters from the WRF model (as inputs variables) to observed hourly ozone concentrations at a monitoring station (as a target). The results show that the CMAQ-CNN model significantly improves the performance of the CMAQ model in both accuracy and bias. The absolute correlation coefficient is improved by 0.16 on average. The CMAQ-CNN model improves simulated ozone peaks for almost all cases and reduces the bias of CMAQ predictions by an average of more than 20 ppb (or 40%). Systematic improvements in CMAQ-CNN simulations suggest that the deep learning model is effective at reproducing accurate estimates of ground-level air quality concentrations. While this study focused on ozone in the United States and outputs of CMAQ, the proposed approach can be applied to any measured air pollution parameters or numerical model in a mesoscale resolution.

For the last task, a hybrid hurricane forecasting model was developed for predicting hurricane characteristics (track and intensity) 24 hours in advance. Dynamical models produce significant model-measurement errors in forecasting hurricanes, the result of the presence of both the chaotic growth of errors in the simulations of initial conditions and deficiencies in the physics of such models. To improve the prediction of hurricanes, this task proposed a novel, hybrid approach that uses a deep-learning approach and an ensemble Kalman filter (EnKF). The goal was to forecast the track and wind intensity of hurricanes 24 hours in advance. Using the output of dynamical hurricane models and observational data, I first developed a hybrid three-step (direction, distance traveled, and intensity) deep learning-based ensemble hurricane forecasting model. I used all tropical cyclones in the Atlantic and Pacific Oceans before 2017 and tested the model for cyclones in 2017. Then, to further reduce the prediction bias of the compiled deep learning-based model, I applied EnKF as a post-processing step. The preliminary results of the hybrid model for 17 tropical storms in 2017 show statistical advantages (a ~13% and ~34% improvement in track and intensity forecast biases, respectively) of official forecasts 24 hours ahead of the National Hurricane Center (NHC) forecasts. Hurricane Harvey served as a case study in this work.

Since deep learning algorithms have become popular data analytic techniques, it is important for atmospheric scientists to have a balanced perception of their strengths and limitations. They can provide a powerful analysis of complex data with well-established procedures. However, despite their enormous success in numerous applications, certain issues related to their applications in air quality forecasting (AQF) require further discussion. In this study, I tried to address significant limitations of CNNs, in two applications presented here: (i) a real-time AQF model, and (ii) a post-processing tool in a dynamical AQF model, CMAQ. For the first case, I used the wavelet transform to reveal the reasons behind the CNN's poor performance during the nighttime, cold months, and high ozone episodes. I found that when the fine wavelet modes (hourly and daily) were relatively weak or the coarse wavelet modes (weekly) were strong, the performance of the CNN model was significantly reduced. For the second case, I used the dynamic time warping (DTW) distance analysis to compare the post-processed results with their CMAQ counterparts (as base model). For CMAQ results that had a consistent DTW distance from the observation, the post-processing approach properly addressed the modeling bias with predicted IOAs more than 0.85. When there was no regularity in the DTW distance of CMAQ-vs-observation, the post-processing approach was unlikely to perform satisfactorily. These techniques will help researchers to become aware of the limitations of deep learning models by considering discrepancies in the input data and their temporal trends.

7.2 Future work

While the CNN model can predict the next 24 hours of ozone concentrations within less than a minute, I identified several limitations of deep learning models for real-time air quality forecasting for further improvement. I suggest that researchers prepare their deep learning modeling configuration based on the temporal trends within input parameters, geographical locations, and variation frequency of the target pollutant. While my study approach might remain valid for other supervised algorithms, I leave a detailed study of other methods as well as unsupervised problem for future work.

Deep neural networks can build accurate air quality forecasting models. However, the contributions of the input variables in the prediction model are generally unknown. This is especially important in analyzing the sensitivity of the air pollutant (e.g., ozone) to the changes of its precursors (e.g., NOx) in an urban environment. I suggest using deep learning to develop a 'dependency' model to quantify the input sensitivity in dynamical model such as CMAQ. This approach can be beneficial in understanding the important factors in selecting efficient air quality reduction policy in local or regional scale.

For future improvement of my hurricane forecasting model/forecasting, I suggest the following directions of study: (i) to develop a global hurricane modeling system; (ii) to develop an image processing CNN model that uses remote sensing data and a more comprehensive set of modeling outputs to generate a direct hurricane forecast; (iii) to apply more environmental variables with a longer history for training the deep learning model(s); (iv) to incorporate rainfall into the hurricane forecasting approach; and (v) to identify the causes of the biases of numerical models in order to tune them for more accurate forecasts.

BIBLIOGRAPHY

1. Pan S, Choi Y, Roy A, Li X, Jeon W, Souri AH. Modeling the uncertainty of several VOC and its impact on simulated VOC and ozone in Houston, Texas. Atmospheric Environment. 2015;120:404-16.

2. Pope III CA. Epidemiological basis for particulate air pollution health standards. Aerosol Science & Technology. 2000;32(1):4-14.

3. Peng H, Lima AR, Teakles A, Jin J, Cannon AJ, Hsieh WW. Evaluating hourly air quality forecasting in Canada with nonlinear updatable machine learning methods. Air Quality, Atmosphere & Health. 2017;10(2):195-211.

4. Eslami E, Pan S, Sadeghi, Bavand, Choi Y, editors. Health and cost impact of air pollution from biomass burning over the United States. AGU Fall Meeting, New Orleans, LA, USA; 2017.

5. Fann N, Alman B, Broome RA, Morgan GG, Johnston FH, Pouliot G, et al. The health impacts and economic value of wildland fire episodes in the US: 2008–2012. Science of the total environment. 2018;610:802-9.

6. Pan S, Roy A, Choi Y, Eslami E, Thomas S, Jiang X, et al. Potential impacts of electric vehicles on air quality and health endpoints in the Greater Houston Area in 2040. Atmospheric Environment. 2019;207:38-51.

7. Podrascanin Z. Setting-up a Real-Time Air Quality Forecasting system for Serbia: a WRF-Chem feasibility study with different horizontal resolutions and emission inventories. Environ Sci Pollut R. 2019;26(17):17066-79.

8. Donnelly A, Misstear B, Broderick B. Real time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmospheric Environment. 2015;103:53-65.

9. Ye FX, Liu XF, Liu YL. Design and Implementation of Real-Time Forecasting System of Urban Air Quality Based on Observation Sites and Thiessen Polygon. 2nd International Conference on Sensors, Instrument and Information Technology (Icsiit 2015). 2015:451-6.

10. Yahya K, Zhang Y, Vukoyich JM. Real-time air quality forecasting over the southeastern United States using WRF/Chem-MADRID: Multiple-year assessment and sensitivity studies. Atmospheric Environment. 2014;92:318-38.

11. Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A. Real-time air quality forecasting, part I: History, techniques, and current status. Atmospheric Environment. 2012;60:632-55.

12. Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A. Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects. Atmospheric Environment. 2012;60:656-76.

13. Chuang MT, Zhang Y, Kang DW. Application of WRF/Chem-MADRID for real-time air quality forecasting over the Southeastern United States (vol 45, pg 6241, 2011). Atmospheric Environment. 2012;60:677-8.

14. Sfetsos A, Bartzis JG. Real time, high resolution regional weather and air quality forecasting system in West Macedonia, Greece. Int J Environ Pollut. 2011;44(1-4):244-51.

15. Kang DW, Mathur R, Rao ST. Real-time bias-adjusted O-3 and PM2.5 air quality index forecasts and their performance evaluations over the continental United States. Atmospheric Environment. 2010;44(18):2203-12.

16. Berge E, Walker SE, Sorteberg A, Lenkopane M, Eastwood S, Jablonska HI, et al. A real-time operational forecast model for meteorology and air quality during peak air pollution episodes in Oslo, Norway. Urban Air Quality - Recent Advances, Proceedings. 2002:745-57.

17. Byun D, Schere KL. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. 2006.

18. Appel KW, Napelenok SL, Foley KM, Pye HO, Hogrefe C, Luecken DJ, et al. Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1. Geoscientific model development. 2017;10(4):1703-32.

19. Wong D, Pleim J, Mathur R, Binkowski F, Otte T, Gilliam R, et al. WRF-CMAQ two-way coupled system with aerosol feedback: software development and preliminary results. Geoscientific Model Development. 2012;5(2):299-312.

20. Schmidhuber J. Deep learning in neural networks: An overview. Neural networks. 2015;61:85-117.

21. Specht DF. A general regression neural network. IEEE transactions on neural networks. 1991;2(6):568-76.

22. Hecht-Nielsen R. Theory of the backpropagation neural network. Neural networks for perception: Elsevier; 1992. p. 65-93.

23. Ma J, Cheng JCP, Lin CQ, Tan Y, Zhang JC. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. Atmospheric Environment. 2019;214.

24. Sun XT, Xu W. Deep Random Subspace Learning: A Spatial-Temporal Modeling Approach for Air Quality Prediction. Atmosphere. 2019;10(9).

25. Li JG, Shao XY, Sun RH. A DBN-Based Deep Neural Network Model with Multitask Learning for Online Air Quality Prediction. J Control Sci Eng. 2019;2019.

26. Zhou YL, Chang FJ, Chang LC, Kao IF, Wang YS. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. J Clean Prod. 2019;209:134-45.

27. Qi ZG, Wang TC, Song GJ, Hu WS, Li X, Zhang ZF. Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-Grained Air Quality. Ieee T Knowl Data En. 2018;30(12):2285-97.

28. Freeman BS, Taylor G, Gharabaghi B, The J. Forecasting air quality time series using deep learning. J Air Waste Manage. 2018;68(8):866-86.

29. Li X, Peng L, Hu Y, Shao J, Chi TH. Deep learning architecture for air quality predictions. Environ Sci Pollut R. 2016;23(22):22408-17.

30. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436-44.

31. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Computation. 2006;18(7):1527-54.

32. Bengio Y, Lamblin P, Popovici D, Larochelle H, editors. Greedy layer-wise training of deep networks. Advances in neural information processing systems; 2007.

33. Nair V, Hinton GE, editors. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10); 2010.

34. Sutskever I, Hinton GE. Deep, Narrow Sigmoid Belief Networks Are Universal Approximators. Neural Computation. 2008;20(11):2629-36.

35. Busseti E, Osband I, Wong S. Deep learning for time series modeling. Technical report, Stanford University. 2012:1-5.

36. Kuremoto T, Kimura S, Kobayashi K, Obayashi M. Time series forecasting using a deep belief network with restricted Boltzmann machines. Neurocomputing. 2014;137:47-56.

37. Qiu X, Zhang L, Ren Y, Suganthan PN, Amaratunga G, editors. Ensemble deep learning for regression and time series forecasting. 2014 IEEE symposium on computational intelligence in ensemble learning (CIEL); 2014: IEEE.

38. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. Journal of Big Data. 2015;2(1):1.

39. Hraskoa R, Pacheco AGC, Krohling RA. Time Series Prediction using Restricted Boltzmann Machines and Backpropagation. Procedia Comput Sci. 2015;55:990-9.

40. Li HO, Hou J, Adhikari B, Lyu Q, Cheng JL. Deep learning methods for protein torsion angle prediction. Bmc Bioinformatics. 2017;18.

41. Zhang JH, Yan J, Infield D, Liu YQ, Lien FS. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Appl Energ. 2019;241:229-44.

42. Li X, Peng L, Yao X, Cui S, Hu Y, You C, et al. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. Environmental pollution. 2017;231:997-1004.

43. Li T, Shen H, Yuan Q, Zhang X, Zhang L. Estimating ground-level PM2. 5 by fusing satellite and station observations: A geo-intelligent deep learning approach. Geophysical Research Letters. 2017;44(23):11,985-11,93.

44. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Commun Acm. 2017;60(6):84-90.

45. Sayeed A, Choi Y, Eslami E, Lops Y, Roy A, Jung J. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. Neural Networks. 2020;121:396-408.

46. Hu QH, Zhang RJ, Zhou YC. Transfer learning for short-term wind speed prediction with deep neural networks. Renew Energ. 2016;85:83-95.

47. Raissi M, Karniadakis GE. Hidden physics models: Machine learning of nonlinear partial differential equations. J Comput Phys. 2018;357:125-41.

48. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313(5786):504-7.

49. Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9(8):1735-80.

50. Zhang LM, Wang B, Fang BW, Ma HR, Yang Z, Xu YY. Two-stage Short-Term Wind Speed Prediction Based on LSTM-LSSVM-CFA. 2018 2nd Ieee Conference on Energy Internet and Energy System Integration (Ei2). 2018.

51. Eslami E, Choi Y, Lops Y, Sayeed A, editors. A Deep Learning Driven Improved Ensemble Approach for Hurricane Forecasting. 2019 ESIP Winter Meeting; 2019: https://esip.figshare.com/articles/Ebrahim_Eslami_Hurricane_pdf/7591775.

52. Eslami E, Choi Y, Lops Y, Sayeed A. A real-time hourly ozone prediction system using deep convolutional neural network. Neural Computing and Applications. 2019:1-15.

53. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks. 1995;3361(10):1995.

54. Sabour S, Frosst N, Hinton GE. Dynamic Routing Between Capsules. Adv Neur In. 2017;30.

55. Pan S, Choi Y, Jeon W, Roy A, Westenbarger DA, Kim HC. 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. 2017;152:362-76.

56. Pan S, Choi Y, Roy A, Jeon W. Allocating emissions to 4 km and 1 km horizontal spatial resolutions and its impact on simulated NOx and O3 in Houston, TX. Atmospheric environment. 2017;164:398-415.

57. Jeon W-B, Lee S-H, Lee H-W, Kim H-G. Process analysis of the impact of atmospheric recirculation on consecutive high-O3 episodes over the Seoul Metropolitan Area in the Korean Peninsula. Atmospheric environment. 2012;63:213-22.

58. Jeon W-B, Lee S-H, Lee H, Park C, Kim D-H, Park S-Y. A study on high ozone formation mechanism associated with change of NOx/VOCs ratio at a rural area in the Korean Peninsula. Atmospheric Environment. 2014;89:10-21.

59. Mazumder R, Hastie T, Tibshirani R. Spectral regularization algorithms for learning large incomplete matrices. Journal of machine learning research. 2010;11(Aug):2287-322.

60. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al., editors. Tensorflow: A system for large-scale machine learning. 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16); 2016.

61. Chen XW, Lin XT. Big Data Deep Learning: Challenges and Perspectives. Ieee Access. 2014;2:514-25.

62. Gu JX, Wang ZH, Kuen J, Ma LY, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recogn. 2018;77:354-77.

63. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.

64. Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE transactions on information theory. 1990;36(5):961-1005.

65. Mallat SG. Multiresolution Approximations and Wavelet Orthonormal Bases of L2(R). T Am Math Soc. 1989;315(1):69-87.

66. Grinsted A, Moore JC, Jevrejeva S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear processes in geophysics. 2004;11(5/6):561-6.

67. Domingues MO, Mendes Jr O, Da Costa AM. On wavelet techniques in atmospheric sciences. Advances in Space Research. 2005;35(5):831-42.

68. Kaheil YH, Rosero E, Gill MK, Mckee M, Bastidas LA. Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines. Ieee T Geosci Remote. 2008;46(9):2692-707.

69. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Machine learning. 2006;63(1):3-42.

70. Quinlan JR. Learning decision tree classifiers. Acm Comput Surv. 1996;28(1):71-2.

71. Quinlan JR. Simplifying Decision Trees. Int J Man Mach Stud. 1987;27(3):221-34.

72. Lariviere B, Van den Poel D. Predicting customer retention and profitability by using random forests and regression forests techniques. Expert Syst Appl. 2005;29(2):472-84.

73. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825-30.

74. Djalalova I, Delle Monache L, Wilczak J. PM2. 5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. Atmospheric Environment. 2015;108:76-87.

75. Neal L, Agnew P, Moseley S, Ordóñez C, Savage N, Tilbee M. Application of a statistical post-processing technique to a gridded, operational, air quality forecast. Atmospheric environment. 2014;98:385-93.

76. Cobourn WG. An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmospheric Environment. 2010;44(25):3015-23.

77. Struzewska J, Kaminski JW, Jefimow M. Application of model output statistics to the GEM-AQ high resolution air quality forecast. Atmos Res. 2016;181:186-99.

78. Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, et al. A description of the advanced research WRF version 2. National Center For Atmospheric Research Boulder Co Mesoscale and Microscale …; 2005.

79. Choi Y, Jeon W, Roy A, Souri A, Diao L, Pan S, et al. CMAQ Modeling Archive for Exceptional Events Analysis, Final report to the Texas Commission on Environmental Quality (TCEQ). 2016.

80. Souri AH, Choi Y, Jeon W, Kochanski AK, Diao LJ, Mandel J, et al. Quantifying the Impact of Biomass Burning Emissions on Major Inorganic Aerosols and Their Precursors in the US. J Geophys Res-Atmos. 2017;122(21):12020-41.

81. Jeon W, Choi Y, Souri AH, Roy A, Diao L, Pan S, et al. Identification of chemical fingerprints in long-range transport of burning induced upper tropospheric ozone from Colorado to the North Atlantic Ocean. Science of the Total Environment. 2018;613:820-8.

82. Gilliland AB, Hogrefe C, Pinder RW, Godowitch JM, Foley KL, Rao ST. Dynamic evaluation of regional air quality models: Assessing changes in O-3 stemming from changes in emissions and meteorology. Atmospheric Environment. 2008;42(20):5110-23.

83. Cheng WYY, Steenburgh WJ. Evaluation of surface sensible weather forecasts by the WRF and the Eta Models over the western United States. Weather Forecast. 2005;20(5):812-21.

84. Hu XM, Nielsen-Gammon JW, Zhang FQ. Evaluation of Three Planetary Boundary Layer Schemes in the WRF Model. J Appl Meteorol Clim. 2010;49(9):1831-44.

85. Kim HC, Lee P, Ngan F, Tang Y, Yoo HL, Pan L. Evaluation of modeled surface ozone biases as a function of cloud cover fraction. Geoscientific Model Development. 2015;8(9):2959-65.

86. David LM, Nair PR. Diurnal and seasonal variability of surface ozone and NOx at a tropical coastal site: Association with mesoscale and synoptic meteorological conditions. J Geophys Res-Atmos. 2011;116.

87. Golaszewski G. Similarity-Based Outlier Detection in Multiple Time Series. Adv Intell Syst. 2020;945:116-31.

88. Marks FD, Shay LK. Landfalling tropical cyclones: Forecast problems and associated research opportunities. Bulletin of the American Meteorological Society. 1998;79(2):305-23.

89. Gopalakrishnan SG, Marks F, Zhang XJ, Bao JW, Yeh KS, Atlas R. The Experimental HWRF System: A Study on the Influence of Horizontal Resolution on the Structure and Intensity Changes in Tropical Cyclones Using an Idealized Framework. Monthly Weather Review. 2011;139(6):1762-84.

90. Balaguru K, Foltz GR, Leung LR. Increasing Magnitude of Hurricane Rapid Intensification in the Central and Eastern Tropical Atlantic. Geophysical Research Letters. 2018;45(9):4238-47.

91. Dorian T, Ward B, Chen YL. Tropical Cyclone Amos (2016) Forecasting Challenges: A Model's Perspective. Trop Cyclone Res Rev. 2018;7(3):172-8.

92. Newman B, Gallion C. Hurricane Harvey: Firsthand Perspectives for Disaster Preparedness in Graduate Medical Education. Acad Med. 2019;94(9):1267-9.

93. Done JM, Holland GJ, Bruyere CL, Leung LR, Suzuki-Parker A. Modeling high-impact weather and climate: lessons from a tropical cyclone perspective. Climatic Change. 2015;129(3-4):381-95.

94. Boas ABV, Ardhuin F, Ayet A, Bourassa MA, Brandt P, Chapron B, et al. Integrated Observations of Global Surface Winds, Currents, and Waves: Requirements and Challenges for the Next Decade. Front Mar Sci. 2019;6.

95. Lee CY, Tippett MK, Sobel AH, Camargo SJ. Autoregressive Modeling for Tropical Cyclone Intensity Climatology. J Climate. 2016;29(21):7815-30.

96. Lee CY, Tippett MK, Camargo SJ, Sobel AH. Probabilistic Multiple Linear Regression Modeling for Tropical Cyclone Intensity. Monthly Weather Review. 2015;143(3):933-54.

97. Racah E, Beckham C, Maharaj T, Kahou SE, Prabhat, Pa C. ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. Adv Neur In. 2017;30.

98. Alemany S, Beltran J, Perez A, Ganzfried S. Predicting Hurricane Trajectories Using a Recurrent Neural Network. Thirty-Third Aaai Conference on Artificial Intelligence / Thirty-First Innovative Applications of Artificial Intelligence Conference / Ninth Aaai Symposium on Educational Advances in Artificial Intelligence. 2019:468-75.

99. Kim S, Kim H, Lee J, Yoon S, Kahou SE, Kashinath K, et al. Deep-Hurricane-Tracker: Tracking and Forecasting Extreme Climate Events. Ieee Wint Conf Appl. 2019:1761-9.

100. Evensen G, van Leeuwen PJ. An ensemble Kalman smoother for nonlinear dynamics. Monthly Weather Review. 2000;128(6):1852-67.

101. Louka P, Galanis G, Siebert N, Kariniotaki G, Katsafados P, Pytharoulis I, et al. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J Wind Eng Ind Aerod. 2008;96(12):2348-62.

102. Tatang MA, Pan WW, Prinn RG, McRae GJ. An efficient method for parametric uncertainty analysis of numerical geophysical models. J Geophys Res-Atmos. 1997;102(D18):21925-32.

103. Dietterich TG. Ensemble methods in machine learning. Multiple Classifier Systems. 2000;1857:1-15.

104. Joodat SHS, Nakshatrala KB, Ballarini R. Modeling flow in porous media with double porosity/permeability: A stabilized mixed formulation, error analysis, and numerical solutions. Comput Method Appl M. 2018;337:632-76.

105. Hamill TM, Whitaker JS, Fiorino M, Benjamin SG. Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter. Monthly Weather Review. 2011;139(2):668-88.

106. Houtekamer PL, Zhang FQ. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. Monthly Weather Review. 2016;144(12):4489-532.

107. Roth M, Hendeby G, Fritsche C, Gustafsson F. The Ensemble Kalman filter: a signal processing perspective. Eurasip J Adv Sig Pr. 2017.

108. Kiaghadi A, Rifai HS. Physical, Chemical, and Microbial Quality of Floodwaters in Houston Following Hurricane Harvey. Environ Sci Technol. 2019;53(9):4832-40.