Application of Deep Learning for Air Quality Predictions, Remote Sensing Processing and Long-Term Climate Index Forecasting

by

Yannic Lops

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

DOCTOR OF PHILOSOPHY

in Atmospheric Sciences

Chair of Committee: Yunsoo Choi
Committee Member: Xun Jiang
Committee Member: Yuxuan Wang
Committee Member: Christoph F. Eick

University of Houston

December 2021

DEDICATION

In dedication to my beloved grandparents, my dear aunt Andrea, and my friend Mitch,

in memoriam.

ACKNOWLEDGMENTS

I would like to express my deep and sincere gratitude to my supervisor, Professor Yunsoo Choi, for giving me the opportunity to join his team and providing invaluable guidance throughout this process, for the continuous support over the course of my Ph.D. work and research, for his patience, motivation, and sharing his immense knowledge.

My sincere thanks to my thesis committee members: Dr. Xun Jiang, Dr. Yuxuan Wang, and Dr. Christoph Eick, for their insightful comments and encouragement that enabled me to widen my research from various perspectives.

I am especially grateful to my current and former AI team members, Dr. Alqamah, Sayeed, Dr. Ebrahim Eslami, Ahmed Khan Salman, Masoud Ghahremanloo on the varying excellent discussions we had in the field of machine learning and atmospheric science which greatly contributed to my research. I would like to also thank Dr. Anirban Roy, Arman Pouyaei, Jia Jung, Ali Mousavinezhad, Bavand Sadeghi, Delaney Nelson, and Deveshwar Singh for working together on various projects.

Most importantly, none of this could have happened without my mother, her unconditional support, her wise counsel, and sympathetic ear. I would like to extend my gratitude to my stepfather, who encouraged me to apply at the University of Houston and who supported me throughout the entire process.

My appreciation also goes out to my entire family and friends for their encouragement and support during my studies. They all provided an important stability in my life while my parents raised me abroad. I cannot express how much your unwavering support meant to me.

ABSTRACT

The advancement and development of new technology provide atmospheric scientists and modelers to acquire an overwhelming amount of data on meteorology and air quality from space, numerical simulations, and in-situ monitoring sites. Integrating these data sources provides unique opportunities to enhance understanding of atmospheric processes to better simulate and forecast these processes. While global climate models and chemical transport models have undergone significant optimizations and improvements over the past decades, they are still unable to provide fully reliable biogenic air quality predictions or long-term climate forecasting. These limitations can be alleviated and addressed by incorporating in-situ measurements and remote sensing products into data assimilation or reanalysis techniques. While ground-based remote sensing measurements provide detailed point observations, they lack the spatial coverage of remote sensing-derived measurements. Unfortunately, these remote sensing measurements experience issues caused by outside factors such as cloud cover contamination and false reflectance. Internal issues involve sensor errors that corrupt or lead to failed measurements of the data.

This study utilizes the advanced capability of several deep learning models for the forecasting of pollen concentrations by up to 7 days; the imputation of remote sensing measurements spatially with partial convolutional neural networks and subsequent revision to incorporate spatio-temporal imputation; and long-term forecasting system of the climate index Nino3.4 by up to 36 months.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Particulate Matter (PM) is a form of aerosol that can originate from dust transport (Wang et al., 2008), biomass burning (Uranishi et al., 2019), and emissions from anthropogenic sources (Kang et al., 2019) and natural sources (Zhang et al., 2016). A large fraction of secondary aerosols is produced from the condensation of precursor gases or nucleation via cloud processes (Ervens et al., 2011). Particulate matter (PM), a major air pollutant worldwide (Koulouri et al., 2008; W. Li et al., 2014; Mukherjee & Agrawal, 2017), and has a significant impact on human health often associated with respiratory and cardiovascular diseases and mortalities (Brunekreef & Holgate, 2002). Pollen is a form of biological aerosol that also has deleterious effects on human health. Allergic rhinitis, for instance, is estimated to affect up to 30% of adults and 40% of children globally (Pawankar et al., 2013), and hay fever and allergic asthma as much as 25% of the population (Traidl-Hoffmann et al., 2003). In addition to aerosols, the emission of greenhouse gases and aerosols affect the atmosphere's physics and chemistry (Mikhaylov et al., 2020; Wu et al., 2018). Based on these impacts on the atmosphere, have led to temperature rise; longer, more intense droughts and storms; increased risk of wildfires (Jones et al., 2020); and sea-level rise (Vu et al., 2018). All these factors will continue to affect water supply and quality (Hashempour et al., 2020), crop and food production (Lal, 2020), and human health (Campbell-Lendrum & Prüss-Ustün, 2019).

As such, it is imperative that proper forecasting models are developed to inform and adopt strategies to mitigate the impact on the environment and human health. Many researchers have

devoted a great deal of effort to develop systems that improve forecasting capabilities in air quality and climate change, they have been limited by the quality and amount of data (Ghahremanloo et al., 2020), the performance of chemical transport and numerical prediction models (Park et al., 2011; Pouyaei et al., 2020), the complexity of reliably simulating complex processes within the climate due to its chaotic yet deterministic fluctuations (Chen et al., 2004; Gupta et al., 2020). To address these limitations, incorporating both in-situ measurements and remote sensing products into data assimilation techniques (Bocquet et al., 2015; Jung et al., 2019) would allow the models to reduce their biases. Unfortunately, their accuracies are limited to the availability and quality of in-situ and remote sensing data products.

This dissertation aims to address these limitations of the non-linear processes within the atmosphere in air quality forecasts, the remote sensing data limitations, and the complexity of long-term climate forecasting by the utilization of Deep Learning (DL) algorithms. DL uses neural networks, which were inspired by an understanding of the brain, which can deal with nonlinear variables and yield higher accuracy than conventional regression and statistical models (Zhang et al., 2016; Gu et al., 2018). Among the various DL algorithms, convolutional neural networks (CNNs) (Krizhevsky et al., 2017) have been among the most complicated, as well as successful and widely used models (LeCun & Bengio, 1995; Schmidhuber, 2015) for various purposes. CNNs use filters to process and convolve data into a more manageable form for processing and extract high-level features of the input data (Lawrence et al., 1997).

The dissertation first utilizes and demonstrates the advanced capabilities of CNN models to forecast pollen concentrations, which are known for their strong seasonal variation and

difficulty for long-term forecasts. Furthermore, the CNN model is implemented for remote sensing processing applications to accurately impute missing data at large spatial distances, both spatially and temporally. Finally, multiple deep CNN models are developed and compared to provide the optimal ensemble of CNN models to forecast the El Niño-Southern Oscillation (ENSO) index by up to 36-months with better accuracy than previous models.

CHAPTER 2

Real-time 7-Day Forecast of Pollen Counts Using a Deep Convolutional Neural Network[1]

2.1 Introduction

Pollen has deleterious effects on human health. Allergic rhinitis is estimated to affect up to 30% of adults and 40% of children (Pawankar et al., 2013), and hay fever and allergic asthma as much as 25% of the population (Traidl-Hoffmann et al., 2003). Effective allergen avoidance has shown improvement in allergy symptoms (Pawankar et al., 2013). For people with allergic symptoms, awareness of where and when elevated concentrations of pollen will occur is critical for their health (Vogel et al., 2008).

Annual pollen concentrations vary considerably due to species-specific and weather-related factors (Jäger et al., 1989). Several studies examined the relationships between meteorology, pollen concentrations, and seasonal pollen trends (García-Mozo et al., 2014; Myszkowska et al., 2014; Rojo et al., 2015; Toro et al., 2015). Temperature has the strongest influence and the strongest relationship to pollen emissions and concentrations (Bartková-Ščevková, 2003). Several studies applied regression analyses and models (Angosto et al., 2005; Fairley & Batchelder, 1986; Jeon et al., 2018; Liu et al., 2015; Vogel et al., 2008; de Weger et al., 2014) to predict pollen concentrations. Studies have spatially (Levetin & Van de Water, 2003) and

[1] Published: Lops, Y., Choi, Y., Eslami, E., & Sayeed, A. (2019). Real-time 7-day forecast of pollen counts using a deep convolutional neural network. *Neural Computing and Applications*, 1-10.

temporally (Wozniak et al., 2017; Zhang et al., 2013) modeled and predicted pollen counts with satisfactory results. Jeon et al. (2018) developed the Community Multiscale Air Quality Modeling System pollen model (CPM) to predict oak pollen concentrations. Evaluations on pollen forecasting applications have shown an accuracy of 50% on average (Bastl et al., 2017). Although few studies have shown a slight improvement of neural networks over regression analyses in air-quality forecasting (Eslami et al., 2019a), few have applied them to forecasting pollen counts (Sánchez et al., 2007; Sánchez-Mesa et al., 2002).

An artificial neural network (ANN) is a layered structure of algorithms. One form of the neural network is the multi-layer perceptron (MLP) with the most basic one consisting of an input layer (data that feed the neural network); hidden layers (transform the inputs into information that the output layer can use); and an output layer (transforms the hidden layer activations into a scale we can define) such as classification and regression. Deep neural networks consist of multiple hidden layers, each of which contains multiple neurons (mathematically mimic a biological neuron using activation functions). These networks were inspired by an understanding of the brain, which is an interconnection of billions of neurons. Deep neural networks dealing with nonlinear variables yield higher accuracy than conventional neural networks and regression models (Zhang et al., 2016; Gu et al., 2018). One of the most complicated neural networks is a convolutional neural network (CNN) (Krizhevsky et al., 2012) which uses filters to convolve input data into multiple convolutions of input data. The motivation for our study is developing a CNN system that forecasts real-time pollen counts with greater accuracy and less processing time than current models.

2.2 Methods

2.2.1 Data

We acquired pollen data from the Houston Department of Health and Human Services (HDHHS) archives. The study focused on daily pollen data from 2009 to 2016. The most common pollen-producing species were comprised of tree and weed categories. Tree and weed pollen were composed of 25 and 14 species, respectively. Pollen concentrations were measured south of the Houston downtown area. Meteorological data were gathered at the Moody Tower station, 4.5 km east-northeast of the pollen station. Figure 2.1 exhibits a map of the study area with the station locations. We obtained meteorological data from the Texas Commission on Environmental Quality (TCEQ), which operates the Continuous Ambient Monitoring Sites (CAMS) in various metropolitan areas within the state of Texas. We selected data from CAMS station 695 (Moody Tower, near Downtown Houston) for its close proximity to the HDHHS pollen measurement station. We extracted the data of surrounding stations CAMS001, CAMS053, CAMS409, and CAMS416 as input for missing data from CAMS695 station. The hourly meteorological data was processed to daily intervals to correspond to the pollen data. The data were comprised of mean temperature (Celsius), total precipitation (mm), mean U and V wind components, mean wind speed (m/s), friction, and radiation (Langley). Relative humidity (%) and pressure (mb) are comprised of daily minimum, mean, and maximum measurements.

We acquired the leaf area index (LAI) data from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard the Terra (EOS AM-1) and Aqua (EOS PM-1) satellites. The spectral resolution of the MODIS data was 500 m, and LAI data comprised of 4-day measurements. LAI measurements were based on a quadrilateral square area that encompassed the entire city of Houston (see Figure 2.1). The center point of the area, the location of the HDHHS pollen station, had a radius of an estimated 45km to each perpendicular side of the measurement area. The total area covered is estimated 2,025 km^2. Since the variations in mean LAI were minimal, we performed linear interpolation between the 4-day measurements to coincide with the daily measurements of the pollen and meteorological data.

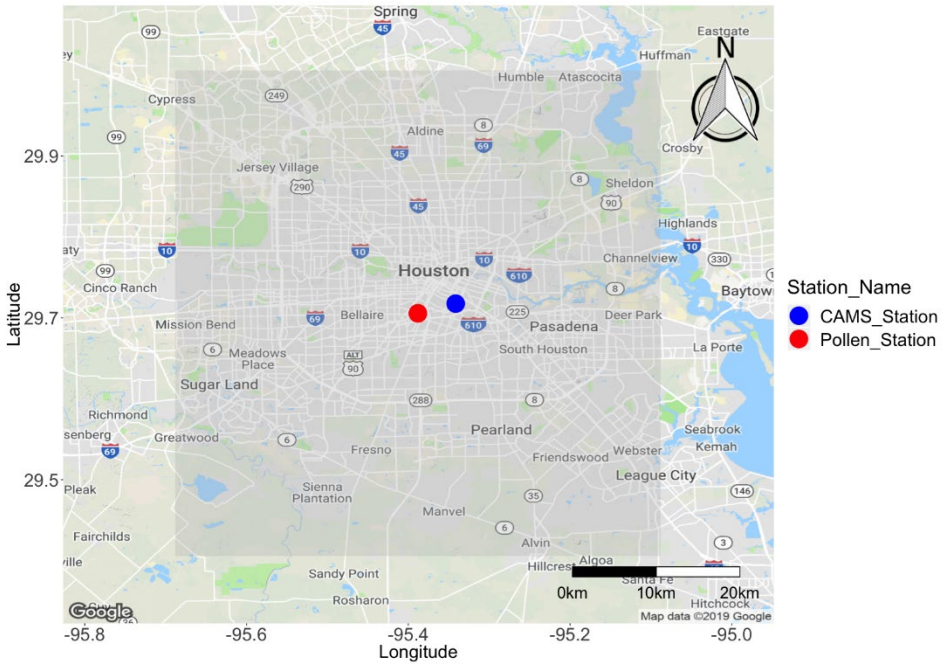

Figure 2.1: Map of the study area in Houston. CAMS Station (Blue) is the location of collected meteorology data. Pollen Station (Red) is the location of collected pollen concentration data. A grey area (centered on the Pollen Station) represents the Leaf Area Index (LAI) process based on MODIS data for pollen flux calculations.

2.2.2 Processed Data

We applied processed data to represent multiple variables representing initial conditions of pollen concentrations. The processed data consist of a meteorological adjustment factor (K_e), normal pollen distribution (C_e), characteristic concentration (C^*), averaged frictional velocity (u_*), and pollen flux (F_p). The meteorological adjustment factor represents the resistance of pollen release based on meteorological conditions (Helbig et al., 2004), comprising of three meteorological factors (temperature, relative humidity, and wind speed) that affect the pollen release from plants. The equation for the adjustment factor is:

$$K_e = 1 - \frac{3}{c1\frac{T}{T_{te}} + c2\frac{RH}{RH_{te}} + c3\frac{WS}{WS_{te}}}. \quad (1)$$

T_{te}, RH_{te}, and WS_{te} represent the threshold values for temperature, relative humidity, and wind speed, respectively. We calculated the adjustment factor from the threshold values of the most common species for the respective groups. Oak pollen comprised over 54% of total tree pollen, and ragweed accounted for over 93% of total weed pollen. Thus, oak (Jeon et al., 2018) and ragweed (Zink et al., 2012) threshold values of temperature, relative humidity, and wind speed are used to represent the respective pollen vegetation. Grass pollen threshold parameters were not available; thus, oak pollen parameters were selected because of the seasonal similarity between the two pollen categories. C_1, C_2, and C_3 are weighting factors that weigh the influence of meteorological resistance. Table T1 lists the threshold values and weighting factors for each pollen category. Adjustment factors for each pollen category were computed.

Normal pollen distribution (C_e) represents the mean normal pollen distribution for each pollen vegetation category by imitating the seasonal pollen cycle for each category. C_e is defined as:

$$C_e = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(d-\mu)^2}{2\sigma^2}}, \quad (2)$$

Where d is the number of consecutive days in which pollen measurements meet or exceed the pollen count, μ is the mean distribution, and σ is the standard deviation for a normal distribution graph. Respective statistical variables were not suitable due to strong variance in the pollen in the time series. Therefore, we manually selected the μ and σ variables where C_e best represents the pollen trend for the years 2009-2012. The μ and σ parameters of tree pollen were 50 and 15, respectively, and those for both grass and weed were 30 and 10.

The pollen flux (Fp) is the daily emission flux of pollen particles for each pollen vegetation type. The computation of pollen flux is:

$$F_p = C_e \cdot C^* \cdot K_e \cdot u_*, \quad (3)$$

where C_e represents the normal pollen distribution, K_e the meteorological adjustment factor, u_* the averaged frictional velocity, and C^* the characteristic concentration. C^* is defined as follows:

$$C^* = \frac{p^q}{LAI \cdot h^c}, \quad (4)$$

Where canopy height (h^c) is the mean canopy height of the vegetation species, the canopy heights for each category were set at 6.38m for the tree (Jeon et al., 2018), 0.1m for grass (Zhang et al., 2013), and 1.0m for weed (Liu et al., 2015). The *LAI* is the computed mean LAI from MODIS satellite image data for the respective time period of the area surrounding the pollen station (see Figure 2.1). p^q is set as 'Pollen Count +1'. From multiple experiments with the data,

we found that the model can be trained more efficiently with values greater than zero. During training, zero values may cause the model to ignore the data, reducing the number of training sets. We added a value of 1 to the pollen count to reduce zero values within the data, preventing the model from becoming a naïve predictor. Naïve predictor ignores the importance of the other input variables. Thus the model placed greater weight on and used last day's pollen count as its forecast and artificially produced favorable statistical results. Previous studies (García-Mozo et al., 2014; Eslami et al., 2019a; Sánchez et al., 2007) had also observed this phenomenon for regression or simple neural network models when they included only pollen grain counts within the dataset.

We normalized all input data to reduce the magnitude between the various input data. This prevents one feature from having more influence than another or causing dramatic changes in the weight matrix when the CNN model was optimized.

2.2.3 Neural Network System

Our CNN model used 24 normalized input variables for the prediction of the next 1-7 day counts of tree, grass, and weed pollen. Input variables comprised of meteorology (13 variables), LAI, fraction of photosynthetically active radiation (FPAR), meteorological adjustment factor, normal pollen distribution, and pollen flux for each of the three pollen categories. See Figure 2.2 for a graphical representation of the pollen forecasting CNN model.

The model was comprised of five main layers: an input layer, two one-dimensional convolutional layers, a fully connected layer, and an output layer. We applied a dropout

(Srivastava et al., 2014) layer between the convolutional layers and optimized the system parameters for each vegetation category (see Table T2 for parameter details). Optimal system parameters were identified using a trial and error method over multiple different parameters ranges consisting of Kernel Size (line segment shape at least 1×2 size), Number of Filters, Learning Rate, Batch Size, and Training Epochs for each pollen category, then evaluated their performance based on 1-day forecast accuracy. Cases with multiple favorable results of differing settings underwent a second set of testing runs to evaluate the stability of the models to further identify the optimal parameters.

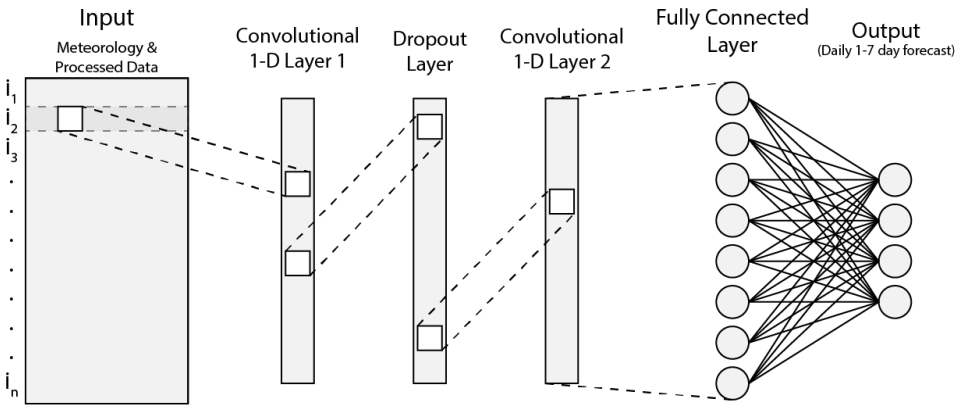

Figure 2.2: Representation of the pollen forecasting convolutional neural network model consisting of an input layer, two convolutional layers, a dropout layer in between the convolutional layers, a fully connected layer, and an output layer.

We trained the model with 2009-2015 data with 15-20% of the training set used for cross-validation purposes for the pollen forecast. The procedure is conducted to avoid overfitting the model. Once the training and validation run was complete, the model received the 2016 normalized data and predicted the next 1 to 7-day pollen for each vegetation category and total pollen prediction.

The CNN model has been compared to Recurrent Neural Network (RNN) (Barbounis et al., 2006; Tsoi & Back 1994) and Deep Neural Network (DNN) for the purposes of forecasting pollen concentration. We implemented a Gated Recurrent Unit (GRU) neural network as a representative and advanced form of an RNN (Chung et al., 2014; Dey & Salemt, 2017; Fu et al., 2016), which is a form of neural network generally suited for temporal sequences. The DNN model uses multiple layers of artificial neurons with dropout layers in between. All models received the same input data with some adjustments to their initial parameters; see Table T3 for parameter details.

The neural network models have been compared based on their statistical prediction capability in forecasting tree, grass, and weed pollen 1 and 7-days ahead. The CNN model has performed consistently better than both the DNN and GRU models in nearly all cases. Furthermore, the mean training time of the CNN model was the fastest of the models tested, with the DNN model close behind and the GRU model taking about 5 times longer than the CNN model. The GRU model was consistently the least accurate in predicting tree and grass pollen for both 1 and 7-day predictions than the other models. The exception was 7-day prediction of weed pollen, where GRU achieved 1% better IOA than the CNN model. The DNN model generally performed 4% less accurately than the CNN model in both 1- and 7-day forecasting of all three pollen categories. See Table T4 in the Appendix section for details on the performances of each model. Thus, this study implemented the CNN model for the purposes of forecasting tree, grass, and weed pollen 1-7 days ahead due to the model's more accurate and stable results and the faster performance in training time.

2.3 Results and Discussion

For our evaluation, we used the Pearson correlation coefficient (*r*), the Index of Agreement (IOA) (Willmott et al., 1985), and a categorical statistics evaluation as presented by (Chai et al., 2013). The categorical statistics evaluation consists of hit rate (HIT), critical success index (CSI), false alarm rate (FAR), equitable threat score (ETS), and proportion of correct (POC). We evaluated each pollen category and the days predicting ahead and compared observed to predicted pollen concentrations using *r* and IOA statistical evaluation methods. The evaluation of categorical statistics evaluation determined how well the model, compared to the observations, captured threshold levels based on the prevalence of allergy symptoms.

2.3.1 Evaluation of the Categorical Statistics

The evaluation of the categorical statistics was based on four quadrants:

N_a. Predictions above and observations below the threshold

N_b. Prediction and observation above the threshold

N_c. Predictions and observations below the threshold

N_d. Predictions below and observations above the threshold

From these quadrants, we evaluated the categorical statistics as follows:

$$HIT = \frac{N_b}{N_b + N_d} \tag{5}$$

$$CSI = \frac{N_b}{N_a + N_b + N_d} \tag{6}$$

$$FAR = \frac{N_a}{N_a + N_b} \tag{7}$$

$$ETS = \frac{N_b - N_r}{N_a + N_b + N_d - N_r} \tag{8}$$

$$N_r = \frac{(N_a + N_b) \cdot (N_b + N_d)}{N_a + N_b + N_c + N_d} \tag{9}$$

$$POC = \frac{N_b + N_c}{N_a + N_b + N_c + N_d}. \tag{10}$$

HIT is the fraction of observed pollen concentrations above the threshold that are predicted correctly by the model (1 is the best). FAR is the fraction of predicted pollen concentrations above the threshold that are false (0 is the best). CSI is the fraction of correctly predicted pollen concentrations above the threshold after the removal of correctly predicted pollen concentrations below the threshold value (1 is the best). ETS measures the performance skill of the model (1 is the best). POC is the fraction of the model forecast that matched the observations above and below the threshold (1 is the best).

For the purpose of the evaluation of the categorical statistics, we defined the threshold levels according to the severity of symptoms and most prevalent pollen species for the respective pollen categories. Oak (Quercus) pollen will represent the tree pollen evaluation of all tree pollen counts (grains m^{-3}). Pollen levels defined by the NAB are based on the percentile ranges of the pollen counts measured by all stations certified by the NAB. The NAB defines pollen counts between 15 and 89 grains m^{-3} as moderate. Soldevilla et al. (2007) categorized biological air quality (BAQ) into four levels (good, acceptable, poor, and bad) based on the frequency of

pollen types and their allergic potential. Poor BAQ is set as a threshold baseline, which refers to pollen types with moderate pollen counts but high allergic potential. Pollen counts of moderate (51-200 grains m^{-3}) for a specific group of tree species (i.e., Cupressus, Pinus, Platanus, Populus, and Quercus) accounted for 81% of the total tree pollen count. Another group of tree species with pollen counts defined as moderate, with 31-50 pollen grains m^{-3}, accounted for roughly 9% of the total tree pollen count. As a compromise for the purpose of the evaluation, we defined the pollen threshold for the three pollen counts as 50.

Grass pollen concentrations from 30 grains m^{-3} day^{-1} to 80 grains m^{-3} day^{-1} substantially increased allergic nose and eye symptoms in children. When grass pollen concentrations exceeded 70 grains m^{-3} day^{-1}, the severity of lung dysfunction symptoms increased (Kiotseridis et al., 2013). For the evaluation, we defined the grass pollen threshold as 30 because of the prevalence of allergy symptoms occurring with grass pollen counts of 30 grains m^{-3}.

In our evaluation, ragweed (Ambrosia) pollen represents the majority of weed pollen, which accounted for about 93% of all weed pollen counts. Ragweed pollen counts as low as 5 grains m^{-3} can cause allergic symptoms (Prank et al., 2013; Banken et al., 1990). Other studies have indicated that symptomatic experiences caused by ragweed pollen occur at higher concentrations ranging from 20-50 grains m^{-3} (Emberlin, 1994; Makra et al., 2011; Martin et al., 2010). The NAB defines pollen counts between 10 and 49 pollen grains m^{-3} as moderate. Thus, as a large number of studies have indicated that patients experience allergy symptoms in conditions within a diverse range, we set the threshold values for ragweed pollen at 20 grains m^{-3} to evaluate the model.

To evaluate the categorical statistics of the total pollen count (sum of the tree, grass, and weed pollen) for the entire year of 2016, we used both the threshold values of each of the pollen categories and the mean threshold of all three categories. Thus, the total pollen model performance had four categorical evaluation thresholds (see Table T5).

The results of the statistical evaluation showed our CNN model (see Table 1) yielded favorable to mixed results for the respective pollen categories. The 1 to 7-day predictions of grass pollen had the least optimal scores for HIT (0.125-0.313), CSI (0.111-0.294), FAR (0.167-0.500), and ETS (0.093-0.271). The grass model, however, had a POC score higher than 0.9 in all predictions. An explanation for this score was the abundance of measured and predicted grass pollen counts falling below the threshold value. The model did not accurately forecast the few cases of threshold exceedances. The ratio of the model over-predicting and correctly forecasting pollen count when exceeding the threshold were nearly equal in most cases. This results in significantly higher FAR than exhibited by the other pollen categories. The low CSI score indicates that the model mostly under-predicted during the threshold exceedances of the grass pollen season.

Table 2.1: Statistical evaluation results of the deep convolutional neural network model based on threshold values for the respective pollen categories, their season ranges, and forecasting days ahead.

		Day 1	***Day 2***	***Day 3***	***Day 4***	***Day 5***	***Day 6***	***Day 7***
Tree Pollen Threshold = 50 **Season:** Jan. 19 – May 12	***HIT***	0.667	0.679	0.603	0.667	0.628	0.641	0.615
	CSI	0.634	0.646	0.573	0.634	0.598	0.610	0.585
	FAR	0.071	0.070	0.078	0.071	0.075	0.074	0.077
	ETS	0.318	0.331	0.262	0.318	0.283	0.295	0.272
	POC	0.739	0.748	0.696	0.739	0.713	0.722	0.704
Grass Po	***HIT***	0.313	0.250	0.250	0.125	0.125	0.125	0.125

	CSI	0.294	0.235	0.211	0.111	0.118	0.111	0.118
	FAR	0.167	0.200	0.429	0.500	0.333	0.500	0.333
	ETS	0.283	0.225	0.198	0.102	0.111	0.102	0.111
	POC	0.967	0.964	0.959	0.956	0.959	0.956	0.959
Weed Pollen Threshold = 20 **Season:** Sep. 6 – Nov. 11	***HIT***	0.885	0.904	0.904	0.923	0.923	0.923	0.904
	CSI	0.868	0.887	0.887	0.873	0.873	0.873	0.870
	FAR	0.021	0.021	0.021	0.059	0.059	0.059	0.041
	ETS	0.576	0.619	0.619	0.546	0.546	0.546	0.562
	POC	0.896	0.910	0.910	0.896	0.896	0.896	0.896

The model yielded favorable results of weed pollen forecasts in all evaluation categories (0.885-0.923 HIT, 0.868-0.887 CSI, 0.021- 0.059 FAR, 0.546-0.619 ETS, and 0.896-0.910 POC). Furthermore, our model consistently produced favorable results in predicting 1 to 7 days ahead. The predictions of tree pollen counts were satisfactory, with all scores experiencing minimal variance at predicting 1 to 7-days ahead (0.603-0.667 HIT, 0. 573-0.646 CSI, 0.071- 0.078 FAR, 0.272-0.331 ETS, and 0.696-0.748 POC).

The evaluation of the categorical statistics for total pollen show favorable results of all four threshold values (0.628-0.716 HIT, 0.597-0.685 CSI, 0.036-0.082 FAR, 0.383-0.541 ETS, and 0.773-0.861 POC). Detailed results appear in Table T5**Error! Reference source not found.** in the Appendix Section of the paper.

The results indicate that categorical statistics were not able to sufficiently evaluate the overall accuracy of the model. Therefore, we determined that the IOA and *r* would be appropriate for an alternative evaluation using statistical analyses.

2.3.2 Index of Agreement and the Pearson Correlation Coefficient

IOA measures the degree of model prediction error and whether our model accurately predicted the peaks of pollen concentrations. The *r* measures the linear correlation between the observed and predicted concentrations of our model. We used both methods to evaluate the accuracy of our model at predicting 1 to 7 days ahead for the 2016 time series. We ran the model through 25 iterations to evaluate the consistency and accuracy of our model. See Figure 2.3 for the model performance in IOA and *r* for the 7-day predictions. For one day ahead prediction of tree pollen, the model achieved mean IOA and *r* of 0.88 and 0.85, respectively. The model accuracy ranged from 0.84-0.91 IOA and 0.79-0.88 *r*. The model accuracy decreased slightly to an IOA and r of 0.76 and 0.71, respectively, by forecasting the 7^{th}day. Although, the model under-predicted the concentrations during the peak tree pollen season (see Figure 2.4). One explanation for this finding is that most of the training data had peak pollen concentrations below 2016 concentrations. To address this issue, the model would need to be trained with more consistent pollen and meteorological data. The model would be trained with more samples, further optimizing its predictive capabilities. These instances of the model not being able to capture tree pollen during the fall and winter seasons are likely outliers which previous years did not have the phenomenon.

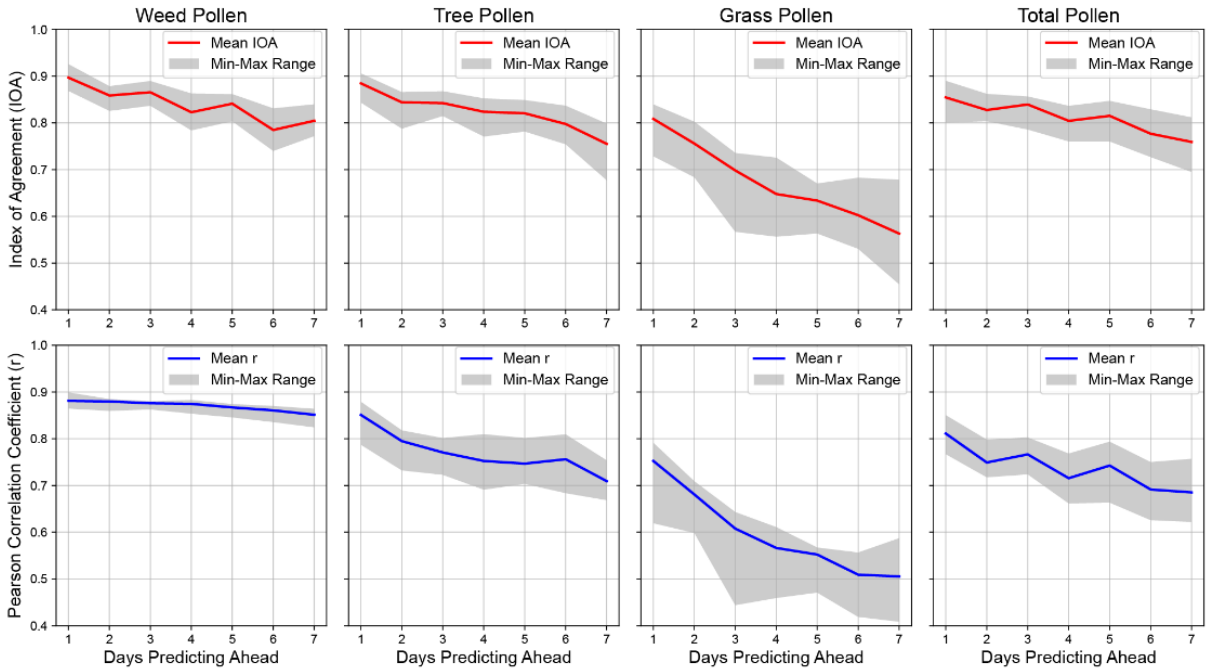

Figure 2.3: Deep convolutional neural network model (CNN) performance in forecasting tree, grass, weed, and total pollen 1-7 days ahead for the entire year of 2016. Red (IOA) and blue (Pearson correlation coefficient) lines represent the mean performance of the model over multiple runs. Shaded areas represent the maximum and minimum performance of the multiple runs for each category.

The model predictions of next-day grass pollen were the least favorable, with a mean IOA score of 0.81 and *r* of 0.75. The range of accuracy for first-day forecasts was also the largest for grass, with an IOA range and *r* range of 0.73-0.84 and 0.62-0.79, respectively. Grass pollen forecast accuracy decreased considerably as forecast time increased. By the 7th day, the accuracy was 0.56 IOA and 0.51 *r*. Weed pollen forecast was most favorable, with a mean IOA and *r* of 0.90 and 0.88, respectively. The model performed more consistently at predicting weed pollen counts. For the 1-day forecasts, IOA score ranged from 0.87-0.93 and *r* score from 0.87-0.90. The accuracy of the model decreased only slightly as forecast time increased. The model over-predicted the weed pollen concentration for 2016 during the peak weed season.

Deep CNN model accuracy substantially declined for the 3-4 day forecasts for the pollen categories. Tree and weed pollen forecast showed minor improvements from 2nd to 3rd-day forecast, suggesting pollen and weather phenomenon of the current day had a delay to the reaction of pollen emissions by plants. A similar phenomenon is observed on an annual temporal scale, where stronger precipitation in one year led to stronger pollen production in the following or second year (Fairley & Batchelder, 1986).

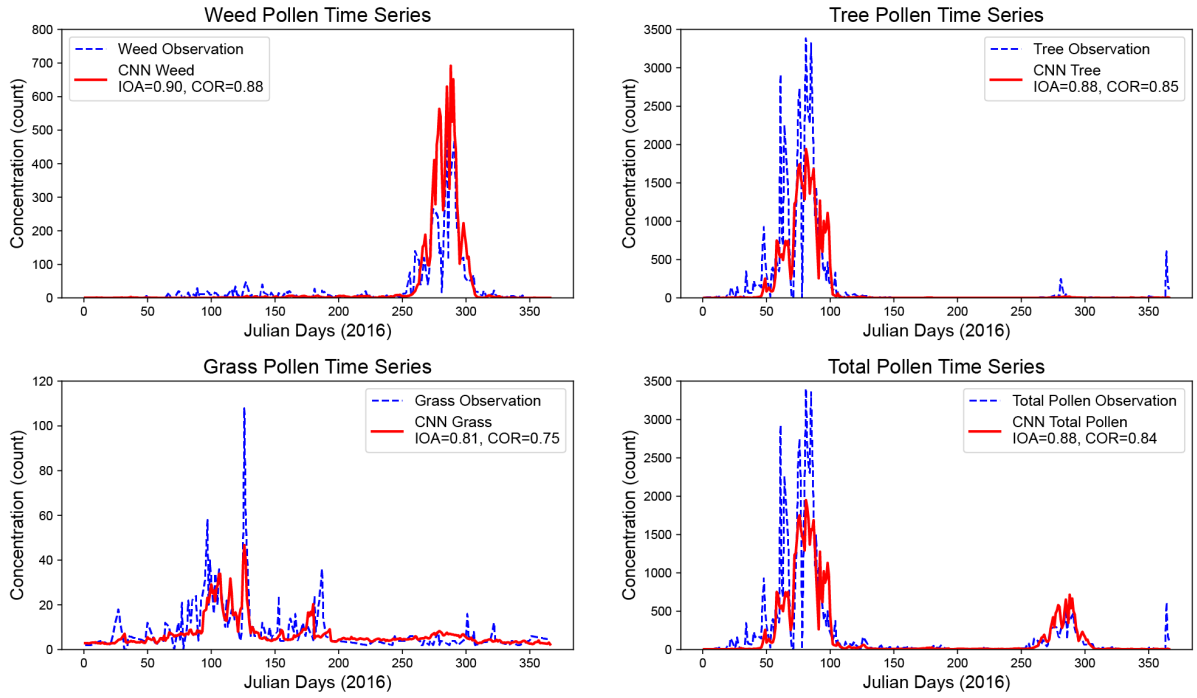

Figure 2.4: Time series of the deep convolutional neural network (CNN) model forecasting of Weed, Tree, Grass, and Total (sum of the tree, grass, and weed pollen) pollen concentrations one day ahead for 2016. Blue indicates the observed pollen concentration, and red indicates forecasted pollen concentration. Respective statistical scores are based on the entire year of 2016.

Results reflected the magnitude of variation in pollen concentrations among the categories from year to year. The average pollen time series of 2009-2015 correlated with that of 2016 by 0.49 for grass, 0.76 for the tree, and 0.84 for weed pollen. Grass pollen strongly varied from year to

year, but those of weed pollen were more consistent, reinforcing the accuracy of the model with regard to the variation among the pollen categories. Hence, the accuracy of the model depends on the stability of the pollen seasons. An evaluation of the model on total pollen 1 to 7-day forecasts for 2016 produced more stable results. This finding can be explained by the inclusion of non- and low-pollen time periods during the year.

The model prediction accuracy of weed pollen counts was slightly less consistent the farther ahead in time the model forecasts. The predictions of counts during the odd days (e.g., 3rd and 5th days) were consistently better than those of the previous days (e.g., 2nd and 4th days). The model results indicate that for 2016, current-day weather had a stronger relationship to the 3rd-day forecast of pollen concentrations than to the 2nd-day forecast. Comparing yearly to seasonal time series for the respective pollen categories, showed on average a 2% decrease in the prediction accuracy for IOA and *r* (see Figure F1 in the Appendix).

The shortcomings of the predictive capability of the current model could be mitigated by implementing certain approaches: (1) Increasing the amount of pollen data concentrations for training; (2) reducing the amount of missing pollen data to improve interpolation accuracy of missing pollen data; (3) increasing temporal resolution of pollen, allowing the evaluation of short-term meteorological effects (e.g., precipitation) on pollen concentrations. Data for individual pollen species were not sufficient for the individual pollen species forecast of the model. Implementation of more layers (deeper CNN) is dependent on the availability of more data. The CNN model's ability to better identify patterns of pollen phenomena was limited to

the availability of data. This phenomenon reflects the limitations of CNN as a forecasting system model discussed by Eslami et al. (2019a).

2.4 Conclusion

Studies have shown that pollen concentrations are related to wind (García-Mozo et al., 2014), temperature (Bartková-Ščevková, 2003), precipitation (Fairley & Batchelder, 1986), and relative humidity (Martin et al., 2010). Despite the meteorological relationships, pollen concentrations vary considerably. This phenomenon reflects the non-linear annual variability of phenological pollen data not always amenable to linear regression modeling (García-Mozo et al., 2014). Despite the pollen concentration variance, our deep CNN was capable of forecasting pollen up to 7 days ahead with sufficient accuracy. On average, our model was able to predict weed pollen concentrations of nearly 0.9 IOA for a 1-day forecast and 0.81 mean IOA for 7^{th}-day forecast. The first-day forecast of tree pollen attained 0.86 IOA and the 7-day forecast 0.78 IOA. Our model produced a satisfactory forecast of grass pollen beyond the next day with an accuracy of at least 0.8 IOA, which dropped considerably to 0.56 IOA in the 7-day forecast. Variation in the forecasting accuracy among the pollen categories related to the variation in their annual pollen seasons. The greater the pollen emissions differed from the previous years, the more difficult it was for the model to accurately forecast pollen concentrations. The model was able to address some of the variability among the pollen concentrations and obtained stronger correlations in the 7-day forecasts than direct correlations between 2016 and previous years.

Our CNN model predicted real-time concentrations of pollen with favorable statistics and generated results within minutes of initiating the model. Therefore, the computational efficiency of the deep CNN algorithm could supplement deterministic and regression models to more accurately and rapidly forecast pollen concentrations - offering a more reliable warning system for populations at high risk of pollen-related allergies.

CHAPTER 3

Application of a Partial Convolutional Neural Network for Estimating Geostationary Aerosol Optical Depth Data[2]

3.1 Introduction

Within East Asia, a significant fraction of the primary and secondary sources of aerosols originate from dust transport (Wang et al., 2008), biomass burning (Uranishi et al., 2019), and emissions from anthropogenic sources (Kang et al., 2019) and natural sources (Zhang et al., 2016). A large fraction of secondary aerosols is produced from the condensation of precursor gases or nucleation via cloud processes (Ervens et al., 2011). Particulate matter (PM), a major air pollutant worldwide (Koulouri et al., 2008; W. Li et al., 2014; Mukherjee & Agrawal, 2017), comes in two aerodynamic diameters of fine particles: less than 10μm (PM_{10}) and less than 2.5μm ($PM_{2.5}$) (USEPA, 2020). As PM is associated with respiratory and cardiovascular diseases and mortalities (Brunekreef & Holgate, 2002), acquiring accurate estimates of PM is important to assess its impact on human health. Estimates have shown that globally over two million deaths resulting from damage to the respiratory system per year are associated with PM pollution (Kim et al., 2015; Shah et al., 2013). A common product of satellites for estimating PM levels (Ghahremanloo et al., 2021a) is the aerosol optical depth (AOD). Although researchers have devoted a great deal of effort to developing systems that improve PM

[2] Published: Lops, Y., Pouyaei, A., Choi, Y., Jung, J., Salman, A. K., & Sayeed, A. (2021). Application of a Partial Convolutional Neural Network for Estimating Geostationary Aerosol Optical Depth Data. *Geophysical Research Letters*, *48*(15), e2021GL093096.

forecasting, they have been limited by the quality and amount of data (Ghahremanloo et al., 2021b) and the performance of chemical transport models (Park et al., 2011; Pouyaei et al., 2020). To address the limitations of models and improve the accuracy of models for atmospheric chemistry and weather forecasting, they have incorporated both in-situ measurements and remote sensing products into data assimilation techniques (Bocquet et al., 2015; Jung et al., 2019), which are limited to the availability and quality of in-situ and remote sensing data products.

While ground-based remote sensing measurements from the Aerosol Robotic Network (AERONET) provide detailed point observations (Holben et al., 1998; 2001), the system lacks the spatial resolution of remote sensing capabilities. Using reflected and emitted radiation at a distance, the process of remote sensing measures the features and characteristics of an area. In the past several decades, studies have reported significant advances in measuring and assessing surface features (Mulla, 2013) as well as forecasting and managing air quality (Mhawish et al., 2018). Satellite remote sensing instruments have contributed essential data pertaining to the global distribution (Gupta et al., 2006; Martin 2008; Lee et al., 2016), evolution (Zhang et al., 2012), and transport (Kim et al., 2017; Wu et al., 2018) of atmospheric pollutants. A limitation of remote sensing is the impact of cloud cover contamination, false reflectance, and significant biases within the data (Choi et al., 2016; Mhawish et al., 2018), which must then be discarded. It can also cause sensor errors that corrupt or lead to failed measurements of the data (Rulloni et al., 2012; Shen et al., 2015). These limitations of remote sensing can compromise both the quality and the scope of data, reducing its ability to estimate and forecast pollutant

concentrations, such as not fully describing spatio-temporal variations of aerosols in the atmosphere (Park et al., 2011).

Studies have applied various methods for imputing missing data with remote sensing such as local linear histogram-matching (Scaramuzza & Barsi 2005), geostatistical approaches such as Kriging (Zhang et al., 2007; Yu et al., 2011), linear regression models (Rulloni et al., 2012), inpainting algorithms (Bertalmio et al., 2000; Bugeau et al., 2010), and deep learning algorithms (Li et al., 2017; Zhang et al., 2018). The latter (LeCun et al., 2015) show significant promise in addressing the limitations of missing data by modeling high-level abstractions within datasets (Bengio, 2009; Deng & Yu, 2014). Among deep learning algorithms, convolutional neural networks (CNNs) (Krizhevsky et al., 2017) are among the most successful and widely used approaches (LeCun & Bengio, 1995; Schmidhuber, 2015) in forecasting (Eslami et al., 2019a; Lops et al., 2019; Sayeed et al., 2020; Zhang et al., 2015), classification (Anthimopoulos et al., 2016; Q. Li et al., 2014), speech recognition (Mikolov et al., 2011), and imputation (Li et al., 2017; Zhang et al., 2018). Many models and methods still have difficulties in imputing remote sensing data with a significant percentage of missing data or large gaps within datasets (Gerber et al., 2018; Srindhuna & Baburaj, 2020; Zhang et al., 2018). To impute remote sensing data with a large percentage of missing data, this paper uses a modified convolutional layer within a CNN model called partial convolution (Liu et al., 2018). Although convolution models have been applied to imputing missing remote sensing data (Li et al., 2017; Zhang et al., 2018), they require a temporal dimension within the dataset or imputed images with a low frequency of missing data. The partial convolution (Liu et al., 2018) approach aims to address the limitations

by accurately imputing images with a significant number of missing data and spatial distances. Another benefit of the partial convolutional neural network (partial CNN) is its use of numerical model data, despite the low correlation to remote sensing data, as the basis for training the deep neural network without significant impact on the accuracy of the model.

3.2 Methods

3.2.1 Data Preparation

The Geostationary Ocean Color Imager (GOCI) is a key instrument aboard the Korean Communication, Ocean and Meteorological Satellite 1 (COMS-1). Obtaining the data in eight bands of visible and near-infrared spectra, the GOCI instrument measures various atmospheric and surface features (e.g., turbidity, chlorophyll, fluorescence signals, and aerosol optical thickness) over East Asia (Choi et al., 2012). The system consists of a geostationary satellite that provides eight images per day at a temporal resolution of one hour, covering a 2,500km × 2,500km area at a spatial resolution of 500m (Ryu e. al., 2012). The GOCI Yonsei aerosol retrieval (YAER) version 2 (V2) product algorithm retrieves 550 nm AOD at a 6 km × 6 km spatial resolution (Choi et al., 2018, 2019). In addition to following a data filtering process to exclude GOCI images with no available data from 2014 to 2018, we have filtered each image to exclude isolated pixel clusters of three or fewer pixels within a defined filter grid to ensure the exclusion of outliers. We acquired missing data masks for the training process of the partial CNN model and restricted the size of GOCI images to 384 × 384 pixels to fit the CMAQ AOD image dimensions used for training the model.

Air quality modeling is commonly carried out by using three-dimensional Eulerian chemical transport models. We implemented the United States Environmental Protection Agency (U.S. EPA) Community Multiscale Air Quality (CMAQ v5.2) model (Byun & Schere, 2006) to estimate and predict the aerosol optical depth (AOD) of a region with no available satellite data. We used the configuration and data of the CMAQ model from Jung et al. (2019). We used CMAQ model runs and the IMPROVE (Interagency Monitoring of Protected Visual Environments) reconstruction method to compute the AOD at 550nm wavelength (Pitchford et al., 2007) over East Asia. To ensure a sufficient number of training samples for the partial CNN model, we applied the GOCI missing data masks over the CMAQ AOD images through a randomization process with image augmentation. The original CMAQ AOD image pixel size was 128 × 174, from which we extracted several images with a 128 × 128 resolution and upscaled them to 384 × 384 pixels using bi-cubic interpolation during image augmentation of the training and validation phase. As increasing the spatial resolution of the domain within the CMAQ model significantly increased the processing time, we performed an interpolation process. Since bi-cubic interpolation smooths concentration variations, however, it had a minimal impact on the overall quality of the image. The augmentation phase ensures that the partial convolution is robust at imputing missing data in various images and that the system does not produce extreme variations within the missing data imputation process. The original domains of GOCI and CMAQ are available in Figure F2 in the supplement to this paper.

We acquired topology data from the GTOPO30 digital elevation model (DEM), developed by the United States Geological Survey (USGS, 2020). We mapped the data, with a spatial

resolution of 1km×1km and a coordinate grid of the elevation data, to respective CMAQ and GOCI coordinate grids. Having coordinate grids ensures the adaptability of the model to various regions and remote sensing input data. Incorporating topology data and global normalization within the partial CNN model have shown to slightly improve the performance of the model (see Figure F3 in the Appendix); a similar case has been shown within other CNN models (Vandal et al., 2017). For the purposes of model comparisons, we focused on the partial CNN model without topology datasets.

3.2.2 Neural Network Structure

We implemented a deep-learning technique called partial convolution (Liu et al., 2018), which utilizes a deep convolutional neural network (deep CNN) (Krizhevsky et al., 2017) as a U-net architecture (Ronneberger et al., 2015) that replaces conventional convolutional layers with partial convolution layers. CNNs process data by convolving the data at each layer, assigning weights and biases at various aspects within the data, and differentiating between them. CNN models reduce data into a more manageable form for processing (without losing important features) and extract high-level features of the input data through the use of kernels during the convolving phase (Lawrence et al., 1997). In addition, we integrated elevation data within a color channel (layer of the primary color of a digital image) of the masked image to spatially impute missing AOD images. Ill-posed problems related to imaging and convolutional processes in inverse problems can be alleviated with the addition of proxy datasets (Jin et al.,

2017; Vandal et al., 2017). We implemented the algorithm in the Keras and TensorFlow environments (Chollet, 2015; Chollet, 2018; Abadi et al., 2016).

We trained the partial convolution model on CMAQ AOD images of model runs for 2014-2017 and extracted missing data masks from the GOCI images. AOD images undergo a normalization process based on the defined minimum and maximum values (0 and 12, respectively) for the entire dataset. Ensuring proper distribution or regularization of the data improves model performance by increasing the cases of rare outliers within the dataset (Sola & Sevilla, 1997; Eslami et al., 2019b; Sayeed et al., 2021). To further increase the training samples and enhance the robustness of the model, we implemented image augmentation of the AOD images (see Table T6). Image augmentation transforms data into modified samples for training, which improves the robustness of the deep learning model (Fawzi et al., 2016). In addition, we randomly selected the GOCI missing data masks, which underwent simple image augmentation before we applied them to a CMAQ AOD image during the augmentation phase. The partial convolution model structure consists of 16 total layers comprised of one input layer, seven partial convolution encoding layers, seven partial convolution decoding layers, and an output layer. Each encoding layer contained a pooling layer, a partial convolution layer with batch normalization, and Rectified Linear Unit (ReLU) activation function. Each decoding layer was comprised of an upsampling layer concatenated with a respective layer in the encoding layer. The upsampled layer was then processed through a partial convolution layer with batch normalization and Leaky ReLU activation. Leaky ReLU prevents information loss and allows

the negative parts of features within the convolution to activate (Maas et al., 2013; Zhang et al., 2017). The final decoding layer was then processed through a regular convolutional layer with Sigmoid activation, which provided the final output image (see Figure 3.1) for the schematic of the system. The loss function used for the model training was the same as that in Liu et al. (2018):

$$L_{total} = L_{valid} + 6L_{hole} + 0.05L_{perceptual} + 120\left(L_{style_{out}} + L_{style_{comp}}\right) + 0.1L_{tv} \quad (11)$$

comprised of pixel hole loss (L_{hole}), pixel valid loss (L_{valid}), perceptual loss ($L_{perceptual}$), raw style output ($L_{style_{out}}$), composited output ($L_{style_{comp}}$), and total variation loss (L_{tv}).

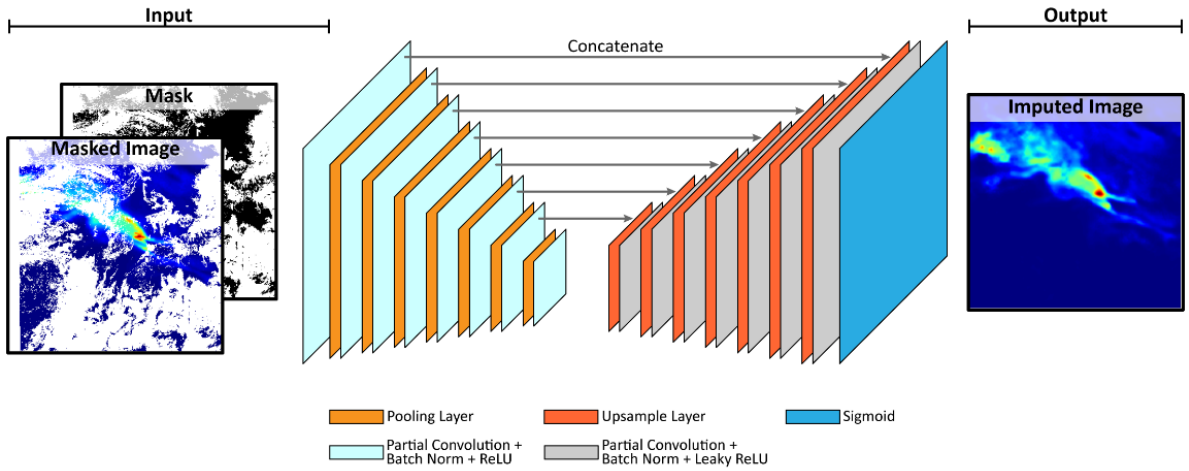

Figure 3.1: Schematic of the partial CNN model for imputing missing remote sensing data.

The training of the model was comprised of four phases: i) an initialization phase of 5 epochs; ii) the first training phase with batch normalization for 15 epochs; iii) the second training phase without batch normalization for 10 epochs, and iv) the final training phase with a reduced initialized learning rate for 10 epochs. The initialization phase increases the training speed of the model. Over several training runs, we observed that the partial CNN model, when

initialized, did not improve its loss score over several epochs, increasing the training time with a minimal reduction in the loss score. When the model was trained for several epochs and reloaded for another training phase, we observed significant improvement in the loss score. Thus, we allowed the model to train for 5 epochs as an initialization process. Batch normalization performed during the first training phase improved the speed and the accuracy of the model training (Bjorck et al., 2018). Checkpoints have been enabled by saving the model with the least validation error from each training phase.

3.2.3 Model Comparisons

We compared the partial convolution model to the Kriging method (Cressie, 1990), which is based on Gaussian process regression and the regionalized variable theory. Kriging assumes that spatial variation in a phenomenon is statistically homogeneous throughout the surface based on available data from nearby locations (Cressie, 1990). The implementation of Spatiotemporal Kriging and CoKriging (Kyriakidis & Journel, 1999) were not utilized due to significantly increased processing time. In addition, we utilized CMAQ model AOD and topology data for CoKriging for the first 3 months within the 2018 GOCI time period. Results indicated a significant reduction in imputation accuracy of the CoKriging model (see Figure F4 in the Appendix). This limitation is significantly impacted by the low correlation (0.52) of CMAQ AOD to the respective GOCI AOD values.

Other models we use to compare our partial CNN model is the Inverse Distance Weighting (IDW) (Lu & Wong, 2008), Radial Basis Function (RBF) (Orr, 1996), K-Nearest Neighbor

(KNN) (Keller et al., 1985), Navier Stokes (NS) method (Bertalmio et al., 2001), and Fast Marching Method (FMM) (Telea, 2004). The inverse distance weighted (IDW) interpolation method assumes that pixels close to each other are more likely to have similar values, and the local influence of available points on predictions falls off with distance (Fisher et al., 1993). RBF is a supervised machine learning (ML) that utilizes linear functions for nonlinear optimizations (Orr, 1996). KNN is also a supervised ML algorithm that focuses on K-nearest patterns in data space (Kramer, 2013). NS works by utilizing fluid dynamics techniques and partial derivatives to estimate the missing data points (Bertalmio et al., 2001). FMM works by propagating an image smoothness as a weighted average along the image gradient by gradually filling missing data points with known data points (Telea, 2004).

3.2.4 Evaluation

We evaluated various models based on various datasets and methods. As we trained the partial convolution model on CMAQ data, we did not conduct an evaluation of the CMAQ data. We evaluated the GOCI images by processing the hourly GOCI images into daily mean GOCI images. We observed a significant number of cases with no AOD values during the first and last hours (due to high solar zenith angle near sunrise and sunset (Choi et al., 2016)) within the eight-hour measurement period; thus, we excluded these images for evaluation purposes. We also excluded cases of one hourly GOCI AOD measurement per day from the evaluation. Thus, the daily mean GOCI AOD was comprised of 2-6 images per day. The processing time of imputing 1581 GOCI images for each of the models is provided in Table T7 (see Appendix).

We evaluated and validated both systems based on the daily mean GOCI images compared to their respective hourly masked GOCI images to the available daily mean mask within the GOCI image. Applying an estimated pixel distance function based on the distance to the nearest available data point, we evaluated the imputation bias of the models to the distance of the nearest data variable (see Figure F5 in the Appendix). We conducted a secondary evaluation based on the imputed GOCI hourly images to the Aerosol Robotic Network (AERONET) data for China (Beijing), South Korea (Gangneung, Gwangju, and Seoul), and Japan (Fukuoka, Niigata, and Osaka). We evaluated the model performances in cases in which no GOCI data were available at the respective location and measurements at the AERONET station were available for 2018.

3.3 Results and Discussion

3.3.1 Imputation of the GOCI Images

The results of the GOCI comparison are based on the daily mean GOCI AOD mask and the difference to the hourly masks for the respective day. We compiled the image data into a one-dimensional format for each hourly measurement from the GOCI for 2018 and evaluated each hourly dataset based on statistical evaluation methods. The dataset was comprised of measurements of the index of agreement (IOA) (Willmott et al., 1985), the correlation coefficient (r) (Benesty et al., 2009), the root mean square error, and the mean absolute error (Chai & Draxler, 2014). In addition, we evaluated the IOA performance of each model based on the percentage of missing data between the daily mean mask and the hourly mask within a

GOCI image. The purpose of this evaluation was to identify how each model performed based on the relative amount of missing data within an image. The GOCI dataset for 2018 comprised of images containing 35.0-99.7% of missing data, with ~45% of the images having 80% or more missing pixels within an image (see Figure F6 in Appendix).

Statistical results of the various models and algorithms for the 2018 GOCI cases are shown in Figure 3.2. For all 2018 GOCI images, the partial CNN model achieved average performance in IOA and *r* while achieving the second-lowest MAE and RMSE of all the models. IDW, RBF, and NS models have similar performance in statistical accuracy. Based on the % of missing NaN values within the evaluated pixels, partial CNN begins to outperform when the percent of missing data begins to exceed 60% (which accounts for ~90% of GOCI images for 2018). The merits and demerits of the models and algorithms based on the results in imputing GOCI AOD are provided in Table T8 (see Appendix).

The partial CNN model has been observed to modify the original GOCI AOD pixel values but still shows a correlation of 0.970 and an IOA of 0.983 with a trend of reducing the original AOD values. Upon further analysis, we found that partial CNN has reduced the bias GOCI AOD in relation to AERONET estimated AOD. Based on the available AERONET measurements, partial CNN has reduced the GOCI values closer to the AERONET estimated AOD (see Figure F7 in Appendix). Line plots indicating the impact of partial CNN's

modification of GOCI AOD are shown in Figure F8 and the statistical comparisons are shown in Table T9 (see appendix).

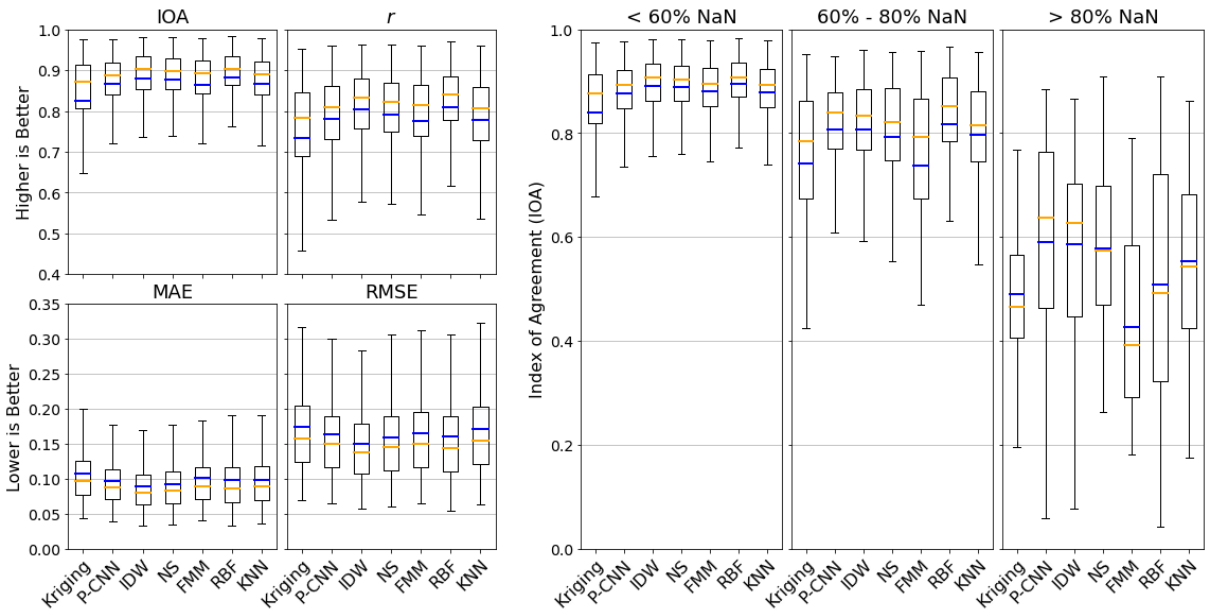

Figure 3.2: Statistical comparison of imputed missing GOCI data from the hourly mask to the daily mean mask. The left figures indicate the statistical performances of the partial CNN model (P-CNN), Kriging (Kriging), Inverse Distance Weighted (IDW), Navier Stokes (NS), Fast Marching Method (FMM), Radial Basis Function (RBF), and K-Nearest Neighbor (KNN) methods. Evaluations are based on the index of agreement (IOA), the correlation coefficient (*r*), the mean absolute error (MAE), and the root mean squared error (RMSE). The right figure indicates the IOA performance of the models, based on the percentage of missing data split into three categories. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range. The blue and yellow horizontal lines represent the mean and median of the dataset, respectively.

We identified three instances in which the partial CNN model failed to properly impute missing data, which can be explained by the distribution of data within an image. If the available data were clustered around the side of an image, the model did not impute missing data accurately on the opposite end of the image. This issue is consistent within several models that we have trained and tested, indicating a limitation of the partial CNN model. These cases are infrequent

within the training and validation dataset; hence, in these instances, the partial CNN model is undertrained. Once we added sample data to the respective three AOD images in a different region within the image, the large biases of the partial CNN model decreased significantly (see Figure F6 in the Appendix). Although these sample cases are maintained within the general model performance presented in Figure 3.2, we excluded them from the results of the pixel distance evaluation (Figure 3.3).

3.3.2 Pixel Distance Evaluation

We evaluated the GOCI imputation comparison by applying a pixel distance mask relative to the closest available data point within the GOCI image. To estimate the pixel distance from the nearest available data point, we applied a filter over multiple iterations to each hourly mask. We calculated the bias of the imputation methods for each pixel, categorized them based on the distance value of the respective pixels from the pixel distance mask (see Figure F6 in the Appendix), and sorted the pixels of each imputed AOD within the evaluation mask at 4-pixel intervals representing ~24 km distances. At each interval, we plotted the difference between the bias of the imputed AOD values to that of the GOCI AOD value in a boxplot. The series of boxplots at each pixel interval provides information on the bias range of each model as the distance increases. We evaluate partial CNN, IDW, RBF, and NS models due to these models being the best performant models within the statistical comparisons. The pixel distance plot (see Figure 3.3) shows that the models performed similarly at distances below 16 pixels (~96 km distance). Compared to the partial CNN model, the other models had a slightly smaller bias-

variance within the nearest pixels to an available data point. Both systems diverged in their bias beyond a 20-pixel distance, with the IDW, NS, and RBF models trending towards a more positive mean bias and a slightly larger variance than the partial CNN model with a negative mean bias and a slightly lower variance. The partial CNN model achieved the lowest bias-variance at a 52-pixel distance (~312 km distance) with a relatively high negative mean bias. Beyond 52-pixel distance, the bias variances of the partial CNN and IDW models are stable, with the partial CNN maintaining a lower overall bias-variance than the other models. We observed a slight trend towards a significantly higher mean bias in the NS and RBF models throughout distance ranges beyond 100 pixels. These results do not include the three cases of the partial CNN (and the respective model outputs) image imputation because of their extreme outliers. The graph of pixel distances in Figure F10 in the Appendix displays partial CNN, Kriging, FMM, and KNN models. As a reference, distributions of bias counts comparing Partial CNN and Kriging, with and without the outliers, appear in Figure F11 in the Appendix.

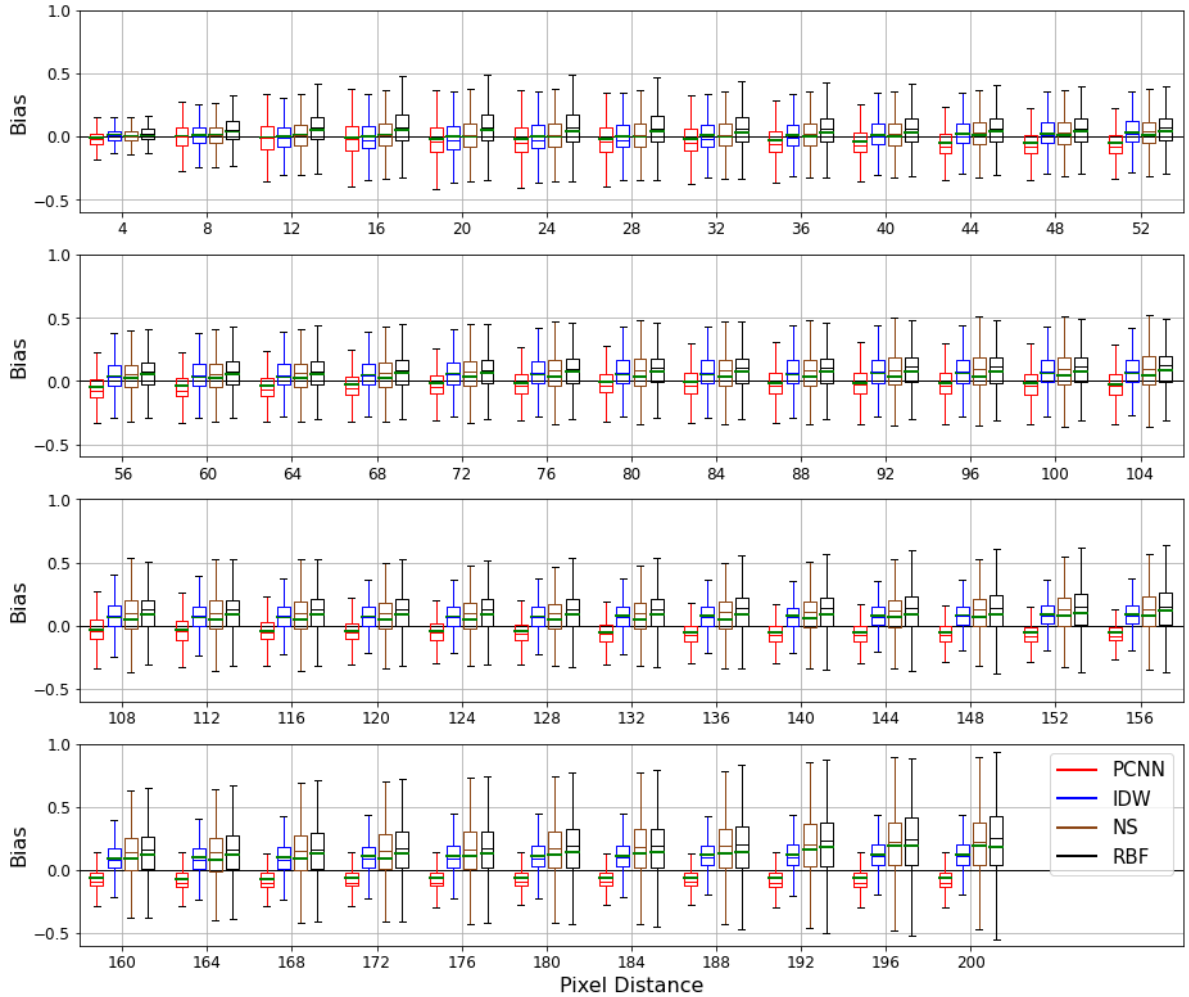

Figure 3.3: Comparison of the bias-variance in the pixel distances of the Partial CNN (PCNN), Inverse Distance Weighting (IDW), Navier Stokes (NS), and Radial Basis Function (RBF) models. The distances are split into four sections at 52-pixel intervals. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The horizontal lines represent the mean bias of PCNN (red), IDW (blue), NS (orange), RBF (black) models, and the median bias (green for all models) of the imputing missing GOCI data. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range.

The negative bias of the partial CNN model can be attributed to the general reduction on the AOD value of the original GOCI image by the partial CNN model. In addition, partial CNN under-predicts at distances further away from the nearest available data point. The positive bias of the other models is explained by the smooth transition between available data points and missing data points at extended distances (see Figure F12 in Appendix). The study by Rossi et

al. (1994) found that Kriging generally experiences a slight positive bias. Seasonal image comparisons exemplify the smoothing effect of the Kriging, IDW, and RBF models. Partial CNN also imputes higher AOD values at medium distances but has cases of under-prediction at long distances and over ocean bodies (see Figure F13 in Appendix).

3.3.3 AERONET

When GOCI data were not available at a respective location, the AERONET results (see Table T10 in Appendix) were based on the available hourly measurements of the AERONET measurements. For the evaluation of AERONET data, the partial CNN performed above average based on all the models. At the Beijing AERONET station, all imputation models experienced high MAE and RMSE, with the NS and RBF models achieving the least biases. Plotting the partial CNN, Kriging, IDW, and RBF imputed GOCI AOD over AERONET AOD, we observed that the partial CNN model frequently estimated AOD values with high positive and negative deviation (see Figure F14 in Appendix). Within the Beijing station, cases of partial CNN over-predicting and Kriging under-predicting have been observed (see Figure F15 in Appendix). Partial CNN's high AOD estimation can be attributed to the combination of the initial GOCI values and mask region within an image due to the convolutional process. Modifying the initial GOCI AOD values by factors of 4, 2, and ½ provide different patterns in the imputation of the partial CNN model (see Figure F16 in Appendix). To further evaluate the impact of partial CNN's imputation based on input image values, we replaced the initial normalized GOCI AOD values with consistent numbers. Results indicate the model provided

similar patterns in the output image, but the individual values have different intensities (see Figure F17 in Appendix). If there are low initial values within a mask, the partial CNN model imputed values higher than expected, overcompensating its imputation in the surrounding region with cases of significant missing data within an image. Partial CNN's over-prediction is based on the pattern it produces in the original masked image and the magnitude of the values within the mask. The statistical bias comparisons of all the models are shown in Table T10 in the Appendix section.

3.4 Conclusion

This research demonstrated the capability of the partial convolutional neural network in the application of imputing missing remote sensing data. All models achieved acceptable results when imputing images comprised of 60% or fewer missing data. Only partial CNN, IDW, NS, and RBF maintained mean IOAs above 0.8 in cases of missing data ranging from 60-80%. When missing data exceeded 80% of the image, the mean IOA of the partial CNN model remained at ~0.6, with a higher IOA score than the other models. FMM has performed the worst, with missing data exceeding 80% within a dataset. From evaluations of the performance based on distance, all models showed similar biases within areas closest to the available data and diverged once a pixel distance of 32 (corresponding to a distance of ~192km) was exceeded. The majority of the models have gravitated to over-predict with higher variance, while the partial CNN model has gravitated to under-predict with lower variance in terms of the bias distribution at each pixel distance. The results of the performances of the models with data from

seven AERONET stations (Beijing, Fukuoka, Gangneung, Gwangju, Niigata, Osaka, and Seoul) in East Asia indicated that the partial CNN model outperformed the Kriging, FMM, RBF, and KNN models for six stations and outperforming all models or only the Niigata station. All models experienced high biases within the Beijing region, occasionally significantly over- or under-predicting AOD, causing a significant deviation from the measured AODs.

The partial CNN model experienced issues with imputing entire images for rare cases of minimal available data with abnormally tight clustering. Data clustering also impacted the other model imputation by outputting AOD values close to zero (e.g., Kriging); the values, however, were not as pronounced as those AOD output bias of the partial CNN model. Such rare cases can be avoided by ensuring that the proper distribution of available data throughout an image regardless of how few data are available. Overall, the imputation performance of the partial CNN model exceeds those of the other models in cases of a high percentage of missing pixels and at long distances from known pixels within an image. Models such as IDW, RBF, and NS performed well with cases of few missing data and imputing regions close to known pixels. Once trained, the partial CNN model imputes large remote sensing datasets in significantly less processing time than the other models. For the accurate imputation of such datasets, the implementation of partial CNN allows the accurate imputation of remote sensing data with large regions of missing data, which will benefit future researchers that conduct other studies such as data assimilation for numerical models and analysis of the impact of air quality on human health. In addition, the partial CNN system can be expanded to impute variables between multiple observation stations over large distances to alleviate the need to use model and

reanalysis processes, saving processing time and resources. Future work involves modifying the partial CNN model to process both temporal and spatial data within remote sensing datasets to further enhance the model's performance.

CHAPTER 4

Spatio-Temporal Estimation of TROPOMI NO2 Column with Depthwise Partial Convolutional Neural Network[3]

4.1 Introduction

Nitrogen oxides (NO_X=NO+NO_2) are one of the major criteria pollutants (Zhang et al., 2016) with significant impact from human activity (Ghahremanloo et al., 2021b). Sources of NO_X come from anthropogenic and natural sources such as the combustion of fossil fuels (Choi et al., 2009), burning of biomass (van der Werf et al., 2006), soil microbial activity (Yienger and Levy, 1995), and lightning (Choi et al., 2009). Nitrogen dioxide (NO_2) has been associated with adverse negative health impacts and effects such as cardiovascular diseases (Liu et al., 2015) and respiratory-related ailments (Zhu et al., 2019).

The concept of remote sensing is to measure the features and characteristics of an area with the utilization of reflected and emitted radiation at a distance. Satellite remote sensing instruments have contributed essential data pertaining to the global distribution (Gupta et al., 2006; Martin 2008; Lee et al., 2016), evolution (Zhang et al., 2012), and transport (Kim et al., 2017; Pouyaei et al., 2020, 2021; Wu et al., 2018) of atmospheric pollutants. Unfortunately, remote sensing has limitations such as low spatial and temporal resolutions (Ghahremanloo et al., 2021c) and measurements issues caused by the impact of cloud cover contamination, false reflectance,

[3] Submitted to Neural Computing and Applications journal

significant biases within the data (Choi et al., 2016; Mhawish et al., 2018). Furthermore, the system can also experience sensor errors that corrupt or lead to failed measurements of the data (Rulloni et al., 2012; Shen et al., 2015).

Several methods have been applied to non-temporally impute missing data within remote sensing images, such as geostatistical approaches (Zhang et al., 2007; Yu et al., 2011), linear regression models (Rulloni et al., 2012), inpainting algorithms (Bertalmio et al., 2000; Bugeau et al., 2010), and deep learning algorithms (Ghahremanloo et al., 2021c; Li et al., 2017; Lops et al., 2021). Deep learning algorithms (LeCun et al., 2015) have shown significant promise in addressing the limitations of missing data by modeling high-level abstractions within datasets (Bengio, 2009; Deng & Yu, 2014). Among various deep learning algorithms, convolutional neural networks (CNNs) (Krizhevsky et al., 2017) have been among the most successful and widely used approaches (LeCun & Bengio, 1995; Schmidhuber, 2015) for various purposes such as forecasting (Lops et al., 2019; Sayeed et al., 2020, 2021; Yeo et al., 2021; Zhang et al., 2015), classification (Anthimopoulos et al., 2016; Lee et al., 2017; Li et al., 2014), speech recognition (Mikolov et al., 2011; Park et al., 2017), and imputation (Li et al., 2017; Zhang et al., 2018; Lops et al., 2021). Many models and methods still have difficulties in imputing remote sensing data with a significant percentage of missing data or large gaps within datasets (Gerber et al., 2018; Srindhuna & Baburaj, 2020; Zhang et al., 2018). Advanced methods of imputation utilize temporal dimensionality to enhance the accuracy of the imputation process (Bae et al., 2018; Carvalho et al., 2016; Gerber et al., 2018).

Although convolution models have been applied to imputing missing remote sensing data with a temporal dimension within the dataset (Zhang et al., 2018), they require data with a low frequency of missing pixels. This paper expands the application of the Partial Convolutional Neural Network (PCNN) (Liu et al., 2018) in imputing missing remote sensing data (Lops et al., 2021) by the addition of temporal dimensionality within the model. The PCNN model has been shown to perform well in imputing images with a significant amount of missing data and spatial distances, which will be further enhanced with the addition of the temporal dimension of the model input. The temporal component of the PCNN is applied through the implementation of depthwise convolutions (Chollet, 2017), in which the convolution process is independently performed for each channel. The Depthwise Partial Convolutional Neural Network (DW-PCNN) aims to address the limitations of the regular PCNN model (Lops et al., 2021) by incorporating the temporal component for imputation, improving the sharpness of the imputed image, and enhancing the accuracy over the regular PCNN.

4.2 Methods

4.2.1 Data Preparation

The TROPOspheric Monitoring Instrument (TROPOMI) is a key instrument aboard the Copernicus Sentinel-5 Precursor (S5P) satellite. The instrument obtains key atmospheric constituents such as ozone (O_3), NO_2, formaldehyde (CH2O), and aerosol data through ten spectra bands of ultraviolet (UV), visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) (Veefkind et al., 2012). The system is a near-polar, sun-synchronous orbit that provides daily global coverage at high spatial resolution (7 km×3.5 km at nadir) (Guanter et al., 2015;

Ludewig, 2021). We performed an initial data filtering process to exclude pixels failing the initial Quality Assurance (QA) value threshold (i.e., QA<0.5) that present an error flag or solar zenith angle exceeding 70°, cloud cover, and air mass factor below 0.1 (Vigouroux et al., 2020). In addition, we have filtered each image to exclude isolated pixel clusters of four or fewer pixels within a defined filter grid to ensure the exclusion of outliers.

We implemented the United States Environmental Protection Agency (U.S. EPA) Community Multiscale Air Quality (CMAQ v5.2) model (Byun & Schere, 2006). The model's domain has a 12 km grid horizontal spacing with 27 vertical layers reaching 100 hPa to estimate and predict the tropospheric column density of NO_2 ($TCDNO_2$) over the Contiguous United States (CONUS). The system uses CB6 and AERO6 chemical mechanisms for the gas-phase and aerosol chemical processes. The model received the 2017 U.S. EPA National Emission Inventory (NEI) (Eyth et al., 2016; Eyth and Vukovich, 2016) with parameterized lightning-induced emission, biogenic emission computed by using the Biogenic Emission Inventory System (BEIS), and biomass burning emission adopted from the Fire Inventory from National Center for Atmospheric Research (FINN) version 1.5 (Wiedinmyer et al., 2014, 2011, 2006). The CMAQ model received simulated meteorological variables from the Weather Research and Forecasting (WRF) model version 4.0 with the National Centers for Environmental Prediction (NCEP). For initial and boundary conditions, the North Americans Regional Reanalysis (NARR) data was utilized. Furthermore, we incorporated indirect soil moisture and the temperature nudging technique (Pleim and Gilliam, 2009; Pleim and Xiu, 2003), as well as a Four-Dimensional Data Assimilation (FDDA) option for the temperature, the water vapor

mixing ratio, and wind components (Hogrefe et al., 2015) to enhance the model performance in simulating meteorological fields. The simulations were performed for the months February-June for 2019 and 2020.

The simulated NO_2 column by CMAQ acts as the basis for preparing training data for the PCNN model. This process is conducted to ensure the data for training contains images without any missing data. The original CMAQ output image pixel size was 299×459, from which we extracted ten images with a 256×256 resolution. We applied several image augmentation processes to ensure enough training samples for the partial CNN model were available. The first image augmentation step is applying a random noise function with Gaussian smoothing on CMAQ images to replicate the pixel variations observed in TROPOMI images. The second phase of the process is the application of basic image augmentation to the updated CMAQ image through the random selection of rotation, flipping, or combination. The final augmentation phase involved the random selection, rotation, flipping, and application of TROPOMI masks to the updated CMAQ images. The augmentation phase ensures that the partial convolution is robust at imputing missing data in various images and that the system does not produce extreme variations within the missing data imputation process.

The domains of TROPOMI and CMAQ within the CONUS are shown in Figure 4.1.

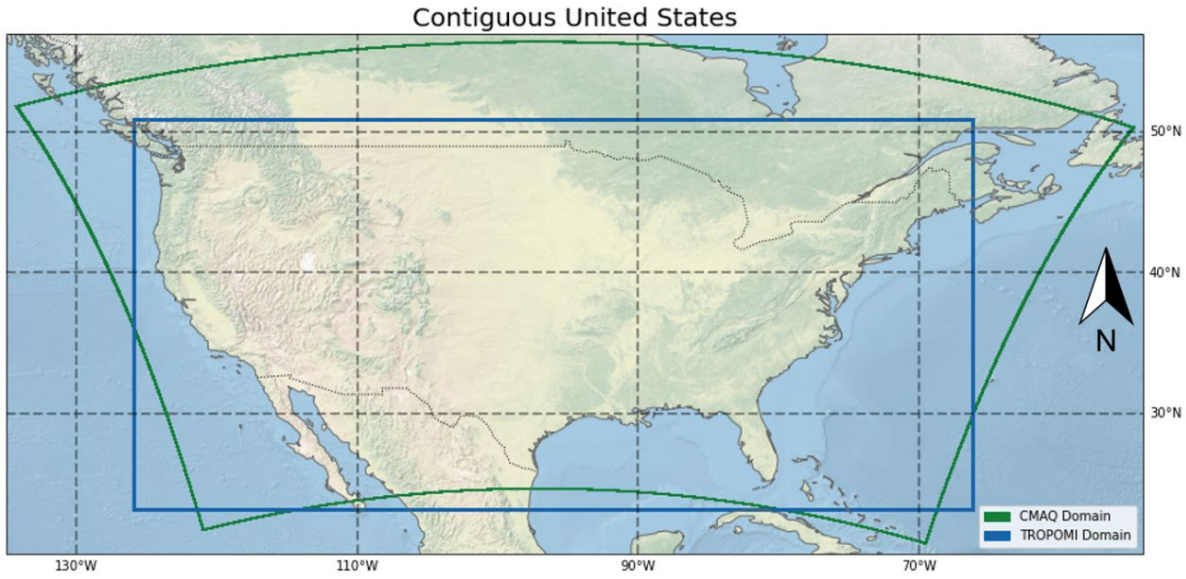

Figure 4.1: Map of the Contiguous United States (CONUS) with domains of the CMAQ model (Green) and TROPOMI measurement (Blue) datasets used for the study.

4.2.2 DW-PCNN Structure

We utilized a deep convolutional neural network (deep CNN) (Krizhevsky et al., 2017) U-net architecture (Ronneberger et al., 2015) that replaces conventional convolutional layers with partial convolution layers (Liu et al., 2018). CNNs process data by convolving the data at each layer over multiple image channels, assigning weights and biases at various aspects within the data, and differentiating between them. The benefits of CNN models are their capability to reduce data into a more manageable form for processing (without losing key features) and extract high-level features of the input data through the use of kernels during the convolving phase (Lawrence et al., 1997).

Single-channel images are defined as gray scale images made of one of the primary colors (Red, Green, and Blue color spaces) from the three channels. Most common digital images use three

channels to collectively form a colored image. The partial CNN (PCNN) model can process one or three channels from an image at the specific instance which the image represents; hence no temporal dimension is considered. The utilization of Recurrent Neural Networks (RNNs) was considered, but due to the size and complexity of the partial CNN model, the utilization of RNNs would have significantly increased the training time of the model without sufficient improvements (Lops et al., 2020). Thus, the temporal dimensionality of the PCNN model is applied through the inclusion of gray-scale images of the TROPOMI images for each channel within the digital image format.

The original partial convolution padding process is designed to gradually reduce the significance of the missing data mask at each encoding phase of the PCNN model. Compared to regular convolutions, the partial convolution process allows the convolution to only depend on the valid pixels, and the normalization is adjusted to the fraction of missing data. During this convolution process, the convolution padding of the mask is applied in unison across all channels. Unfortunately, utilizing different masks for each channel with the conventional convolution kernel causes the convolution kernel to process all the channels at once as one unified mask. This process does not reduce the significance of the mask at a gradual pace but at a much faster rate than expected and losing the individual feature of the mask, thus reduce the potential performance of the model. To address this limitation, we replaced the conventional 2D convolution layer (within the partial convolution) with depthwise convolutions (Chollet, 2017) within the encoding phase of the PCNN model (see Figure 4.2). A comparison of the regular convolution padding and depthwise padding can be found in Figure F18. The decoding

phase of the U-net architecture is unchanged since the mask significance is already removed at the final encoding process of the model. We implemented the algorithm in the Keras and TensorFlow environments (Chollet, 2015; Chollet, 2018; Abadi et al., 2016).

The PCNN model was trained on CMAQ $TCDNO_2$ images of model runs for 2018-2019 and extracted missing data masks from the TROPOMI images. CMAQ TROPOMI images underwent a partial normalization process by dividing the entire dataset by a set value of $1x10^{17}$ based on slightly above the maximum $TCDNO_2$ value within the TROPOMI 2019 dataset. This ensures proper distribution or regularization of the data to improve model performance by reducing the significance of rare outliers within the dataset (Sola & Sevilla, 1997; Eslami et al., 2019; Sayeed et al., 2021). To further increase the training samples and enhance the robustness of the model, we implemented two phases of image augmentation (transforms data into modified samples) (Fawzi et al., 2016) of the CMAQ images for training (see Table T11). The first phase involves applying a modified form of white noise within the CMAQ images to better represent the pixel variations of the original TROPOMI images (see Figure F19). The second phase involved basic rotation and flipping of the image as well as adding a randomized linear function to add a positive or negative value shift of the image. TROPOMI $TCDNO_2$ missing data masks also underwent basic augmentation (randomized flipping and rotation) and overlaid for each CMAQ image.

The partial convolution model structure consists of 16 total layers comprised of one input layer, seven depthwise partial convolution encoding layers, seven partial convolution decoding layers, and an output layer. Each encoding layer contained a pooling layer, a depthwise partial

convolution layer with batch normalization. Each decoding layer was comprised of an upsampling layer concatenated with a respective layer in the encoding layer. The upsampled layer was then processed through a partial convolution layer with batch normalization. Each encoding and decoding layer used the Leaky Rectified Linear Unit (ReLU) activation function (negative slope coefficient = 0.5) due to the input data containing both negative and positive values. Leaky ReLU prevents information loss and allows the negative parts of features within the convolution to activate (Maas et al., 2013; Zhang et al., 2017). The final decoding layer was then processed through a regular convolutional layer with the hyperbolic tangent (tanh) activation function, which provided the final output image (see Figure 2) for the schematic of the system. The loss function used for the model training was the same as that in Liu et al. (2018):

$$L_{total} = L_{valid} + 6L_{hole} + 0.05L_{perceptual} + 120\left(L_{style_{out}} + L_{style_{comp}}\right) + 0.1L_{tv} \quad (12)$$

which comprises pixel hole loss (L_{hole}), pixel valid loss (L_{valid}), perceptual loss ($L_{perceptual}$), raw style output ($L_{style_{out}}$), composited output ($L_{style_{comp}}$), and total variation loss (L_{tv}).

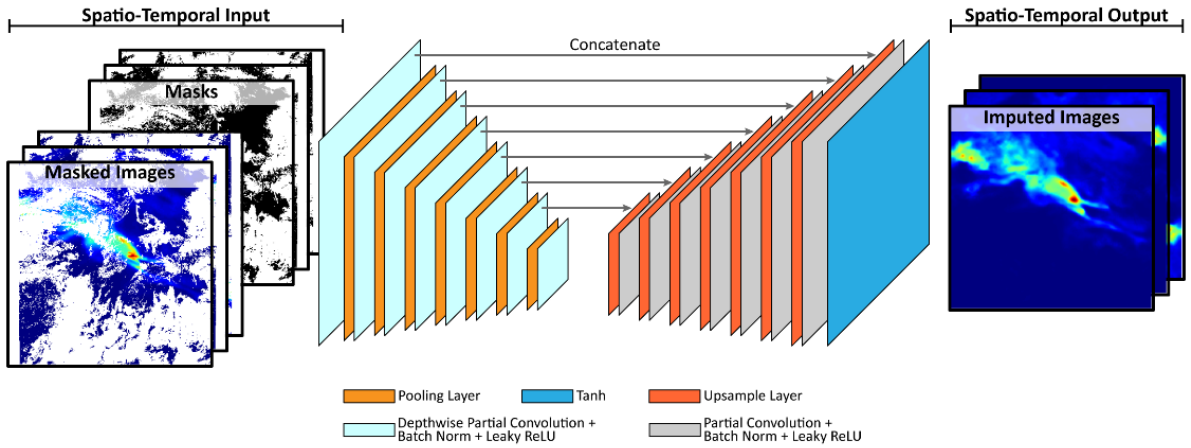

Figure 4.2: Schematic structure of the Depthwise Partial CNN (DW-PCNN) model for imputing missing remote sensing data.

The training of the models was comprised of three phases: i) the first training phase with batch normalization for 400 epochs with a learning rate of 0.001; ii) the second training phase without batch normalization for 800 epochs with a learning rate of 0.001, and iii) the final training phase with a reduced initialized learning rate of 0.0001 for 800 epochs. Batch normalization is performed during the first training phase to improve the initial speed of the model training by more effectively reducing the loss (Bjorck et al., 2018). The second and third phases exclude batch normalization to better optimize loss and reduce the potential bias of the model for imputation. Based on internal tests, the three-phase training has led to an improvement in loss optimization and a reduction in overall training time. The optimizer used for the model is the adaptive moment estimation (Adam) (Kingma & Ba 2014) stochastic gradient descent method, which adaptively estimates the first and second-order moments. Checkpoints have been enabled by saving the model with the least validation error from each training phase.

4.2.3 Model Comparisons

For comparing the performance of DW-PCNN, we use non-temporal based imputation methods such as Inverse Distance Weighting (IDW) (Lu & Wong, 2008) and regular PCNN without depthwise convolutions. The IDW interpolation method assumes that pixels close to each other are more likely to have similar values and the local influence of available points on predictions falls off with distance (Fisher et al., 1993). These two models were the best performants in a previous study for non-temporal imputation of Geostationary Ocean Color Imager (GOCI) Aerosol Optical Depth (AOD) images (Lops et al., 2021). Implementing Spatio-Temporal Kriging (ST-Kriging) method (Kyriakidis & Journel, 1999) as a direct comparison to the DW-PCNN was not possible due to the size of the datasets leading to significant processing time and requiring more memory than the High-Performance Computing system could allocate. We utilized IDW for weekly mean TROPOMI images to fill in any remaining missing data and integrated this result with CoKriging (Kyriakidis & Journel, 1999). The IDW-CoKriging coupled system utilizes IDW imputed weekly mean (as a substitute to the temporal mean of the dataset) and fed to the CoKriging process as a co-variable. CoKriging takes advantage of the covariance of the potential relationship of regionalized variables (weekly mean within filled missing datasets by IDW) during the imputation process. Kriging is based on Gaussian process regression and assumes that spatial variation in a phenomenon is statistically homogeneous throughout the surface based on available data from nearby locations (Cressie, 1990). Both Kriging and IDW (weighting power = 5) models were based on the *gstat* package (Pebesma, 2004).

4.2.4 Evaluation

We evaluated the various models based on various datasets and methods. Since we trained the PCNN model on CMAQ data, we did not conduct an evaluation based on the CMAQ data. All daily images had measurements of $TCDNO_2$ values within the 2019 and 2020 study periods. We evaluated the imputation models based on TROPOMI NO_2 images by processing the daily TROPOMI images into the weekly moving average of TROPOMI images. A strong temporal correlation (0.96 *r* for 2019 and 2020) was observed in the weekly averages between daily variations (see Figure F20); thus, we performed a weekly shift format (0.69 and 0.68 *r* for 2019 and 2020, respectively) as input for the spatio-temporal imputation models. The evaluation of TROPOMI $TCDNO_2$ is conducted by applying TROPOMI $TCDNO_2$ daily missing masks on the weekly mean $TCDNO_2$ images. The models are then tasked to accurately impute the $TCDNO_2$ images. In addition, we apply an estimated pixel distance function based on the distance to the nearest available data point within each image mask. This process is performed to evaluate the imputation bias of the models to the distance of the nearest data variable (see Figure F21).

4.3 Results and Discussion

4.3.1 Imputation of the TROPOMI Images

The results of the TROPOMI NO_2 are based on the weekly mean mask and the difference to the daily masks for the respective day. We compiled the image data into a one-dimensional format for each daily measurement from the TROPOMI measurements for 2019 and 2020 separately and evaluated each daily dataset based on statistical evaluation methods. The dataset

was comprised of measurements of the index of agreement (IOA) (Willmott et al., 1985), the correlation coefficient (*r*) (Benesty et al., 2009), the root mean square error (RMSE), and the mean absolute error (MAE) (Chai & Draxler, 2014). In addition, we evaluated the IOA performance of each model based on the percentage of missing data within the TROPOMI $TCDNO_2$ image. The purpose of this evaluation was to identify how each model performed based on the relative amount of missing data within an image. The TROPOMI dataset for 2019 comprised of images containing 1-20% of missing data with ~17% of images for 2019 and 30% of images for 2020 having 10% or more missing pixels within an image (see Figure F22).

Statistical results of the various models and algorithms for the TROPOMI 2019 cases are shown in Figure 4.3 and TROPOMI 2020 cases are shown in Figure 4.4. For both 2019 and 2020 TROPOMI $TCDNO_2$ images, the DW-PCNN model achieved the best overall performance in IOA (0.81 for 2019 and 0.82 for 2020) and *r* (0.74 for both 2019 and 2020) of all the models. The default PCNN model with spatio-temporal data achieved the lowest MAE (5.77×10^{14} molecules/cm^2) and RMSE (8.45×10^{14} molecules/cm^2) statistical results for 2019, while DW-PCNN had the lowest MAE (4.55×10^{14} molecules/cm^2) and RMSE (6.21×10^{14} molecules/cm^2) scores for 2020. For 2019, there was minimal difference between IDW and the coupled IDW-CoKriging imputation model for overall statistical comparisons (0.15% and 0.04% for IOA and *r*, respectively). Due to the reduced processing resources required, we focused on the IDW model for 2020 comparisons. Based on the percentage of missing values within the evaluated pixels, the DW-PCNN model outperformed all models. For 2019, DW-PCNN outperformed

the other methods 6-34% with images with more than 10% missing data, 10-14% with images between 5% and 10% missing data, and 4-7% with images with less than 5% missing data. For 2020, DW-PCNN outperformed the other methods 8%-11% with images with more than 10% missing data, 4-11% with images between 5% and 10% missing data, and 2-11% with images with less than 5% missing data. In contrast, the IDW-CoKriging model and default PCNN models had mixed results with no clear indicator of which model is better overall (see Table T12). Furthermore, both DW-PCNN and default PCNN with spatio-temporal data were able to impute a TROPOMI image with over 95% missing data significantly better than the over models (see Figure F23). The merits and demerits of the models and algorithms based on the results in imputing TROPOMI NO_2 are provided in Table T13.

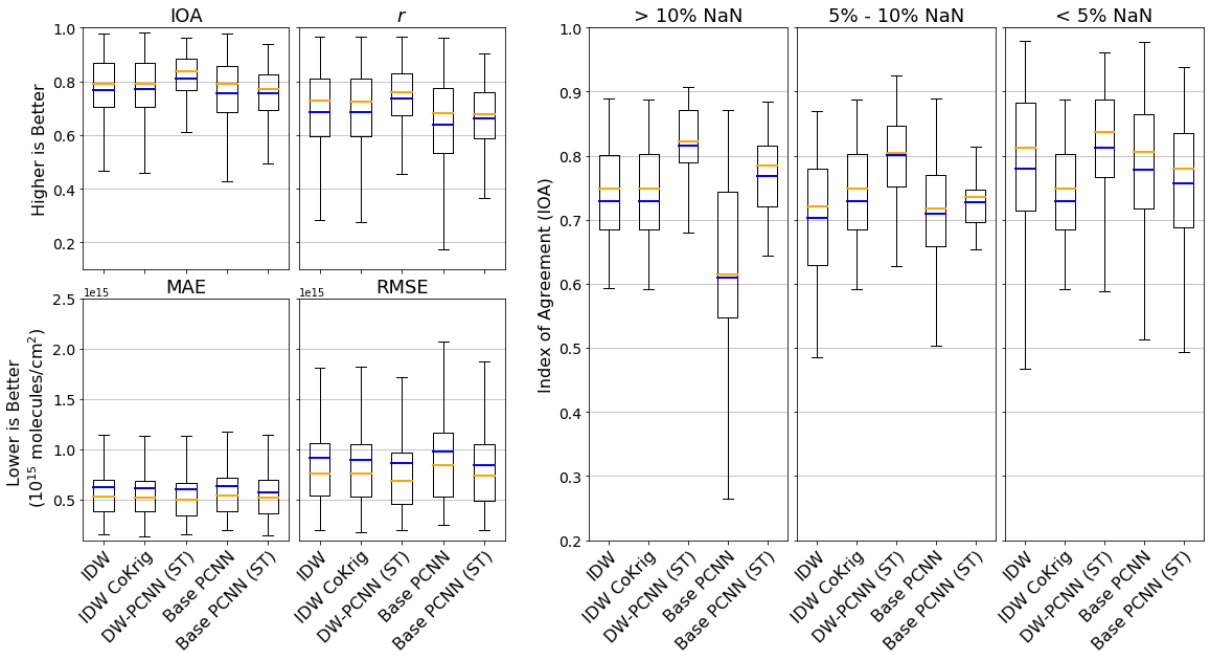

Figure 4.3: Statistical comparison of imputed missing TROPOMI 2019 data from the daily mask to the weekly mean mask. The left figures indicate the statistical performances of Inverse Distance Weighting (IDW), IDW with

CoKriging (IDW CoKrig), Depwthwise Partial CNN (DW-PCNN), default Partial CNN without spatio-temporal data (Base-PCNN), and PCNN with spatio-temporal data (Base-PCNN (ST). Evaluations are based on the index of agreement (IOA), the correlation coefficient (*r*), the mean absolute error (MAE), and the root mean squared error (RMSE). The right figure indicates the IOA performance of the models, based on the percentage of missing data split into three categories. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range. The blue and yellow horizontal lines represent the mean and median of the dataset, respectively.

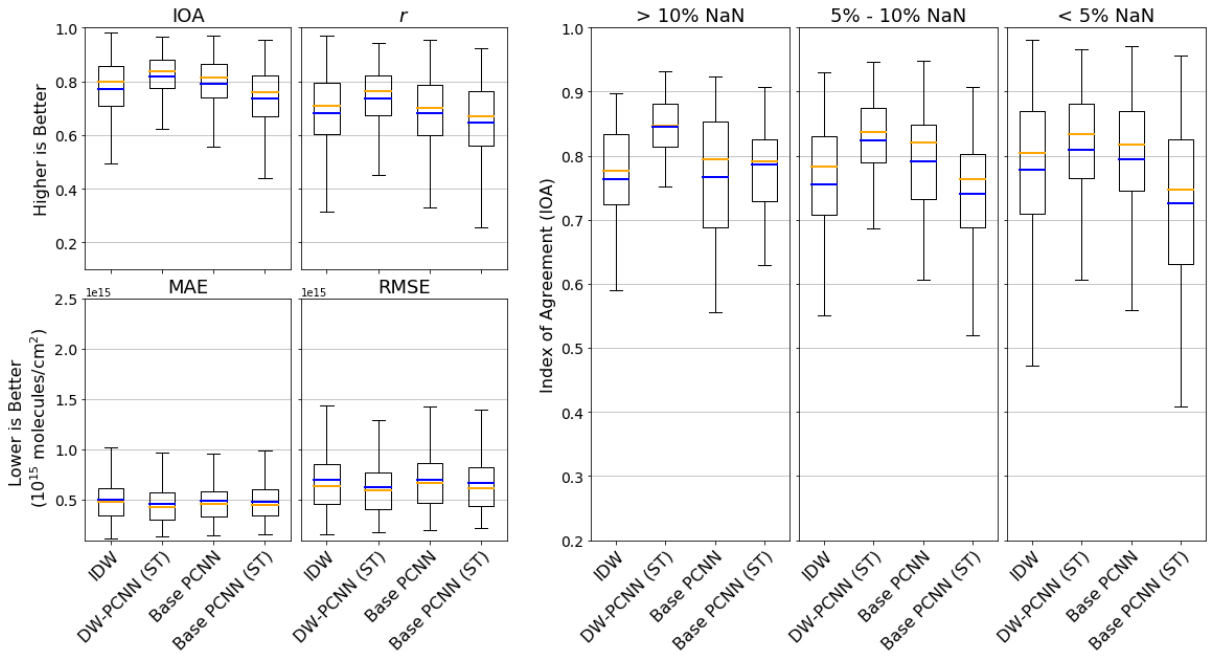

Figure 4.4: Statistical comparison of imputed missing TROPOMI 2020 data from the daily mask to the weekly mean mask. The left figures indicate the statistical performances of Inverse Distance Weighting (IDW), Depwthwise Partial CNN (DW-PCNN), default Partial CNN without spatio-temporal data (Base-PCNN), and PCNN with spatio-temporal data (Base-PCNN (ST). IDW with CoKriging has been excluded due to the minimal overall performance compared to default IDW. Evaluations are based on the index of agreement (IOA), the correlation coefficient (*r*), the mean absolute error (MAE), and the root mean squared error (RMSE). The right figure indicates the IOA performance of the models, based on the percentage of missing data split into three categories. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range. The blue and yellow horizontal lines represent the mean and median of the dataset, respectively.

4.3.2 Pixel Distance Evaluation

We evaluated the TROPOMI imputation comparison by applying a pixel distance mask relative to the closest available data point within the TROPOMI image. To estimate the pixel distance from the nearest available data point, we applied a Euclidean process for each daily mask and rounded to the nearest integer to represent the pixel distance. We calculated the bias of the imputation methods for each pixel, categorized them based on the distance value of the respective pixels from the pixel distance mask (see Figure F21), and sorted the pixels of each imputed $TCDNO_2$ within the evaluation mask at 4pixel intervals representing ~28km x 14km distances. At each interval, we plotted the difference between the biases of the imputed $TCDNO_2$ values to that of the TROPOMI $TCDNO_2$ value as a boxplot. The series of boxplots at each pixel interval provides information on the bias range of each model as the distance increases. We evaluate DW PCNN, IDW CoKriging, IDW, PCNN, and PCNN with spatio-temporal data (ST) models. The pixel distance plot (see Figure 4.5 and Figure 4.6) shows that the PCNN models had similar variance and range of biases throughout the pixel distance range. The default PCNN without spatio-temporal data (PCNN) had a slight negative bias, while the DW-PCNN and PCNN (ST) had a slight positive bias. We used IDW CoKriging (IDW CoKrig) in the comparison due to the slightly lower overall bias over the default IDW model. IDW CoKriging had an overall negative bias, with some fluctuations, but the variance and range of the bias narrowed down more significantly than the PCNN models as the distance increased.

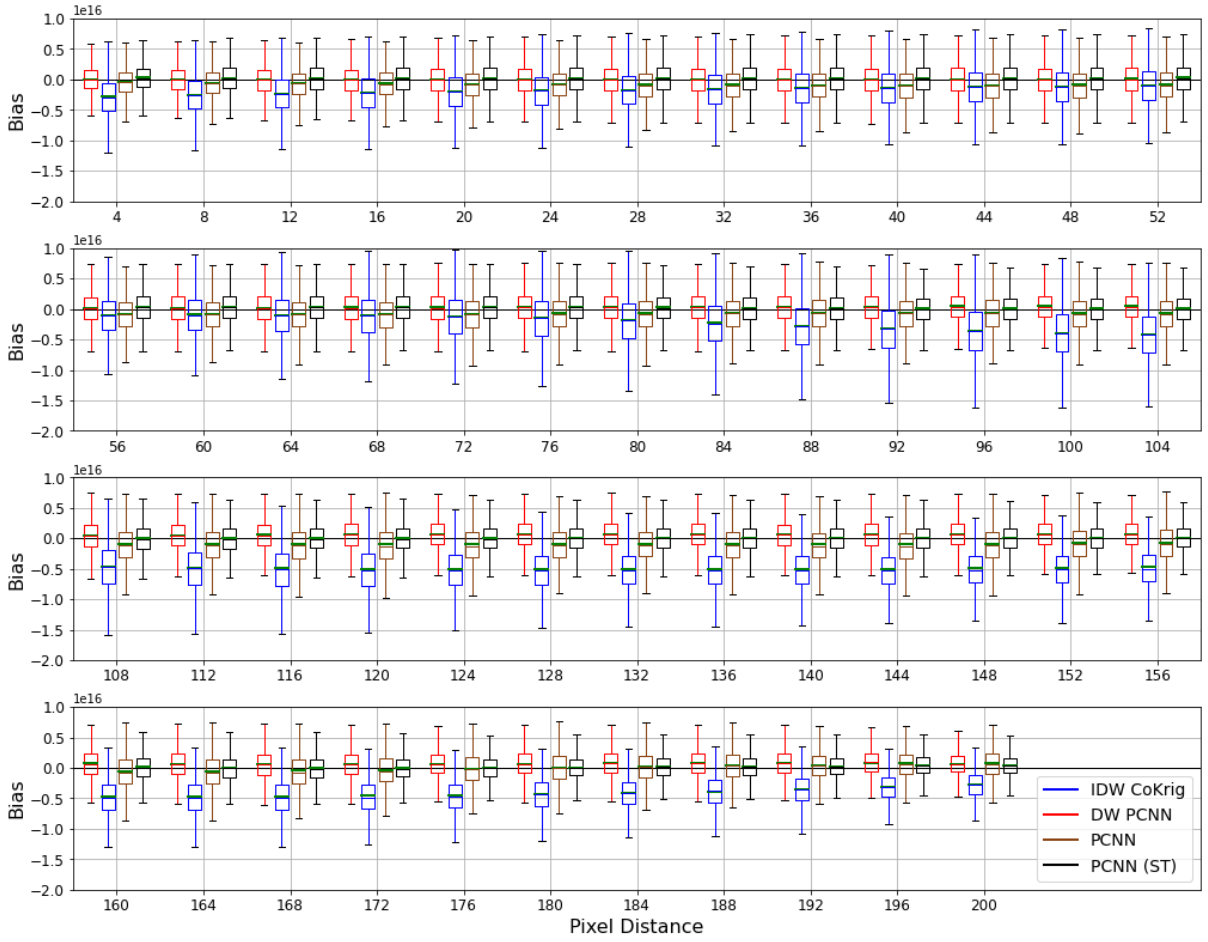

Figure 4.5: Comparison of the bias-variance in the pixel distances of TROPOMI 2019 NO_2 of the Depthwise Partial CNN (DW PCNN), coupled Inverse Distance Weighting with CoKriging (IDW CoKrig), default Partial CNN without spatio-temporal data as input (PCNN), and default Partial CNN with spatio-temporal data as input (PCNN (ST)). The distances are split into four sections at 52-pixel intervals. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The horizontal lines represent the mean bias of DW PCNN (red), IDW CoKrig (blue), PCNN (brown), PCNN (ST) (black) models, and the median bias (green for all models) of the imputing missing TROPOMI data. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range.

For TROPOMI 2020, the biases of the model imputations were significantly lower than the 2019 biases over distance. PCNN (ST) had the largest bias within near pixel distances of measured TROPOMI data but decreased as the pixel distance increased with a slight positive bias compared to the PCNN and DW PCNN models. In contrast to the other models, PCNN

experienced an increase in overall bias and range as the pixel distance increased beyond the 24-pixel (~168km) distance threshold. IDW also achieved relative stable biases over the distance ranges. Despite the slightly larger bias range in relation to DW PCNN and PCNN (ST), IDW had the most consistent mean bias without a positive or negative trend compared to all the models. DW PCNN had the narrowest bias range of the models across all the pixel distances for the 2020 TROPOMI dataset.

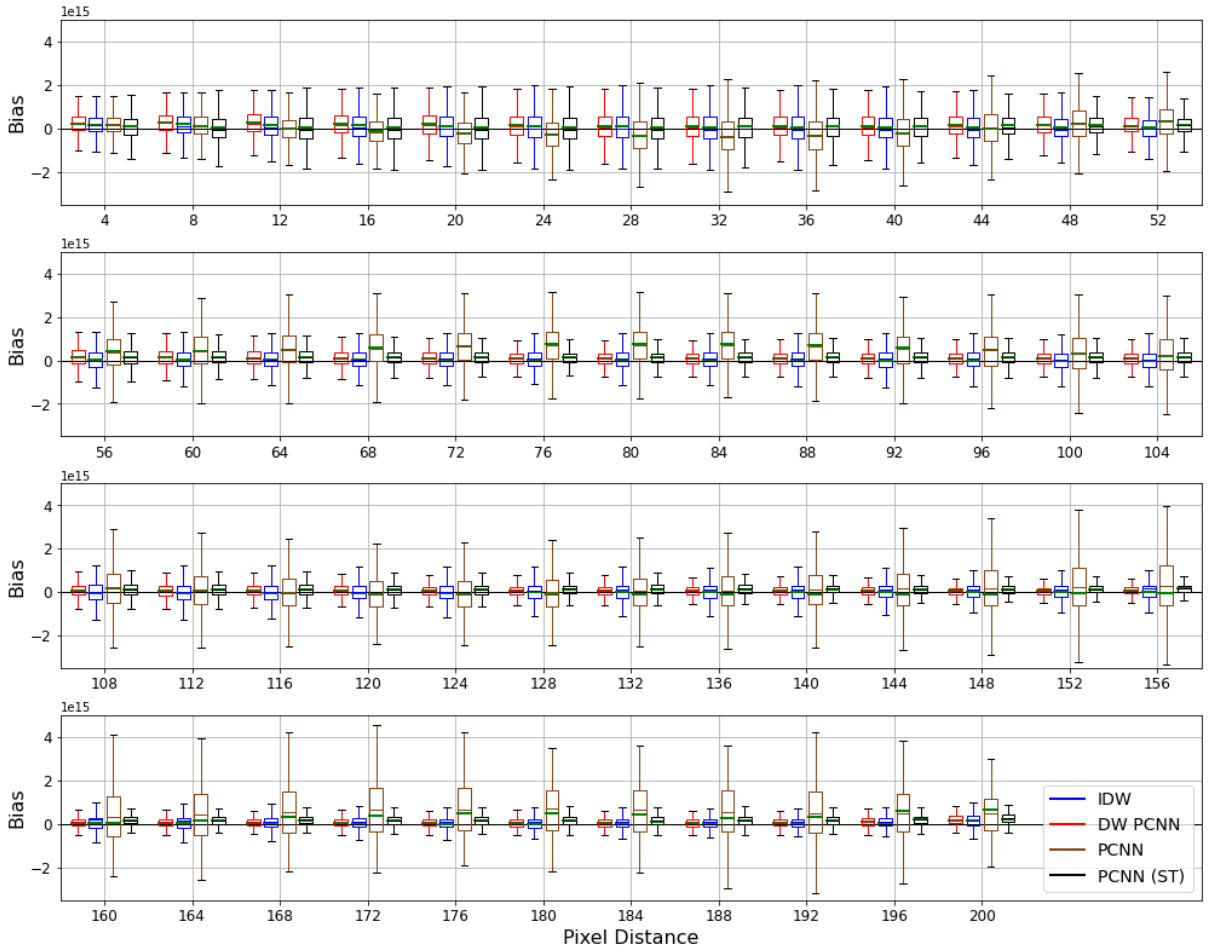

Figure 4.6: Comparison of the bias-variance in the pixel distances of TROPOMI 2020 NO_2 of the Depthwise Partial CNN (DW PCNN), coupled Inverse Distance Weighting (IDW), default Partial CNN without spatio-temporal data as input (PCNN), and default Partial CNN with spatio-temporal data as input (PCNN (ST)). The distances are split into four sections at 52-pixel intervals. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The horizontal lines represent the mean bias of DW PCNN (red), IDW (blue), PCNN (brown), PCNN (ST) (black) models, and the median bias (green for all models) of the imputing missing TROPOMI data. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range.

4.4 Conclusion

This research demonstrated the improved capability of the depthwise partial convolutional neural network in the application of spatio-temporal imputation of missing remote sensing data. Both 2019 and 2020 TROPOMI $TCDNO_2$ imputation results demonstrated DW-PCNN as the

best performing model to impute TROPOMI $TCDNO_2$. The addition of spatio-temporal data for the PCNN model has shown significant improvements over the regular PCNN model (without temporal data) with datasets containing large percentages of missing data and at extended distances. Despite the improvements of adding temporal dimensionality within the input for the PCNN model, the mask padding of the regular convolution process does show limitations and has still led to some biases. The implementation of depthwise convolutions, where masks at each image channel are padded separately, has shown further improvement and demonstrates the importance of maintaining the individual channel masks and gradual feature reduction over the conventional method. Furthermore, DW PCNN was the only model to maintain a mean IOA above 0.8 for all statistical comparisons for both TROPOMI 2019 and 2020 $TCDNO_2$ datasets. The current limitation of DW-PCNN is the constrained number of channels (only three channels) the model can process, thus limiting the temporal samples and the addition of co-variables that can facilitate the imputation performance.

While the bias-distance comparison has shown PCNN ST perform as well as DW PCNN, the statistical comparisons have shown PCNN ST have a lower correlation in reconstructing TROPOMI $TCDNO_2$ than the default PCNN model. This can be explained by the output of the PCNN ST model having much smoother transitions than the PCNN or DW PCNN models and slight under-prediction of high $TCDNO_2$ concentrations and over-prediction of low $TCDNO_2$ column concentrations. This phenomenon may minimally impact the bias evaluation but does impact the correlation and IOA scores.

For a spatial imputation algorithm without temporal dimensionality, IDW performed consistently compared to the PCNN and PCNN ST models. While IDW was not able to surpass the PCNN models in different metrics, it also did not perform the worst in the respective metrics. The major limitation of IDW is the computation cost of such large datasets, especially when required to take all available samples within the TROPOMI image. As such, IDW and other interpolation-based algorithms (e.g., Kriging) require exponentially more computation power and resources as the dataset size increases spatially and with the addition of temporal dimensionality. For smaller dataset sizes, these algorithms do demonstrate competitive performance to the PCNN model (refer to Lops et al., 2021). As datasets increase in size, dimensionality, and missing data, the benefits of deep learning algorithms for imputation purposes also increase.

Once trained, the DW PCNN model imputes large remote sensing datasets in significantly less processing time than the interpolation-based algorithms and performs much better than the default PCNN models with and without temporal dimensionality of datasets. For the accurate imputation of such datasets, the implementation of DW PCNN allows the accurate imputation of remote sensing data with large regions of missing data, which will benefit future researchers that conduct other studies such as data assimilation for numerical models, emission studies, and human health impact analyses from air pollution. To further enhance the imputation accuracy of the DW PCNN model, the expansion of the number of image channels the PCNN can receive needs to be increased. Furthermore, with the increase in input channels, the addition of more

temporal information and co-variables as input will allow the model to further improve the imputation capability over previous iterations and other algorithms.

CHAPTER 5

Development of Deep Convolutional Neural Network Ensemble Models for 36-Month ENSO Forecasts[4]

5.1 Introduction

Climate change persists to be a growing problem as greenhouse gases (GHGs) and aerosols continue to flood the atmosphere and affect its physics and chemistry (Mikhaylov et al., 2020; Wu et al., 2018). These changes in the atmosphere have led to temperature rise; longer, more intense droughts and storms; increased risk of wildfires (Jones et al., 2020); and sea-level rise (Vu et al., 2018). All these factors will continue to affect water supply and quality (Hashempour et al., 2020), crop and food production (Lal, 2020), human health (Campbell-Lendrum & Prüss-Ustün, 2019), shelter, wildlife, and much more.

The state of El Niño-Southern Oscillation (ENSO) is based on the sea surface temperature (SST) patterns and interannual fluctuation over the equatorial Pacific Ocean. ENSO has been shown to have impacts and global teleconnections on regional temperature, precipitation, and mid-tropospheric atmospheric circulation (Diaz et al., 2001) and has been used as a predictor for regional weather (Trenberth & Stepaniak, 2001). Nino3.4 is the most used index to represent the state of ENSO that covers the equatorial region 5°N – 5°S, and 170°W – 120°W (Bunge & Clarke, 2009). As such, dynamical and statistical models have been developed over the past

[4] Under preparation for submission

decades to predict seasonal ENSO but are not able to reliably predict ENSO due to its chaotic yet deterministic fluctuations (Chen et al., 2004; Gupta et al., 2020). As scientists attempt to find the solution to climate change, many are turning to deep learning (DL) for insight into how to predict and mitigate climate change. Artificial neural networks (ANNs) constitute one application of DL to climate modeling, and they allow the creation of climate simulations to reveal inter-variable relationships within the atmosphere (Labe & Barnes, 2021; Tzuc et al., 2020). They also allow pertinent simulations of atmospheric physics and chemistry (Lauret et al., 2016).

Deep neural networks (DNNs) are one of the common applications of DL to climate science. DL algorithms (LeCun et al., 2015) have already shown significant promise in various applications by modeling high-level abstractions within datasets (Bengio, 2009; Deng & Yu, 2014). DNNs can be utilized for highly accurate climate forecasting, which allows for improved climate adaptation and mitigation. DNNs have been utilized to parameterize individual variables and "subgrid processes" to demonstrate their discrete impacts on the model (O'Gorman & Dwyer, 2018; Yuval & O'Gorman, 2020). Convolutional Neural Networks (CNN) (Krizhevsky et al., 2017) have been among the most successful and widely used neural networks thus far (LeCun & Bengio, 1995; Schmidhuber, 2015) for various purposes such as air quality and meteorological forecasting (Ghahremanloo et al., 2021; Lops et al., 2019; Sayeed et al., 2020, 2021; Yeo et al., 2021; Zhang et al., 2015), classification (Anthimopoulos et al., 2016; Lee et al., 2017; Li et al., 2014), speech recognition (Mikolov et al., 2011; Park et al., 2017), and imputation of temporal and spatial datasets (Ghahremanloo et al., 2021; Li et al.,

2017; Zhang et al., 2018; Lops et al., 2021). CNNs have also been utilized to forecast climate (Fedotova & Luferov, 2019; Ham et al., 2019; He et al., 2019; Mu et al., 2019) and other types of networks for climate forecasting include physics-informed neural networks (PINNs) (de Wolff et al., 2021), and very simple neural networks (Mooers et al., 2021) among others. Furthermore, other applications of DL to climate modeling can be seen in the incorporation of DL into numerical modeling technology (Partee et al., 2021) and the utilization of the DL platform Tensorflow to model climate via multiple linear regression analysis (Kim, 2019).

5.2 Methods

5.2.1 Climate Data

In relation to climate data, it is observed and understood that climate and their respective indices have teleconnections on the regional climates and other climate indices (Diaz et al., 2001; Yuan et al., 2018). Hence, the constraints of featuring only time-series data would limit the capability of the model to identify the spatial and temporal teleconnections within climatic data. The implementation of 2D-CNNs for forecasting allows the utilization of climate model and reanalysis datasets with both spatial and temporal dimensions – thus including the potential teleconnections of climate. For the study, 2D-CNN is utilized for climate forecasting due to the convolution kernels being able to extract spatial- and temporal features from the datasets and provide a more generalized model capability for long-term forecasts. The integration of diverse climatological data from model and reanalysis datasets offers enhancements in data dimensions

and allows the expansion of forecasting of climate indices (e.g., Nino3.4) by up to 36 months with sufficient accuracy.

5.2.2 Data Preparation and Processing

Datasets used for training and evaluating the 2D-CNN model are from the Coupled Model Intercomparison Project (CMIP5), ECMWF Reanalysis v5 (ERA5), Global Ocean Data Assimilation System (GODAS), and Simple Ocean Data Assimilation (SODA) to forecast Nino3.4 temperature anomalies, which represents the El Niño-Southern Oscillation (ENSO) index. For the model forecasting, we utilized 9 ocean and atmospheric parameters for the forecasting of Nino3.4. For ocean parameters, we used Sea Surface Temperature (SST), Mean Potential Temperature (MPT) within the 300-meter depth, Meridional (OV) and Zonal (OU) Current velocities. We computed the monthly SST anomaly (SSTA) and MPT anomaly (MPTA) based on subtracting the long-term climatology (between 1900 and 2000 time-period) to the monthly SST and MPT values. For atmospheric parameters, we used Precipitation (PR), Pressure at Sea Level (PSL), Surface Temperature (ST), and Meridional (AV) and Zonal (AU) velocities of wind. CMIP5 datasets contained both atmospheric and ocean variables, while the respective reanalysis datasets had to be merged from different sources. For ocean parameters of reanalysis data (SST, MPT, OV, and OU), we used SODA and GODAS, while atmospheric parameters (PR, PSL, ST, AV, and AU) of reanalysis data was acquired from ERA5. Due to the different sources, the reanalysis datasets were temporally aligned, processed, and merged for SODA and GODAS periods.

The datasets for all parameters were spatially constrained with 60° N and 55° S. The calculation of Nino3.4 from SSTA are based on the mean of the constrained region within 5° N – 5° S, and 170° W – 120° W (Chen et al., 2004; Gupta et al., 2020). To validate the anomaly calculation, we compared the National Oceanic and Atmospheric Administration's (NOAA) Nino3.4 index (NOAA, 2021) to NOAA Extended Reconstructed Sea Surface Temperature V5 (https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html) in (which the SSTA calculation was tested) with a correlation of 0.996 (see Figure F24). We then compared the calculated NOAA SSTA to the GODAS SSTA (see Figure F25) to validate the reanalysis data processing for SSTA. Due to the large datasets, we constrained the data size by applying down-sampling. The down-sampling process utilized bi-cubic and inter-area interpolation of the spatial datasets to spatial dimensions of 32 by 96 pixels in the latitude and longitude dimensions, respectively. The model receives twelve months of input for each parameter (for a total of 108 samples) to forecast 1-36 lead months in Nino3.4. Thus, each sample the model receives is comprised of 32×96×108 (331,776 values total) to forecast up to 36 months. Normalization was based on defined min and max variables for each input parameter (see Table T14). This ensures proper distribution or regularization of the data to improve model performance by reducing the significance of rare outliers within the dataset (Sola & Sevilla, 1997; Eslami et al., 2019; Lops et al., 2021; Sayeed et al., 2021).

5.2.3 Coupled Model Intercomparison Project

Coupled Model Intercomparison Project phase5 (CMIP5) is comprised of more than 20 modeling groups to develop and study community-based coupled atmosphere-ocean general circulation models (AOGCMs). The goal of these simulations was to help improve our understanding of climate change and variability (Taylor et al., 2021). We acquired CMIP5 datasets from the Earth System Grid Federation (ESGF) managed by the Department of Energy Lawrence Livermore National Laboratory (LLNL) (data available at https://esgf-node.llnl.gov/search/cmip5/).

Table T15 lists the CMIP5 IDs, modeling groups and model time-period. We only extracted the first ensemble member (the first realization, initialization, and set of perturbed physics, which is denoted "r1i1p1") from each historical experiment from the CMIP5 model outputs (Peng et al., 2020). The outputs were split between the ocean and atmospheric domains, which have been merged for the data processing for the forecasting CNN model.

5.2.4 ECMWF Reanalysis v5

The European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) atmospheric variables were used as input for the Deep CNN model. ERA5 is the 5th generation climate reanalysis dataset from ECMWF and provides a detailed record of the global atmosphere, oceans, and land surface. ERA5 reanalysis is based on the Integrated Forecasting System (IFS) Cy41r2 with 4DVar data assimilation. The ERA5 datasets have an hourly monthly average temporal resolution and spatial resolution of ~31km. The vertical resolution of the

dataset is comprised of 137 hybrid sigma-pressure vertical levels, with the top-level located at 0.01 hPa (~80km altitude). At the time of writing, the datasets from January 1979 to the present (September 2021) have been made available to the public. The reanalysis is expected to cover the time period from January 1950 to the present (Hersbach et al., 2020; Hoffmann et al., 2019; Urraca et al., 2018).

5.2.5 Global Ocean Data Assimilation System

The Global Ocean Data Assimilation System (GODAS) was implemented and is managed by the National Centers for Environmental Prediction (NCEP). The reanalysis datasets have a spatial resolution of 0.3 degrees latitude and 1-degree longitude for the entire quasi-global domain (75°S–65°N) configuration of the Geophysical Fluid Dynamics Laboratory's (GFDL) Modular Ocean Model version 3 (MOM.v3). GODAS receives temperature observations from Expendable Bathythermographs (XBTs) (Ishii & Kimoto, 2009), Tropical Atmosphere Ocean (TAO) array in the tropical Pacific (McPhaden, 1995), Triangle Trans-Ocean Buoy Network (TRITON) in the tropical Indian Ocean (Hase et al., 2008), Prediction and Research Moored Array in the Tropical Atlantic (PIRATA) (Servain et al., 1998), and Argo profiling floats (Gould et al., 2004) for ocean data assimilation (Huang et al., 2008). GODAS provides pentad and monthly ocean analyses and data products from 1979 to the present. The data is hosted and maintained by the NOAA's Climate Prediction Center (Huang et al., 2010) (available at http://www.cpc.ncep.noaa.gov/products/GODAS).

5.2.6 Simple Ocean Data Assimilation

The Simple Ocean Data Assimilation (SODA) is a reanalysis dataset consisting of gridded variables for the global ocean since the beginning of the 20th century. The reanalysis datasets have a spatial resolution of 0.25 degrees latitude and 0.4 degrees longitude with 40 vertical levels (Carton & Giese, 2008). SODA 2.2.4 datasets were acquired, which is based on the Parallel Ocean Program numerics (Smith et al., 1992) with data assimilation (Giese & Ray, 2011). SODA 2.2.4 provides monthly ocean analyses and data products from 1871 to 2010. The data is hosted and maintained by the Asia-Pacific Data-Research Center (APDRC) (available at http://apdrc.soest.hawaii.edu/datadoc/soda_2.2.4.php)

5.2.7 Neural Network Structure

We utilized a deep convolutional neural network (deep CNN) (Krizhevsky et al., 2017) with fully connected layers for the forecasting of the NINO3.4 climate anomaly. CNNs process data by convolving the data at each layer over multiple image channels, assigning weights and biases at various aspects within the data, and differentiating between them. The benefits of CNN models are their ability to reduce data into a more manageable form for processing (without losing important features) and extract high-level features of the input data with kernels during the convolving phase (Lawrence et al., 1997). The utilization of Recurrent Neural Networks (RNNs) would have potentially increased the training time of the model without sufficient improvements (Lops et al., 2020). Thus, due to the size and complexity of the Deep CNN forecasting model, we did not utilize RNNs for the performance comparisons. The implemented

algorithms for CNN models were run on the Keras and TensorFlow 2 environments (Chollet, 2015; Chollet, 2018; Abadi et al., 2016).

The structure of the CNN models comprised of three 2D-Convolutions with kernel sizes of 4×6, strides 2×3, and padding set to '*same*' and one flattening layer with a dropout of 10%. The flattening layer is connected to a dense layer with 128 neurons and dropout of 10%. The final layer is the output layer with 36 neurons. Both dense models have the Leaky-ReLU activation function. Leaky ReLU, with a negative slope coefficient of 0.5, prevents information loss and allows the model to predict negative features and values (Maas et al., 2013; Zhang et al., 2017) that represent the ENSO index. We tested multiple models with swish (Ramachandran et al., 2017) and ReLU (Xu et al., 2015) activation functions, a combination of the number of filters (8/8/8 or 8/16/32) for the 2D-convolution layers and the inclusion of max-pooling layers (Nagi et al., 2011) for each of the first two 2D-convolution layers. The schematic of the model is shown in Figure 5.1.

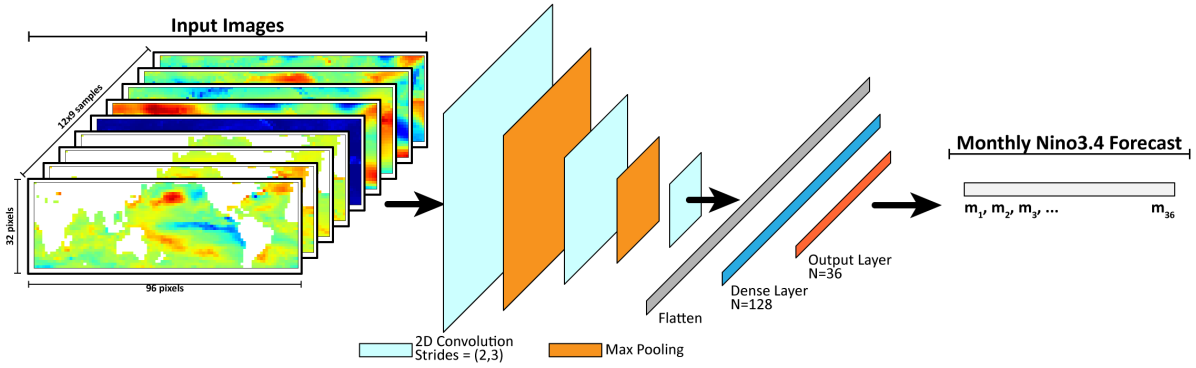

Figure 5.1: Basic schematic of the CNN models for forecasting Nino3.4. Differences between models are whether max pooling is enabled, the number of filters (8/8/8 or 8/16/32), and the activation functions (Swish or ReLU) for the convolution layers.

We utilized the Index of Agreement (IOA) (Willmott et al., 1985) as the loss function for the models. Utilizing IOA as the loss function has been shown to improve model performance for varying long-term forecasting applications (Sayeed et al., 2020; 2021). The mathematical expression of the IOA loss function is as follows:

$$\mathrm{IOA} = 1 - \frac{\sum(O_i - P_i)^2}{\sum(\mathrm{abs}(O_i - \bar{O}) + \mathrm{abs}(P_i - \bar{O}))^2} \quad (12)$$

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

The Deep CNN forecasting model underwent two training phases. The first phase involved training the models on CMIP5 images comprising of Ocean and Atmospheric components with validation checkpoint enabled (saving the model with the lower validation error at each iteration) at 100 epochs. During this phase, we split the training and validation datasets at an 80%/20% distribution. The second phase involved utilizing transfer learning (Torrey & Shavlik, 2010) to retrain the model with SODA ocean and ERA5 atmospheric variables without checkpoint or validation process enabled. The models were retrained on SODA datasets (1980/01-2008/09) for 50 epochs and tested on GODAS from Oct. 2008 (based on 1st month lead time) to Sept. 2021 (based on 36th month lead time).

5.3 Results and Discussion

For the evaluation of the model's forecasting capability, we used the Pearson correlation coefficient (*r*) (Benesty et al., 2009) and IOA (Willmott et al., 1985) to directly compare the

predicted Nino3.4 to the observed Nino3.4 during the testing time period. We tested multiple Deep CNN models to identify optimal parameters for the model to forecast the Nino3.4 index by up to 36 months. Model hyper-parameters (Bergstra & Bengio, 2012) variables consisted of Swish and ReLU activation functions, the inclusion or exclusion of max-pooling layers, and the number of filters arrangement for the convolution layers at 8/8/8 or 8/16/32. Initial testing on the performance of the models was not impacted by the increase in neurons for the first dense layer; thus we set the number of neurons to 128 for the first dense layer. Each model received the same datasets for initial training on CMIP5 and transfer learning utilization on SODA ocean reanalysis with ERA5 reanalysis atmospheric datasets. Results show that some models diverge in performance for different months within the 36-month period of predicting Nino3.4.

Models were identified and labeled by their model parameters. '*C*' designates the convolution layers with the number referring to the number of filters within the convolution layer. The filters are unchanged for the models, which are 4×6 with strides 2×3. '*m*' designates a max-pooling layer between the convolution layers and *Swish* (*ReLU*) are the activation functions used for the convolution layers. Thus, a model with designation *C8m8m8|Swish* has three convolution layers with 8 filters and swish activation function, and max pooling layers between the convolution layers.

The models are evaluated based on their 6^{th}, 12^{th}, 18^{th}, 24^{th}, 30^{th}, and 36^{th}-month forecasting score for IOA and *r*. Due to the divergence of the model performances, to identify the model with the overall best performance, we utilize a cumulative score of the evaluation months for IOA and *r* (see Figure 5.2). The highest cumulative IOA score was achieved by the model

C8m16m32|Swish with a cumulative IOA score of 3.742, while the model C8/8/8|Swish achieved the highest cumulative *r* score (see Figure 5.3). Both models achieved the highest IOA and r scores overall. For 6 and 12-month forecasting accuracy, the model C8m16m32|Swish achieved the highest scores for IOA and *r*. The model achieved 0.83 and 0.68 IOA for 6 and 12-month forecast, respectively, while the model achieved 0.72 and 0.52 *r* for 6 and 12-month forecasts, respectively. Both C8/8/8|Swish and C8/16/32|Swish models achieved very similar performance for forecasting 18 months ahead. Both models achieved an IOA of 0.63 and the C8/8/8|Swish model achieved an *r* of 0.59 while the C8/16/32|Swish achieved an *r* of 0.58. The C8m8m8|Relu model achieved the highest IOA and r (0.67 and 0.47, respectively) of all the models for 24-month forecasting. For both 30 and 36 month forecasts, C8/8/8|Swish model achieved the highest r of all the models (0.46 and 0.48, respectively). In terms of 30-month forecast, the C8m16m32|Swish model had the highest IOA (0.63) while the C8/8/8|Swish model achieved the highest IOA of 0.61 for the 36-month forecast. See Table T16 for the statistical scores for each model and respective lead months at 6-month intervals.

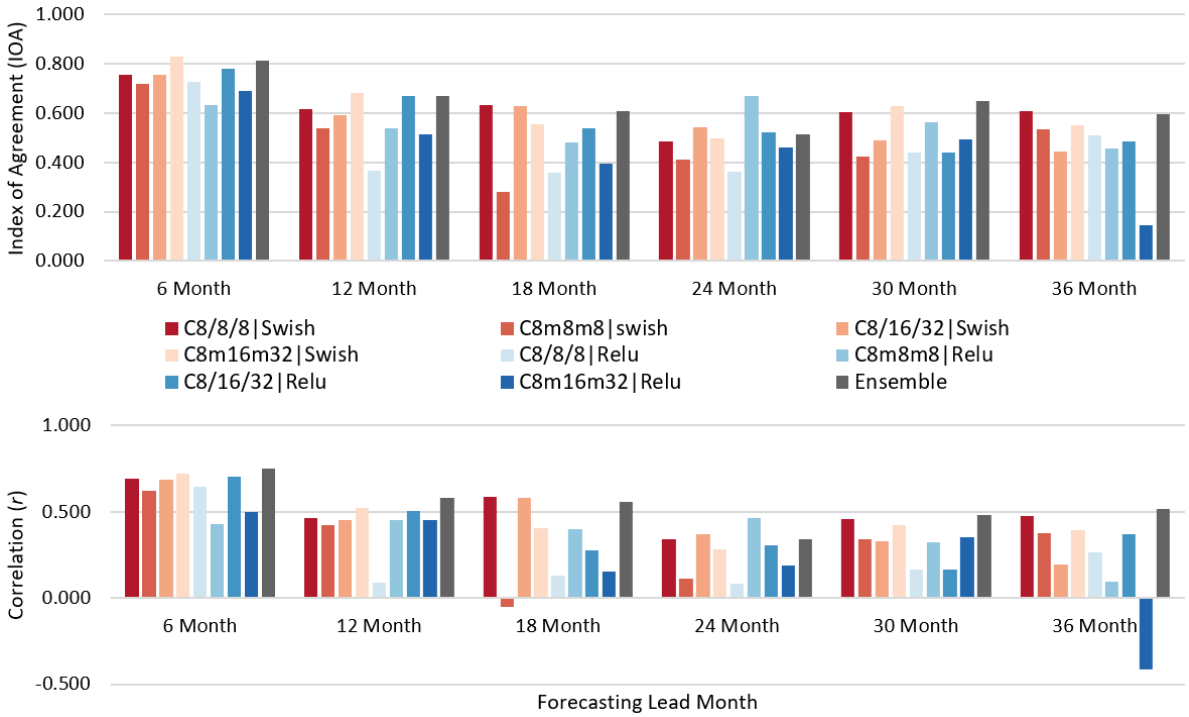

Figure 5.2: Statistical comparisons (IOA and r) of the Deep CNN models based on the lead forecasting months at 6-month intervals. The ensemble is the combined mean output of C8/8/8|Swish and C8m16m32|Swish models.

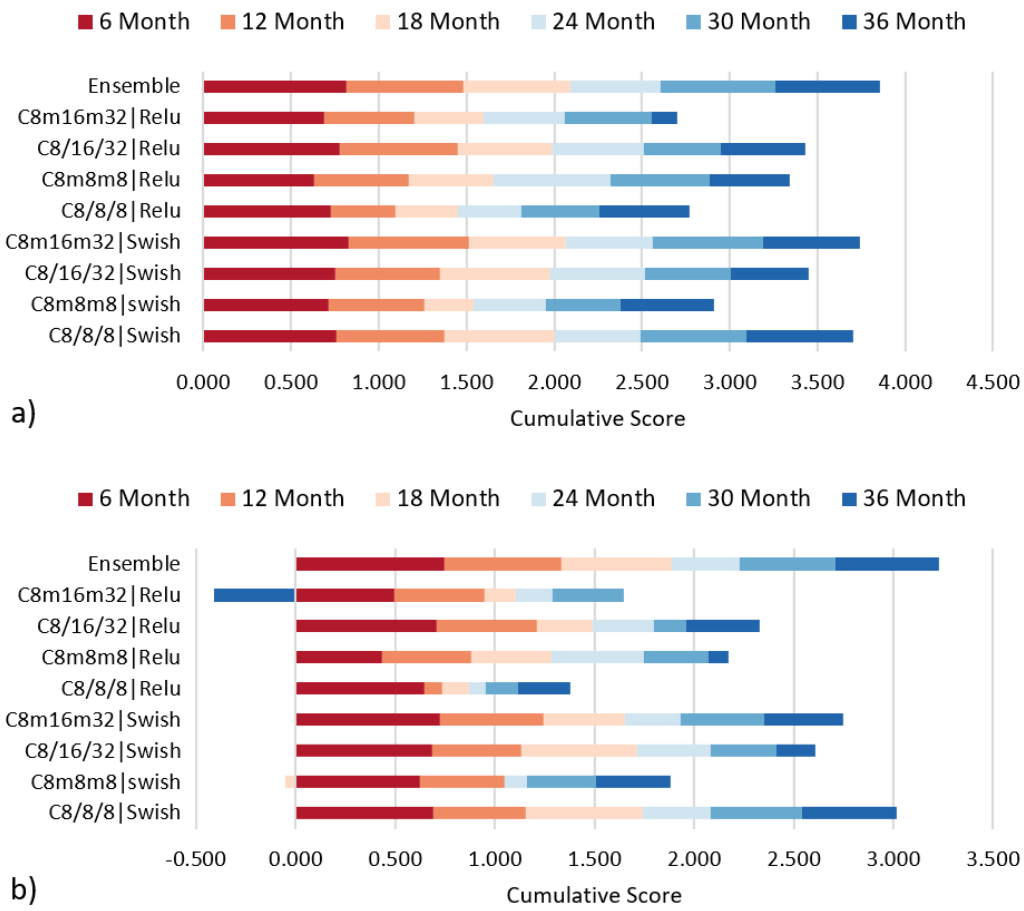

Figure 5.3: Cumulative scores of Index of Agreement (a) and correlation (b) for the different models based on the forecasting lead month at 6 month intervals.

Based on these results, we combined the outputs of the C8/8/8|Swish and C8m16m32|Swish models as an ensemble model output. The ensemble model achieved a 4% and 7% improvement in cumulative IOA and *r* score, respectively, over the C8/8/8|Swish model. With an r of 0.52 for 36-month forecast, the ensemble output provides a moderate correlation performance on predicting the Nino3.4 index.

5.4 Conclusion

ENSO has been shown to have impacts on regional weather and global teleconnections within regional climates. As GHGs and aerosols continue to be emitted into the atmosphere, climate change will persist in being a growing problem due to shifts in local climates and the intensity of climate extremes within regions. These factors will impact human health, water availability, food production, wildlife, and coastal regions from sea level rise. Is it a growing concern to develop dynamical and statistical models to forecast climate over extended periods for better adaption and mitigation strategies from adverse weather and climate. Due to ENSO's global teleconnection and impact on regional climates, scientists have worked on forecasting ENSO, but it has proven difficult due to its chaotic yet deterministic fluctuations (Chen et al., 2004; Gupta et al., 2020). DL has shown promise in identifying non-linear relationships within input datasets for long-term forecasting of air quality and weather forecasting (Ghahremanloo et al., 2021; Lops et al., 2019; Sayeed et al., 2020, 2021; Yeo et al., 2021; Zhang et al., 2015). In this research, we utilize different deep 2D-CNN systems to forecast Nino3.4 SST anomalies by up to 36 months with moderate performance. While the deep 2D-CNN models had different performances based on the 6-month interval lead time, we identified two models with the best overall cumulative IOA and *r* scores. Creating an ensemble forecast by merging the output of the two models, we have shown the system's exceptional forecasting accuracy and capability to forecast 36 months ahead with a moderate score of 0.52 *r* where 0.5 is often forecasting threshold of a model for sufficient forecasting capability (Ham et al., 2029).

While forecasting ENSO (i.e., Nino3.4) remains difficult and predictions remain uncertain, this study demonstrates the potential of DL has shown promise to extend moderately accurate forecasts within (or even beyond) 36 month lead times. Through further optimizations of the model parameters, selection and preparation of input variables, and extension of a historical time period for training the models, there is potential to improve the forecasting capability of the system. Through combination and optimization of model outputs into an ensemble system, the accuracies can be further improved by each model overcoming the shortfalls of the other models in their forecasting.

CHAPTER 6

Conclusion

In this study, deep 1D- and 2D-convolutional neural networks (CNN) and U-net-based 2D-convolutions were used for various atmospheric applications in long-term air quality and climate forecasting, and imputation of remote sensing datasets over East Asia, the Contiguous United States, and globally.

The first task utilized a deep 1D-CNN neural network to forecast pollen concentrations from three different groups of plants: Tree, Grass, and Weed pollen. The Deep CNN algorithm obtained a relatively high index of agreement (IOA) and Pearson correlation coefficient (r) of up to 0.90 and 0.88, respectively. Categorical statistic evaluations based on defined threshold levels also show satisfactory results for the model forecasting capability. The model achieved a Critical Success Index as high as 0.887 for weed pollen, 0.646 for tree pollen, and 0.294 for grass pollen. Forecasts of grass pollen exhibit the largest decrease in accuracy because of the strong variance in annual pollen concentrations. Forecasts of weed pollen exhibit the greatest consistency, with the model achieving a 7-day forecast correlation and index of agreement of 0.82 and 0.77, respectively, during the peak season. This correlates with the consistency of annual and seasonal trends of weed pollen within the study area. Compared to the conventional modeling approaches, the deep convolutional neural network shows a promising ability to predict pollen for multiple days to allow individuals with allergies to take proper precautions during high pollen days.

The second task utilized advanced U-net-based convolutional neural network architecture to spatially impute missing remote sensing Aerosol Optical Depth (AOD) from the Geostationary Ocean Color Imager (GOCI). The model utilized partial convolutions, which allows the convolution to only depend on the valid pixels and the normalization is adjusted to the fraction of missing data. The model has been trained with Community Multiscale Air Quality (CMAQ) simulated AOD data with masks extracted from the GOCI datasets. The partial CNN model has outperformed various models and algorithms for imputing GOCI images with a significant amount of missing data and distance to the nearest known pixel within the GOCI image. Once trained, the model required significantly less processing time and fewer resources than the other models and methods. The model allows the accurate imputation of remote sensing images within significant amounts of missing data. One limitation of the model was the lack of temporal dimensionality in the imputation process; hence the third task was implemented to address the limitation of the model.

The third task expanded the partial CNN model's capability to integrate temporal dimensionality within the datasets. The model structure was similar to the previous model in the second task but utilizes depthwise convolutions over the conventional convolutions to partially convolve the different masks independently. This change allowed the model to impute TROPOspheric Monitoring Instrument (TROPOMI) tropospheric total column NO_2 at much better quality than the default partial CNN model with the same spatio-temporal datasets. Another feature of the task was the expansion of the image augmentation process on CMAQ data to simulate the pixel variances of the TROPOMI datasets. The combination of depthwise

convolutions and image augmentation allowed the model to exceed the performances of the default partial CNN models by up to 11% and 15% in IOA and *r*, respectively. The model was also capable of imputing images with over 95% of missing pixels with sufficient accuracy, where the other algorithms and partial CNN models were not capable of such a difficult feat.

The fourth task utilized the knowledge of the previous tasks to develop a deep 2D-CNN model to forecast ENSO by up to 36 months in advance. The model incorporated various model simulation outputs from the Coupled Model Intercomparison Project phase5 (CMIP5) with various input variables reflecting the state of the climate. The variables used for the model forecasting were Sea Surface Temperature (SST) and Mean Potential Temperature anomalies, ocean current velocities, atmospheric precipitation, surface pressure, surface temperature, and atmospheric wind velocities near the surface to predict Nino3.4. Multiple models were designed and compared in their performance for forecasting Nino3.4 by up to 36 months. Two models were identified to have the best overall performance n IOA and r for forecasting Nino3.4 from 1-36 month lead times. The ensemble, which involved merging the output of the models, provided 4-7% better performance than the two best individual models. The ensemble output addressed the limitation of the individual models in specific lead months and was able to achieve a correlation of 0.52 for the 36-month forecast, indicating a moderate forecasting capability.

APPENDIX

A: Tables

Table T1: Parameters of temperature (T_{te}), relative humidity (RH_{te}), and wind speed (WS_{te}) thresholds for computing the meteorological adjustment factor (k_e). Thresholds are based on the tree, grass, and weed pollen categories. Weighting factors (c1, c2, c3) are applied to their respective meteorological thresholds.

Parameters	Vegetation		
	Tree	Grass	Weed
T_{te}	8	8	19.85
WS_{te}	2.5	2.5	2.9
RH_{te}	90	90	60
c1, c2, c3	0.5, 1.0, 2.0	0.5, 1.0, 2.0	0.5, 1.0, 2.0

Table T2: Parameters of the deep convolutional neural network for each vegetation pollen category.

Parameters	Vegetation		
	Tree	Grass	Weed
Testing Size	15%	20%	20%
Batch Size	100	100	100
Drop Out	0.4	0.1	0.2
Number of Filters	16	16	64
Hidden Units	32	64	128
Kernel Size	8	12	2
Learning Rate	0.001	0.02	0.01
Training Epochs	1000	600	500

Table T3: Parameters comparison and changes of the CNN, GRU, and DNN models. '/' designates the parameter differences between tree, grass, and weed pollen respectively. '+' designates the different layers with number of hidden units for each layer.

Parameters	Model Parameters		
	CNN	GRU	DNN
Drop Out	0.4/0.1/0.2	None	0.2
Hidden Units	32/64/128	64+32	192+128+64
Training Epochs	1000/600/500	200	1000/600/500

Table T4: Statistical performance comparison using Index of Agreement (IOA) and Pearson correlation coefficient (r) of the CNN, GRU, and DNN models in forecasting 1- and 7-day tree, grass, and weed pollen.

Model	Evaluation	Performance for 1Day Forecast			Performance for 7Day Forecast		
		Tree	Grass	Weed	Tree	Grass	Weed
CNN	IOA	0.88	0.81	0.90	0.76	0.56	0.80
	r	0.85	0.75	0.88	0.71	0.51	0.85
GRU	IOA	0.81	0.77	0.80	0.64	0.45	0.81
	r	0.73	0.68	0.85	0.69	0.48	0.80
DNN	IOA	0.88	0.81	0.79	0.72	0.54	0.78
	r	0.80	0.73	0.85	0.67	0.46	0.84

Table T5: Statistical evaluation results for the total pollen generated by the deep convolutional neural network model for the entire year of 2016. Results are based on the threshold values of respective pollen categories and the mean of the pollen thresholds and forecasting days ahead.

		Day1	Day2	Day3	Day4	Day5	Day6	Day7
Total Pollen Threshold = 20	HIT	0.631	0.646	0.646	0.651	0.641	0.656	0.626
	CSI	0.600	0.612	0.615	0.620	0.613	0.627	0.598
	FAR	0.075	0.080	0.074	0.073	0.067	0.066	0.069
	ETS	0.389	0.399	0.404	0.409	0.404	0.420	0.389
	POC	0.776	0.781	0.784	0.787	0.784	0.792	0.776
Total Pollen Threshold = 30	HIT	0.679	0.691	0.722	0.716	0.673	0.673	0.685
	CSI	0.659	0.671	0.688	0.682	0.641	0.637	0.661
	FAR	0.043	0.043	0.064	0.065	0.068	0.076	0.051
	ETS	0.509	0.523	0.538	0.531	0.484	0.478	0.510
	POC	0.844	0.850	0.855	0.852	0.833	0.831	0.844
Total Pollen Threshold = 33 (Mean)	HIT	0.684	0.696	0.709	0.709	0.684	0.684	0.684
	CSI	0.667	0.675	0.685	0.671	0.642	0.657	0.663
	FAR	0.036	0.043	0.058	0.052	0.062	0.068	0.044
	ETS	0.525	0.532	0.541	0.526	0.492	0.506	0.518
	POC	0.852	0.855	0.858	0.852	0.839	0.844	0.850
Total Pollen Threshold = 50	HIT	0.662	0.676	0.649	0.662	0.662	0.689	0.628
	CSI	0.641	0.654	0.627	0.641	0.636	0.667	0.608
	FAR	0.049	0.048	0.050	0.049	0.058	0.047	0.051
	ETS	0.506	0.521	0.492	0.506	0.500	0.535	0.471
	POC	0.850	0.855	0.844	0.850	0.847	0.861	0.836

Table T6: Image augmentation settings used for the training of the partial CNN model.

Parameter	Setting	Description
Rotation	10	Range for random rotations in degrees
Width Shift	0.1	Fraction of total width
Height Shift	0.1	Fraction of total height
Shear	0.05	Shear Intensity
Zoom	0.05	Fraction of random zoom
Horizontal Flip	True	Random horizontal flip
Vertical Flip	True	Random vertical flip
Fill mode	Reflect	Fills empty spaces after augmentation
Upscaling	Bi-cubic	Uses bicubic interpolation for upscaling

Table T7: Estimated imputation time to impute GOCI AOD dataset of various models and the platforms used to process the respective models.

Model	Hardware Details	Training Time	Imputation Time (1581 Images)	Total Time
Kriging	24 processors per node Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz	-	~ 7 hours (~15.94 seconds per image)	~ 7 hours
Partial CNN	16 processors per node Intel(R) Xeon(R) CPU E5-2603 v4 @ 1.70GHz Tesla K40c GPU	~ 7.9 hours	~ 1 minute (~0.04 seconds per image)	~ 7.9 hours
Inverse Distance Weighting	96 processors per node (Intel(R) Xeon(R) Platinum 8160 @ 2.1GHz processors (system managed by TACC)	-	~ 35 hours (~79.70 seconds per image)	~ 35 hours
Navier Stokes	24 processors per node Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz	-	~ 0.34 hours (~0.77 seconds per image)	~ 0.34 hours
Fast Marching Method	24 processors per node Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz	-	~ 0.19 hours (~0.43 seconds per image)	~ 0.19 hours
Radial basis Function	20 processors per node Intel(R) Xeon(R) CPU E5-2580 @ 2.50GHz	-	~168.3 hours (~383.23 seconds per image)	~168.3 hours
K-Nearest Neighbor	20 processors per node Intel(R) Xeon(R) CPU E5-2680 @ 2.50GHz	-	~280.1 hours (~637.80 seconds per image)	~280.1 hours

Table T8: Table summarizing the merits and demerits of the models utilized for imputing GOCI datasets.

Methods	Features	Merits	Demerits
Inverse distance weighting (IDW)	Inversed distance weighted-average within the search radius	• Good accuracy • Low bias at small distances	• No temporal variable • Overall prediction close to the mean value • Easily influenced by outliers
Kriging	Utilizing semi-variance for the spatial distribution	• Good overall performance	• No temporal variable • Overall prediction close to the mean value • Bad accuracy for distant pixels
K-Nearest Neighbor (KNN)	Nearest neighbor within the search radius	• Low bias at small distances	• No temporal variable • Low spatial variation
Radial Basis Function (RBF)	Generates functions based on the spatial distribution	• Good overall performance	• No temporal variable • Computation time is proportional to the amount of missing data • Bad accuracy for distant pixels • Requires extensive testing for optimal parameters
Navier Stokes Method (NS)	Utilizes techniques from computational fluid dynamics	• Computationally efficient • Good accuracy to performance	• No temporal variable • Original input data is slightly modified • Bad accuracy for distant pixels
Fast Marching Method (FMM)	Heuristic operation of gradually filling missing pixels from known pixels	• Computationally efficient	• No temporal variable • Low accuracy with more missing data • Original input data is slightly modified • Bad accuracy for distant pixels
Partial Convolutional Neural Network (PCNN)	Partially convolves data to reduce the significance of the mask	• Good accuracy to performance • Optimal for a high percentage of missing data • Low bias at large distances	• No temporal variable • Original input data is slightly modified • Requires prior training

Table T9: Statistical results of daily mean GOCI AOD and GOCI AOD modification of Partial CNN (PCNN) data to the AERONET measurements of seven stations in East Asia in 2018.

AERONET Station	Index of Agreement (IOA)		Pearson Correlation (*r*)		Mean Absolute Error (MAE)		Root Mean Square Error (RMSE)	
	GOCI	PCNN	GOCI	PCNN	GOCI	PCNN	GOCI	PCNN
Beijing	0.942	**0.948**	0.966	**0.973**	0.096	**0.076**	0.130	**0.112**
Fukuoka	0.749	**0.775**	0.846	**0.874**	0.112	**0.078**	0.143	**0.109**
Gangneung	0.806	**0.871**	0.874	**0.922**	0.100	**0.075**	0.150	**0.113**
Gwangju	**0.696**	0.560	**0.810**	0.737	0.132	**0.128**	**0.189**	0.203
Niigata	0.694	**0.804**	0.725	**0.859**	0.157	**0.090**	0.228	**0.136**
Osaka	0.372	**0.619**	0.526	**0.764**	0.188	**0.104**	0.249	**0.141**
Seoul	0.870	**0.881**	0.923	**0.937**	0.104	**0.085**	0.142	**0.122**

Table T10: MEA and RMSE results of the models in imputing GOCI AOD to the AERONET measurements of seven stations in East Asia in 2018.

	AERONET Station	**Kriging**	**PCNN**	**IDW**	**NS**	**FMM**	**RBF**	**KNN**
Mean Absolute Error (MAE)	Beijing	0.140	0.123	0.118	**0.109**	0.115	0.111	0.114
	Fukuoka	0.074	0.070	0.074	0.070	**0.069**	0.084	0.092
	Gangneung	0.099	**0.087**	0.090	0.098	0.108	0.110	0.106
	Gwangju	0.099	0.085	**0.075**	0.084	0.085	0.085	0.087
	Niigata	0.087	**0.073**	0.084	0.086	0.093	0.104	0.126
	Osaka	0.096	**0.091**	0.098	0.104	0.098	0.118	0.127
	Seoul	0.124	0.121	0.111	0.109	0.119	**0.108**	0.114
Root Mean Square Error (RMSE)	Beijing	0.252	0.241	0.232	0.213	0.220	**0.194**	0.214
	Fukuoka	0.111	0.107	0.108	0.100	**0.098**	0.125	0.133
	Gangneung	0.158	0.166	**0.154**	0.165	0.179	0.189	0.180
	Gwangju	0.138	0.116	**0.103**	0.120	0.118	0.113	0.121
	Niigata	0.129	**0.104**	0.125	0.127	0.148	0.142	0.189
	Osaka	0.167	**0.139**	0.147	0.152	0.140	0.171	0.177
	Seoul	0.211	0.198	0.187	0.183	0.189	**0.173**	0.184

Table T11: Image augmentation settings used for the training of the partial CNN model.

Parameter	**Setting**	**Description**
Random Noise	Mean=0, Deviation=1-2% of max NO_2 variable per image	Applies random noise mask based on set deviation range with a 0 mean distribution
Horizontal Flip	True	Random horizontal flip
Vertical Flip	True	Random vertical flip

Rotation	90-degree steps	Rotates images at 90-degree intervals

Table T12: Table showing the statistical evaluation results of imputing TROPOMI images by the Inverse Distance Weighting (IDW), coupled IDW with CoKriging (IDW CoKriging), regular Partial Convolutional Neural Network (Base PCNN), Base PCNN with Spatio- and temporal-datasets (ST), and Depthwise PCNN (DW-PCNN) models. The evaluations are based on the Index of Agreement (IOA), correlation (*r*), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Models	**2019**				**2020**			
	IOA	***r***	**MAE**	**RMSE**	**IOA**	***r***	**MAE**	**RMSE**
IDW	0.77	0.68	6.23×10^{14}	9.15×10^{14}	0.77	0.68	5.02×10^{14}	6.98×10^{14}
IDW CoKriging	0.77	0.68	6.13×10^{14}	9.00×10^{14}	-	-	-	-
Base PCNN	0.76	0.64	6.40×10^{14}	9.76×10^{14}	0.79	0.68	4.89×10^{14}	7.02×10^{14}
Base PCNN (ST)	0.76	0.66	**5.77×10^{14}**	**8.45×10^{14}**	0.74	0.65	4.85×10^{14}	6.65×10^{14}
DW-PCNN	**0.81**	**0.74**	6.05×10^{14}	8.64×10^{14}	**0.82**	**0.74**	**4.55×10^{14}**	**6.21×10^{14}**

Table T13: Table summarizing the merits and demerits of the models utilized for imputing GOCI datasets.

Methods	**Features**	**Merits**	**Demerits**
Inverse distance weighting (IDW)	Inversed distance weighted-average within the search radius	• Good accuracy • Optimal for a low percentage of missing data • Low bias at small distances	• No temporal variable • Overall prediction close to the mean value • Easily influenced by outliers
IDW - CoKriging	Utilizing weekly mean imputed images as an additional variable for Kriging imputation	• Good accuracy • Optimal for a low percentage of missing data • Low bias at small distances • Incorporates semi-temporal variable	• No direct temporal variable • Overall prediction is the same as IDW • Increased computation and processing time
Partial Convolutional Neural Network (PCNN)	Partially convolves data to reduce the significance of the mask	• Optimal for a low percentage of missing data • Low bias at large distances	• No temporal variable • Original input data is modified • Requires prior training • Distortions with images with high % missing data

Partial Convolutional Neural Network with Spatio-Temporal Data (PCNN-ST)	Partially convolves data to reduce the significance of the mask. Receives different images for each channel	• Good accuracy to performance • Optimal for a high percentage of missing data • Low bias at large distances	• Original input data is modified • Requires prior training • Significant smoothing in the output and slight over-prediction • Mask features are not properly padded during the encoding phases
Depthwise Partial Convolutional Neural Network (DW-PCNN)	Partially convolves data to reduce the significance of the mask with depthwise convolutions	• Very good accuracy to performance • Optimal for all cases • Low bias at large distances • Mask features are maintained for each channel	• Original input data is slightly modified • Requires prior training • Requires longer training than regular PCNN models

Table T14: List of defined maximum and minimum variables for the data normalization process of the input data.

Parameter	SSTA & MPTA	UO & VO	PR	PSL	TS	UA & VA
Max	10	2.5	0.007	10500	320	14
Min	-13	-2	-0.00004	97000	220	-12

Table T15: List of CMIP5 model IDs used for training the Deep CNN models for forecasting Nino3.4.

CMIP ID	Modeling Group	Model Time Period
CanESM2	Canadian Centre for Climate Modelling and Analysis	1850/01-2005/12
CMCC_CM	**Centro Euro-Mediterraneo per I Cambiamenti Climatici,**	1850/01-2005/12
CMCC_CMS		1850/01-2005/12
HadGEM2-AO	National Institute of Meteorological Research & Korea Meteorological Administration	1860/01-2005/12
HadGEM2-CC	Met Office Hadley Centre	1859/12-2005/11

HadGEM2-ES		1859/12-2005/12
IPSL-CM5A-MR	Institut Pierre-Simon Laplace	1850/01-2005/12
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	1850/01-1999/12

Table T16: Table showing the statistical scores, Index of Agreement (IOA) and Correlation Coefficient (*r*), of the Deep CNN models in forecasting Nino3.4 SST anomaly between 6 and 36 months at 6-month intervals. The ensemble is the merging of the 'C8/8/8 Swish' and 'C8m16m32 Swish' model outputs. 'C' designates the convolution layers with the number referring to the number of filters within the convolution layer. 'm' designates a max-pooling layer between the convolution layers and Swish (ReLU) are the activation functions used for the convolution layers.

	Forecast Lead	C8/8/8 Swish	C8m8m8 swish	C8/16/32 Swish	C8m16m32 Swish	C8/8/8 Relu	C8m8m8 Relu	C8/16/32 Relu	C8m16m32 Relu	Ensemble
Index of Agreement (IOA)	**6 Month**	0.76	0.72	0.75	**0.83**	0.73	0.63	0.78	0.69	0.81
	12 Month	0.62	0.54	0.59	**0.68**	0.37	0.54	0.67	0.51	0.67
	18 Month	**0.63**	0.28	**0.63**	0.55	0.36	0.48	0.54	0.40	0.61
	24 Month	0.49	0.41	**0.54**	0.50	0.36	0.67	0.52	0.46	0.51
	30 Month	0.60	0.42	0.49	0.63	0.44	0.56	0.44	0.49	**0.65**
	36 Month	**0.61**	0.54	0.44	0.55	0.51	0.46	0.49	0.14	0.60
	Cum. Score	3.70	2.91	3.45	3.74	2.77	3.34	3.43	2.70	**3.85**
Correlation Coefficient (*r*)	**6 Month**	0.69	0.62	0.68	0.72	0.65	0.43	0.71	0.50	**0.75**
	12 Month	0.46	0.42	0.45	0.52	0.09	0.45	0.51	0.45	**0.58**
	18 Month	**0.59**	-0.05	0.58	0.41	0.13	0.40	0.28	0.15	0.56
	24 Month	0.34	0.11	0.37	0.28	0.08	**0.47**	0.31	0.19	0.34
	30 Month	0.46	0.34	0.33	0.42	0.16	0.32	0.16	0.35	**0.48**
	36 Month	0.48	0.38	0.19	0.40	0.27	0.10	0.37	-0.41	**0.52**
	Cum. Score	3.02	1.83	2.61	2.75	1.38	2.17	2.33	1.23	**3.23**

B: Figures

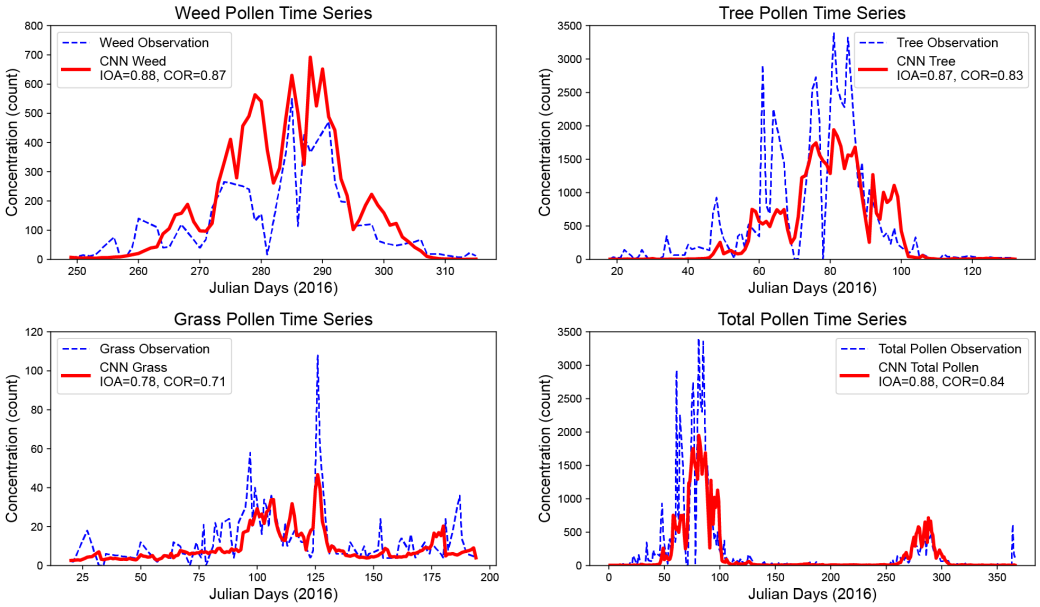

Figure F1: Season time series of the deep convolutional neural network (CNN) model forecasting of Weed, Tree, and Grass pollen concentrations one day ahead for 2016. The time series of total pollen is for the entire year of 2016. Blue indicates the observed pollen concentration, and red indicates forecasted pollen concentration. Respective statistical scores are based on the respective seasons during the year 2016.

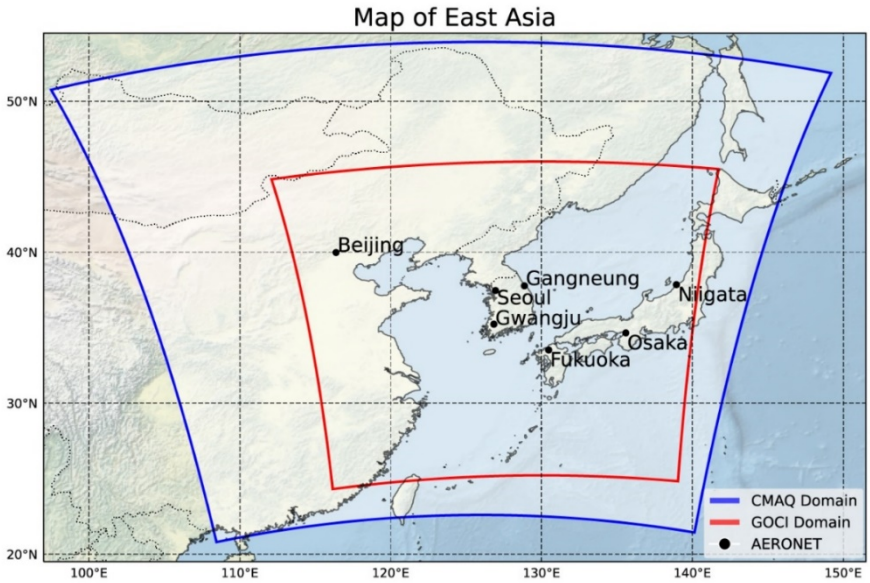

Figure F2: Map of East Asia with the domains for the CMAQ model (blue) used for training the partial CNN model and GOCI images (red) for evaluations of both the partial CNN model and the Kriging method. Black dots indicate the locations and city names of the AERONET stations used to evaluate the imputation accuracy of GOCI images.

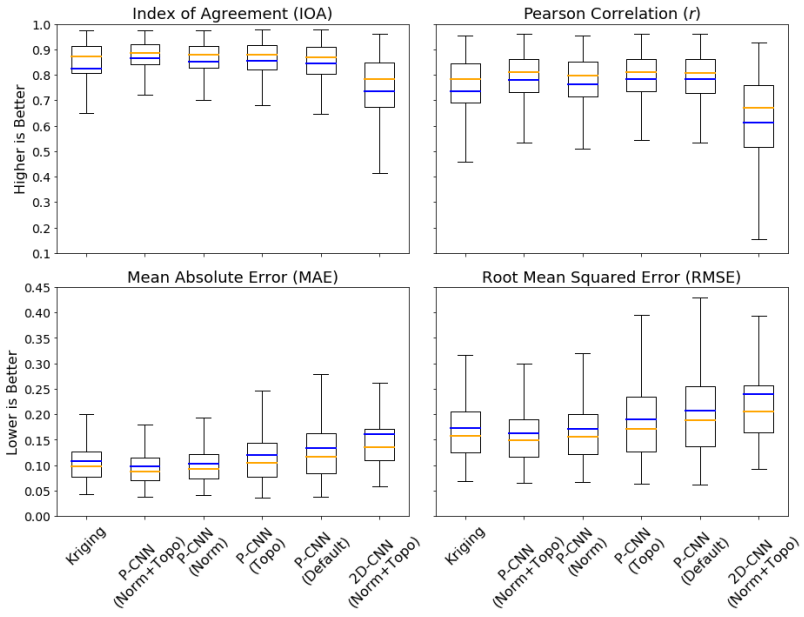

Figure F3: Boxplots showing the performance of various Partial CNN models based on including or excluding globally defined normalization (Norm) and/or topology layer (Topo) within a color channel of the masked image. As a reference, the Kriging model and convolutional neural network (2D-CNN) are shown.

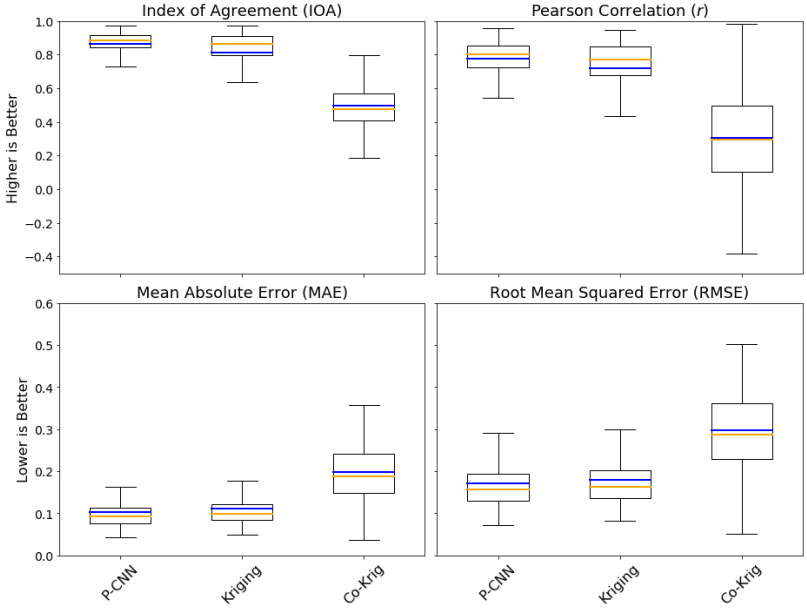

Figure F4: Boxplot indicating the statistical performance comparison of Co-Kriging, ordinary Kriging, and Partial CNN (P-CNN) models.

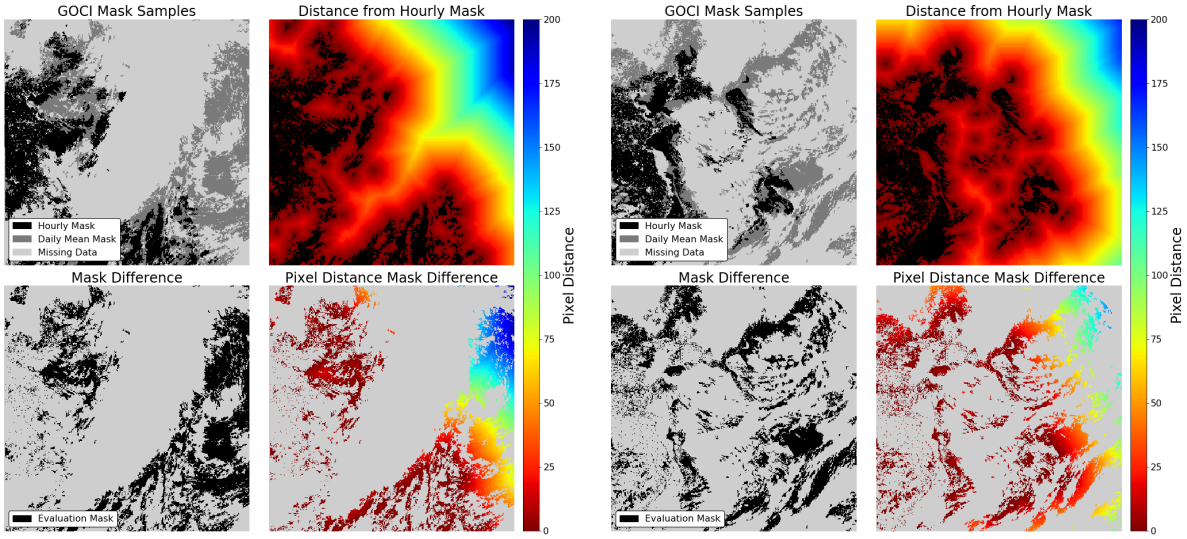

Figure F5: Demonstration of the method for evaluating the pixel distance. Hourly and daily mask differences are used for evaluating the performance of the models for imputing missing GOCI data. Pixel distances are estimated based on the distance from the available hourly mask. Displaying imputation performance, the evaluation mask is overlaid on the pixel distance from the hourly mask, which represents the pixels.

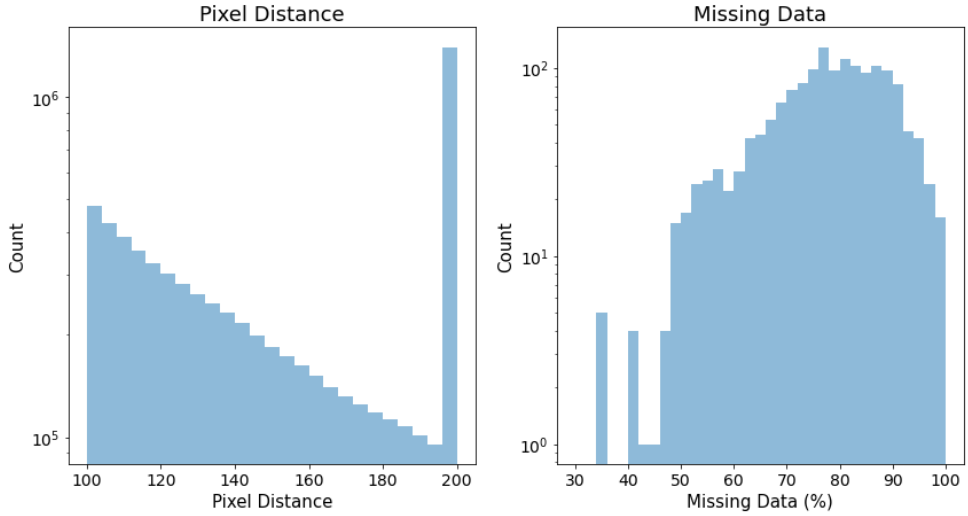

Figure F6: Histograms showing the distribution from 100 to 200 or more pixel distances (left) and percent of missing data (right) within the GOCI image.

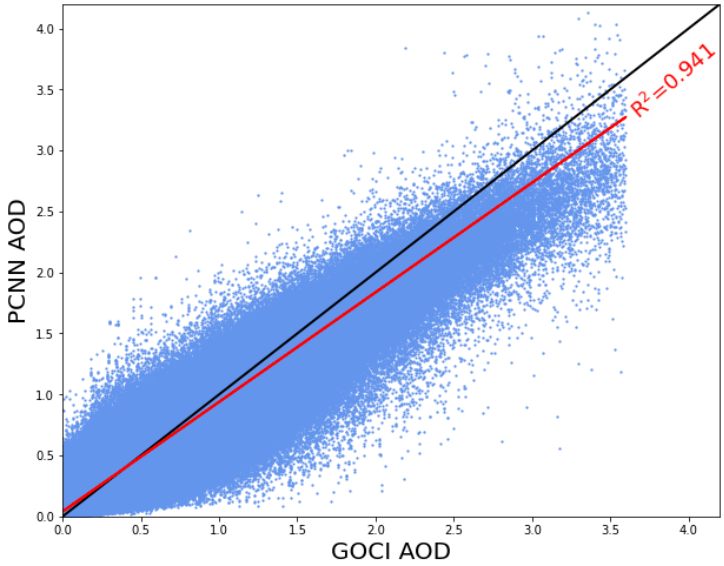

Figure F7: Scatterplot of Partial CNN's (PCNN) modification of measured GOCI AOD. The line of best fit (red) achieved a coefficient of determination (R^2) of 0.941.

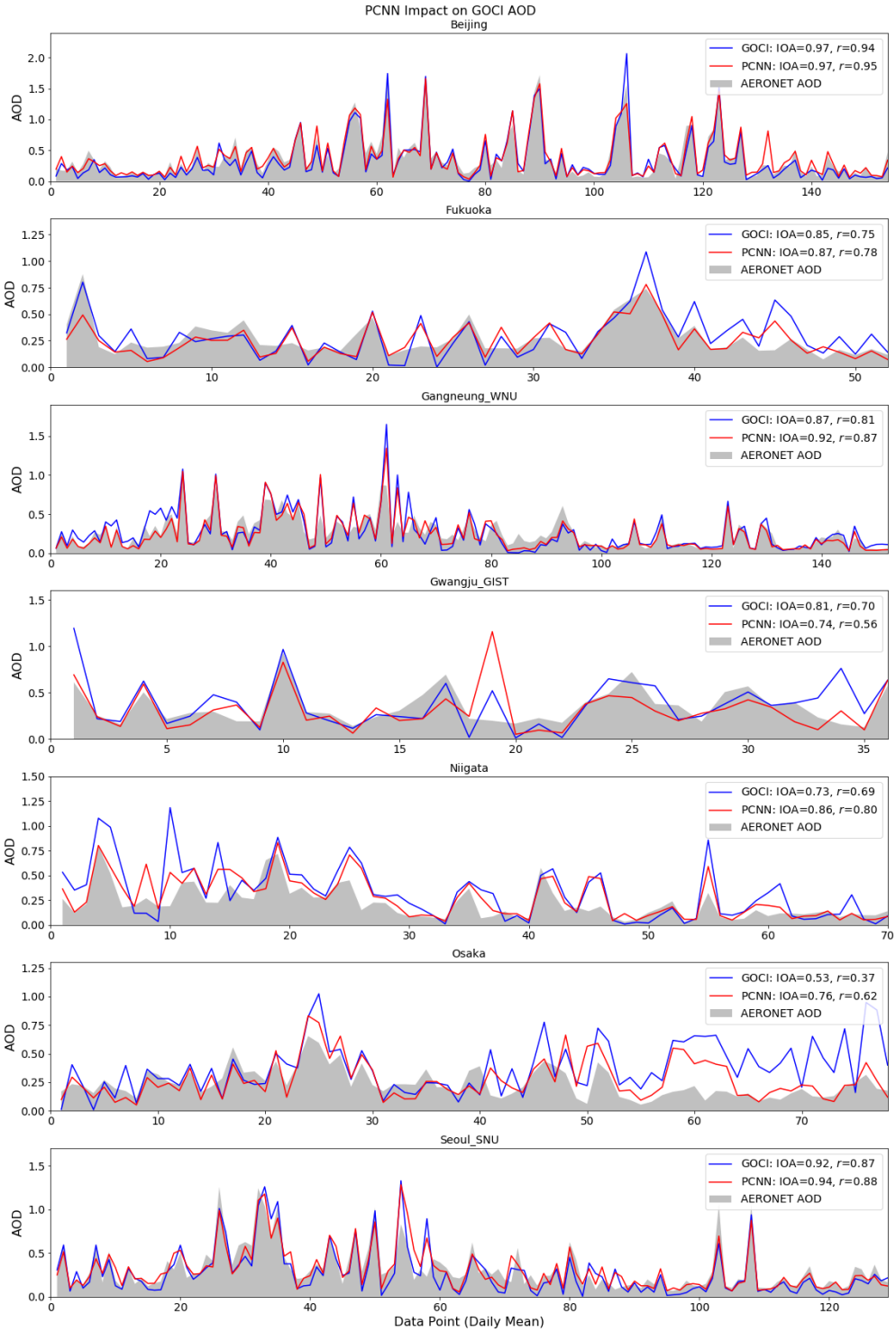

Figure F8: Line plots of daily mean GOCI AOD (blue) and Partial CNN modification of GOCI AOD (red) to the seven AERONET station AOD estimations (grey) for 2018.

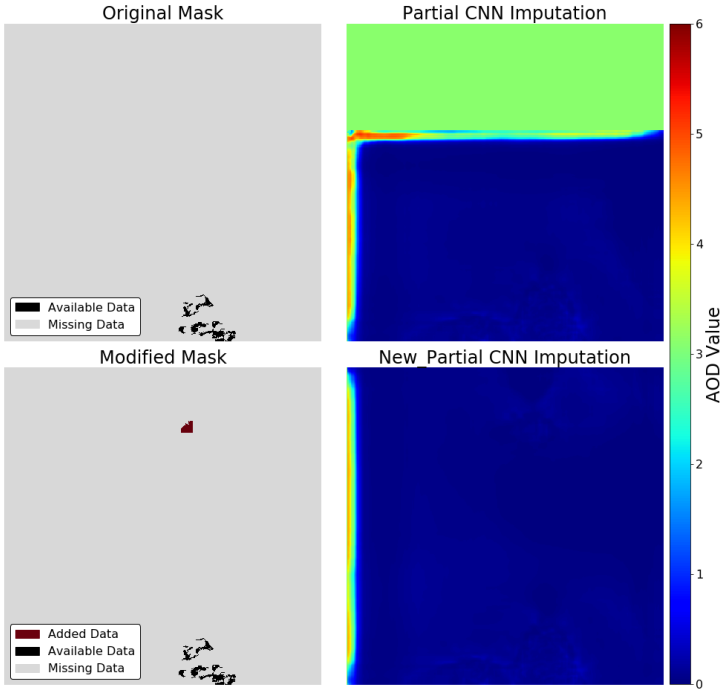

Figure F9: Demonstration of the partial CNN model unable to properly estimate the GOCI image (top). This phenomenon is due to the clustering of available data (black) on one side of the image. A modified mask demonstrates the application of sample data (red) to the image for imputation. The partial CNN model bias issue has been partially resolved. For optimal performance, available data need to be distributed throughout the image.

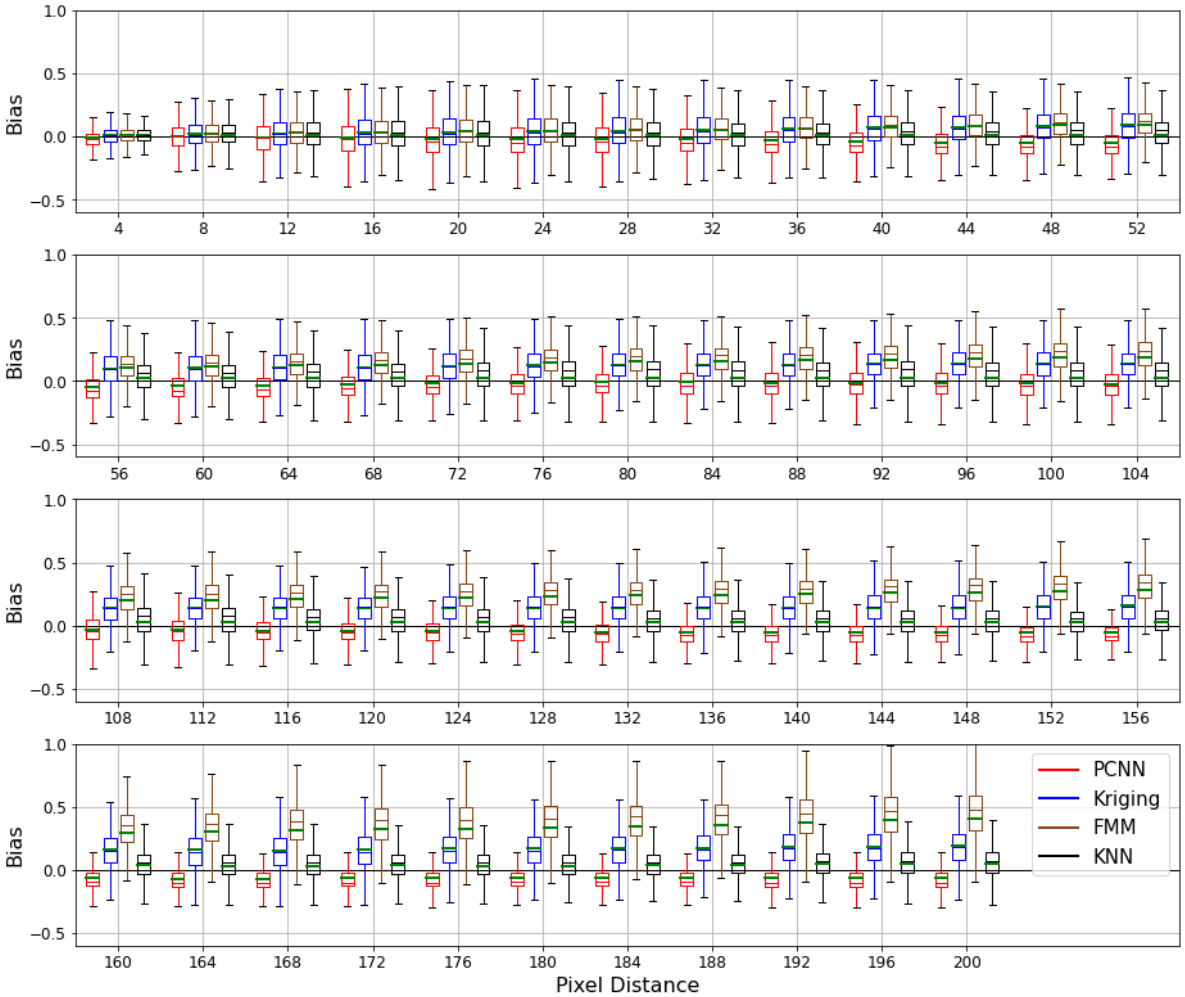

Figure F10: Results of the bias-variance among pixel distances of the Partial CNN (PCNN), Kriging, Fast Marching Method (FMM), and K-Nearest Neighbor (KNN) models. The distances are split into four sections at 52-pixel intervals. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The horizontal lines represent the mean bias of PCNN (red), IDW (blue), NS (orange), RBF (black) models, and the median bias (green for all models) of the imputing missing GOCI data. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range.

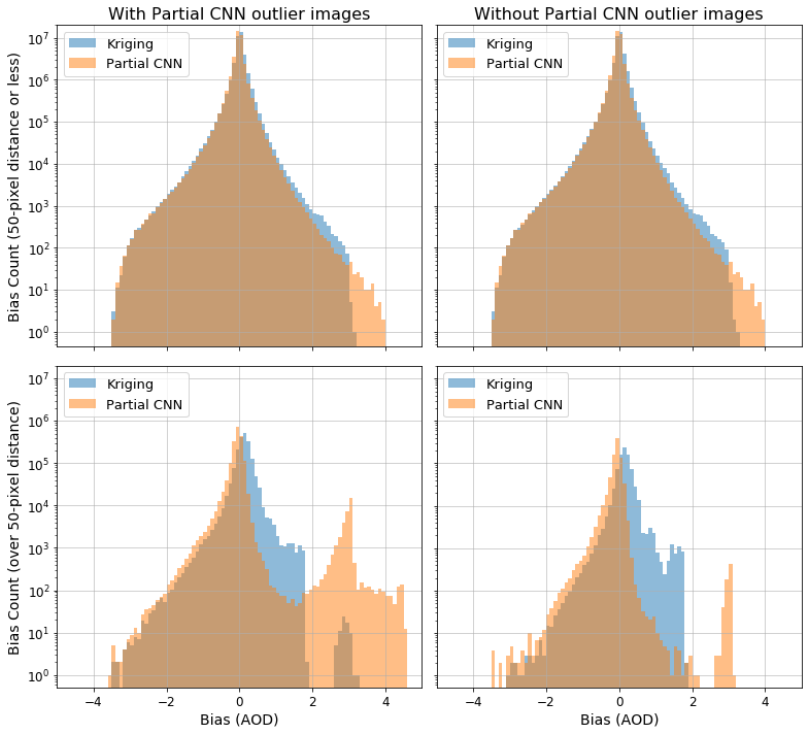

Figure F11: Histograms of the partial CNN and Kriging methods at distances of 50 pixels or less and over 50 pixels. Histograms in the left column include the three cases of high bias estimation from the partial CNN model. Histograms in the right column exclude the three cases of high bias from the partial CNN model.

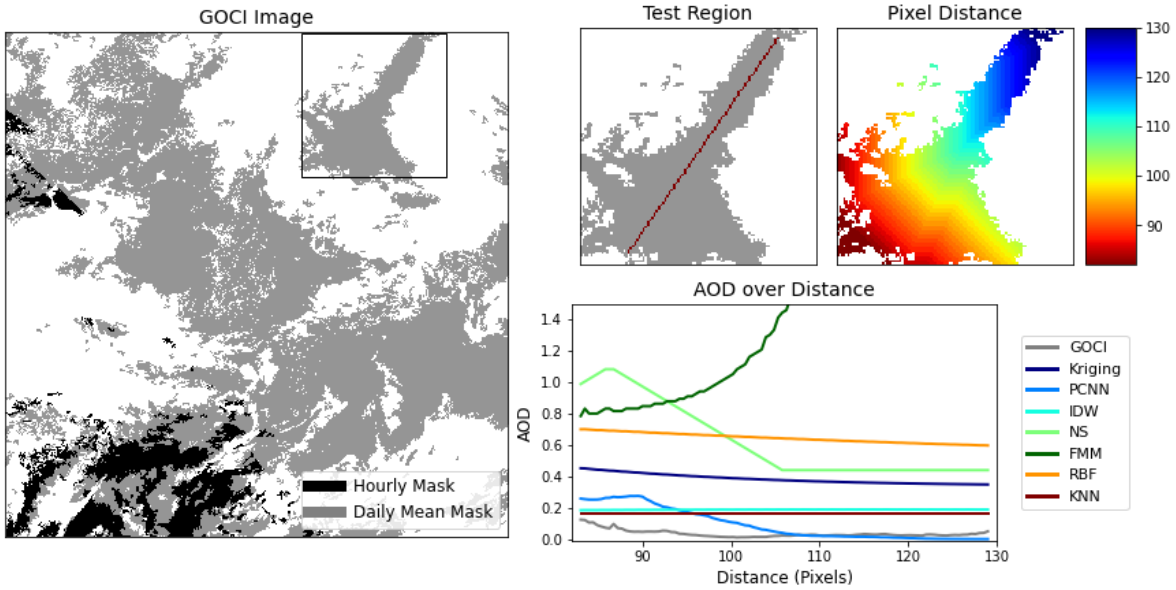

Figure F12: Distance AOD imputation comparison of the imputation models in estimating the daily mean GOCI image based on the GOCI hour mask of the respective day.

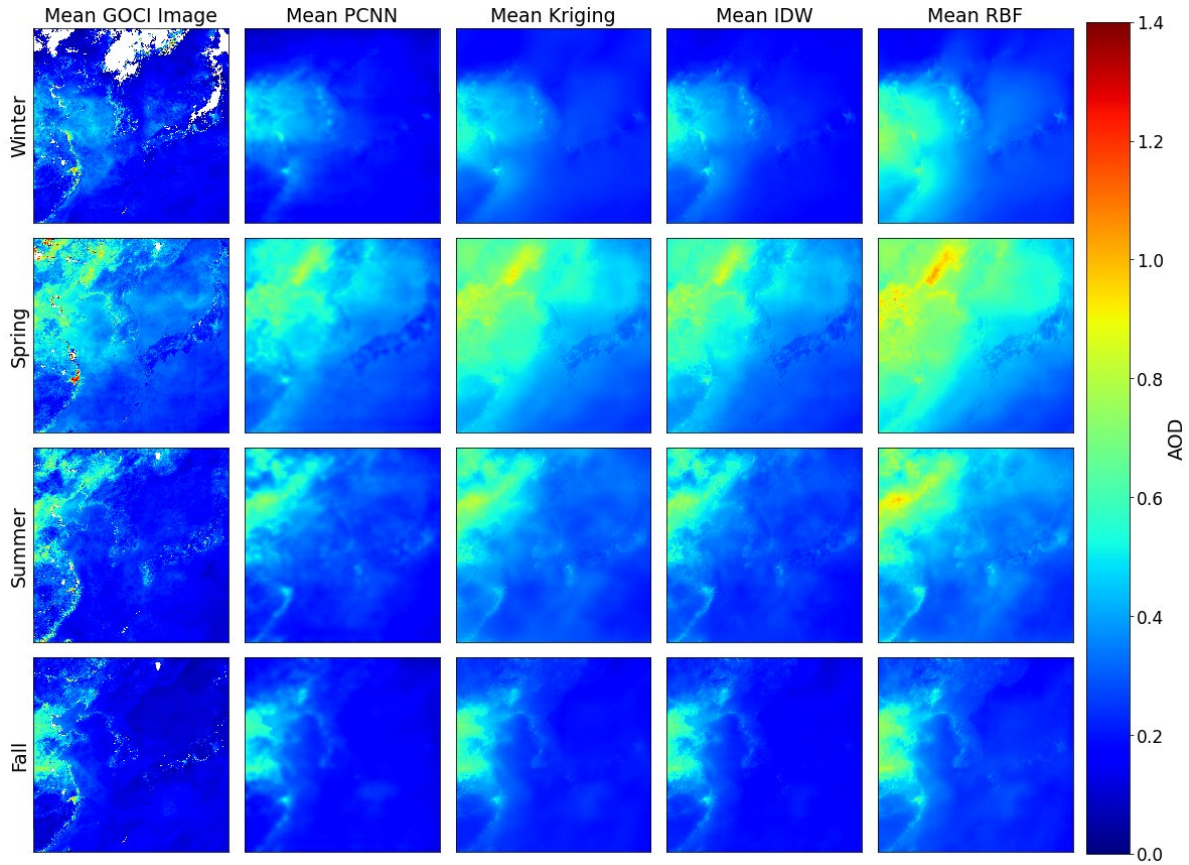

Figure F13: Images of seasonal means of GOCI AOD, Partial CNN (PCNN), Kriging, Inverse Distance Weighting (IDW), and Radial Basis Function (RBF). The seasons are divided into Winter (DJF), Spring (MAM), Summer (JJA), and Fall (SON).

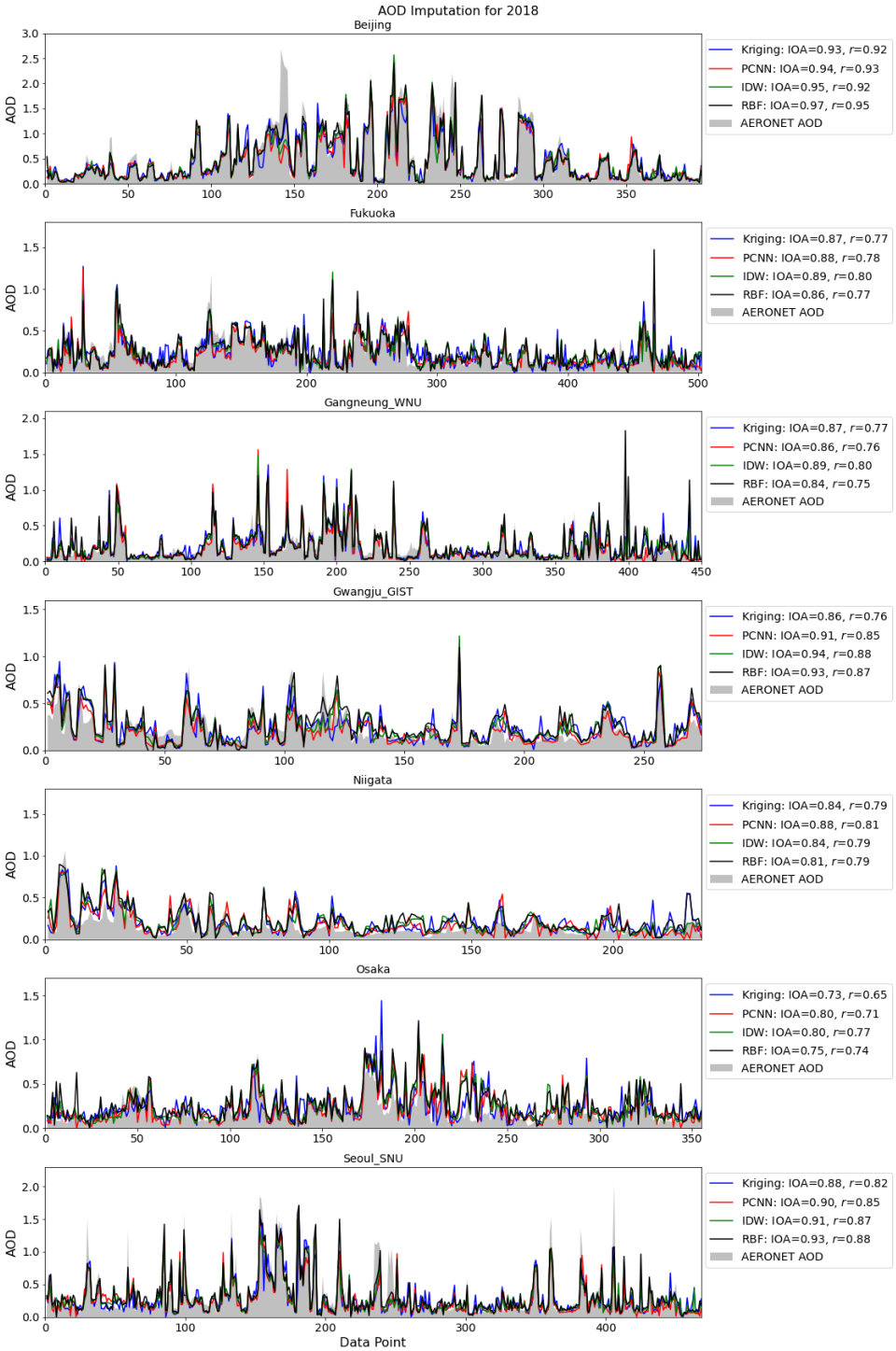

Figure F14: Line plots of imputed Kriging AOD (blue), Partial CNN AOD (red), Inverse Distance Weighting (green), and Radial Basis Function (black) to the seven AERONET station AOD estimations (grey) for 2018 where GOCI AOD measurements were not available.

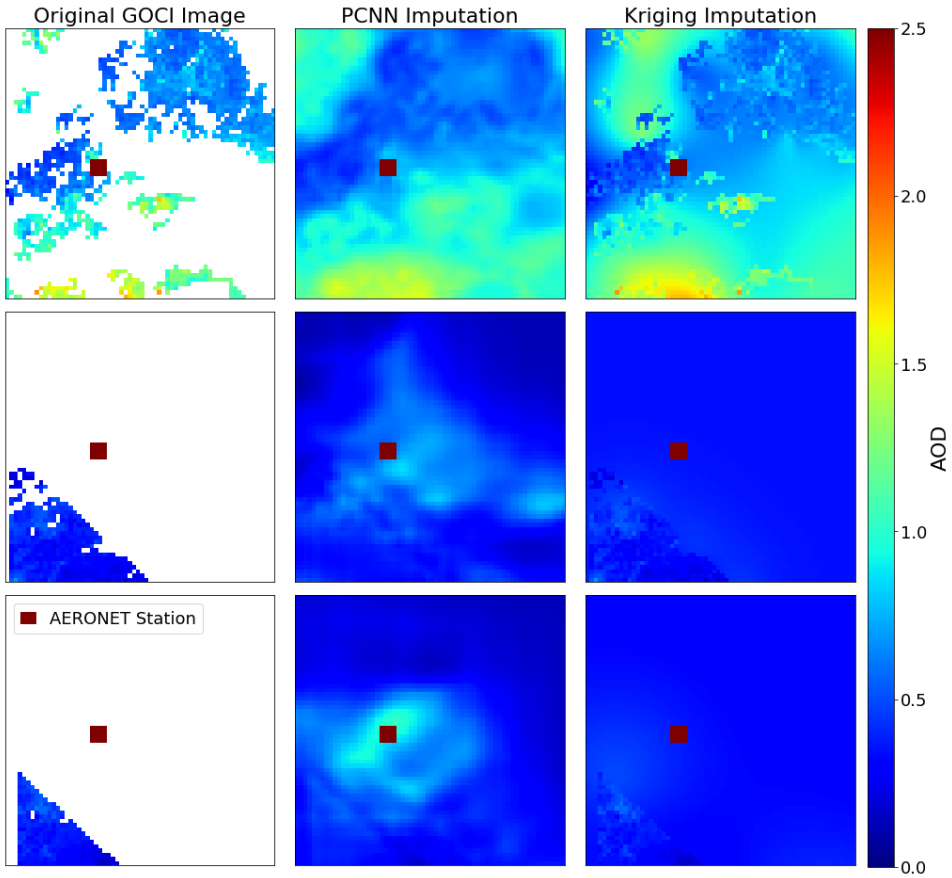

Figure F15: Sample cases of the bias of the partial CNN model in significantly overestimating AOD levels at the Beijing AERONET station.

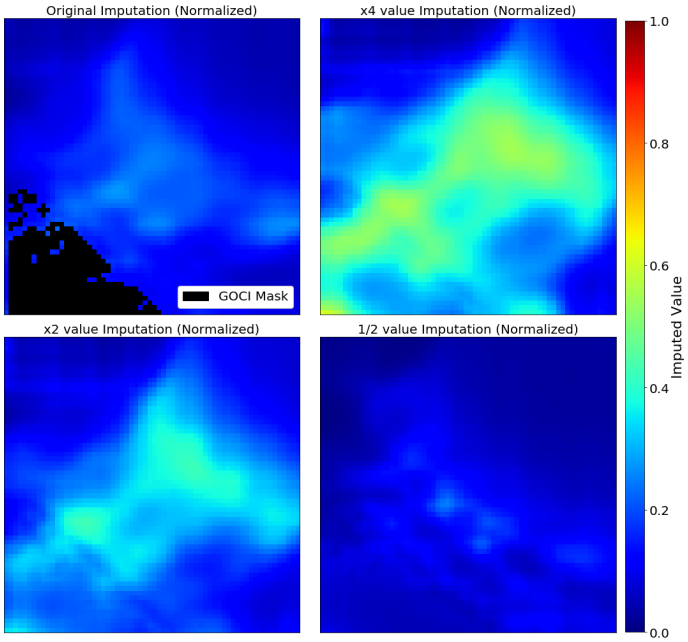

Figure F16: Partial CNN imputation impact based on initial values within the GOCI mask of the Beijing station region. Imputed results are normalized within a range of 0-1 and do not directly represent AOD. Values are based on the modified AOD values across the mask region prior to imputation.

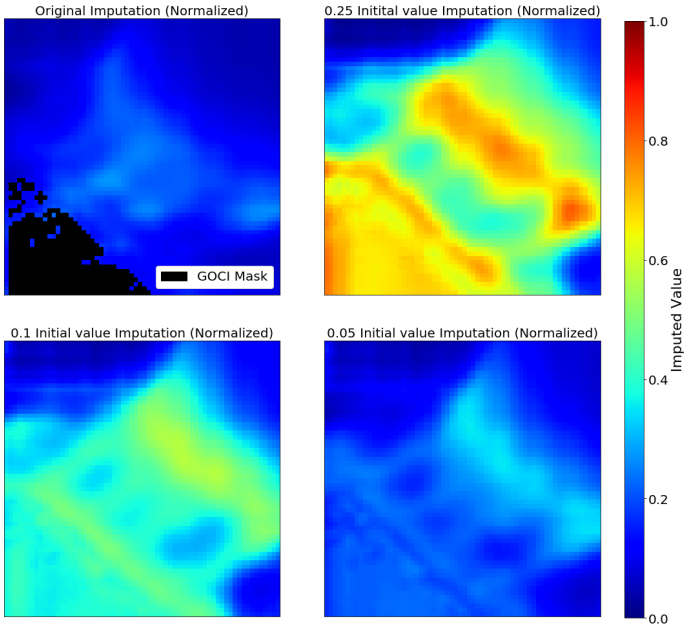

Figure F17: Partial CNN imputation impact based on initial values within the GOCI mask of the Beijing station region. Imputed results are normalized within a range of 0-1 and do not directly represent AOD. Values are based on a consistent value across the mask region prior to imputation.

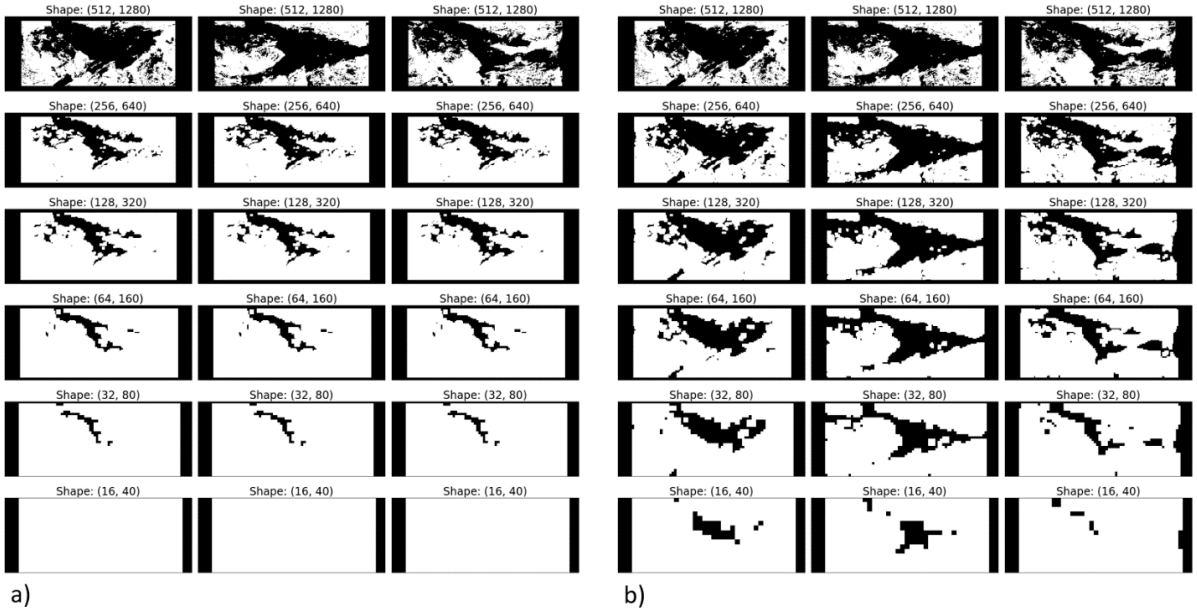

Figure F18: Mask padding comparison of the regular convolution (a) and depthwise convolution (b) padding. The columns represent one gray scale image within the three color channels (Red, Green, Blue) within a digital image. The regular convolution padding convolves the different masks as one unified mask after the first encoding layer. The depthwise convolution maintains the different mask features due to the padding occurring for each channel individually.

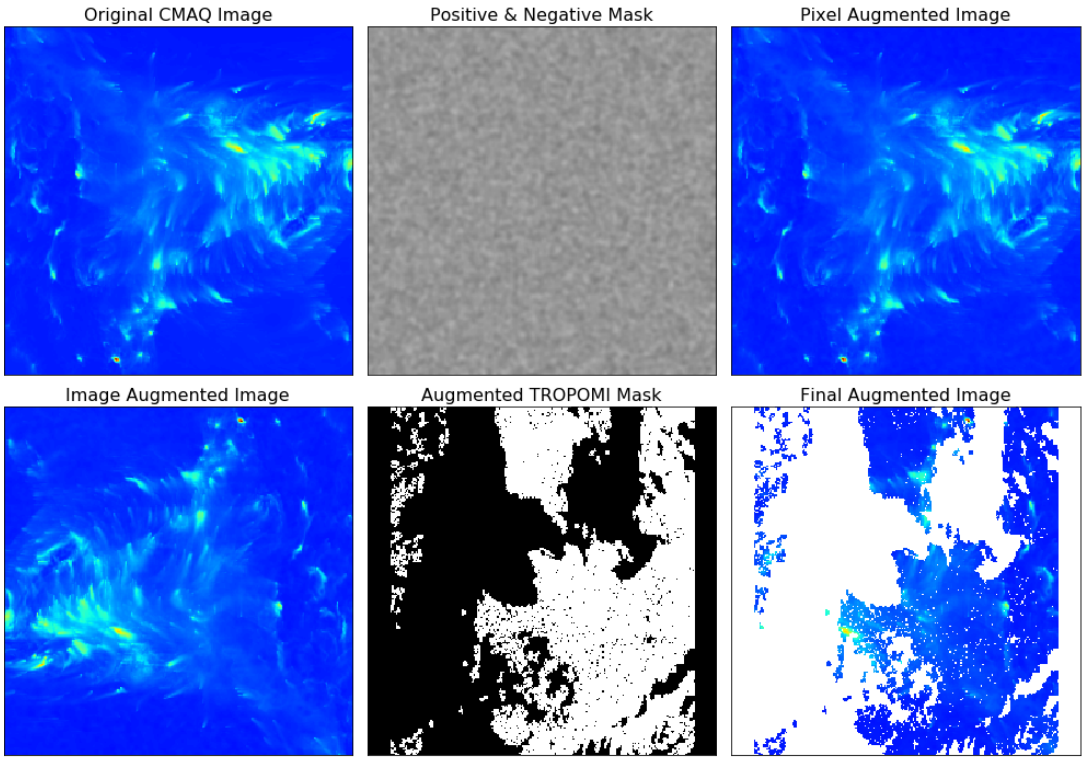

Figure F19: Sample of image augmentation process on CMAQ images performed for the training of the DW-PCNN model. The first phase of augmentation adds random positive and negative features (Positive & Negative Mask) to simulate the pixel variations within the TROPOMI remote sensing images. Basic image augmentation parameters (such as flip and rotate) to the image are then applied with a randomly selected mask extracted from TROPOMI 2019 datasets. The mask is then applied to the augmented image and used as input for the model to impute.

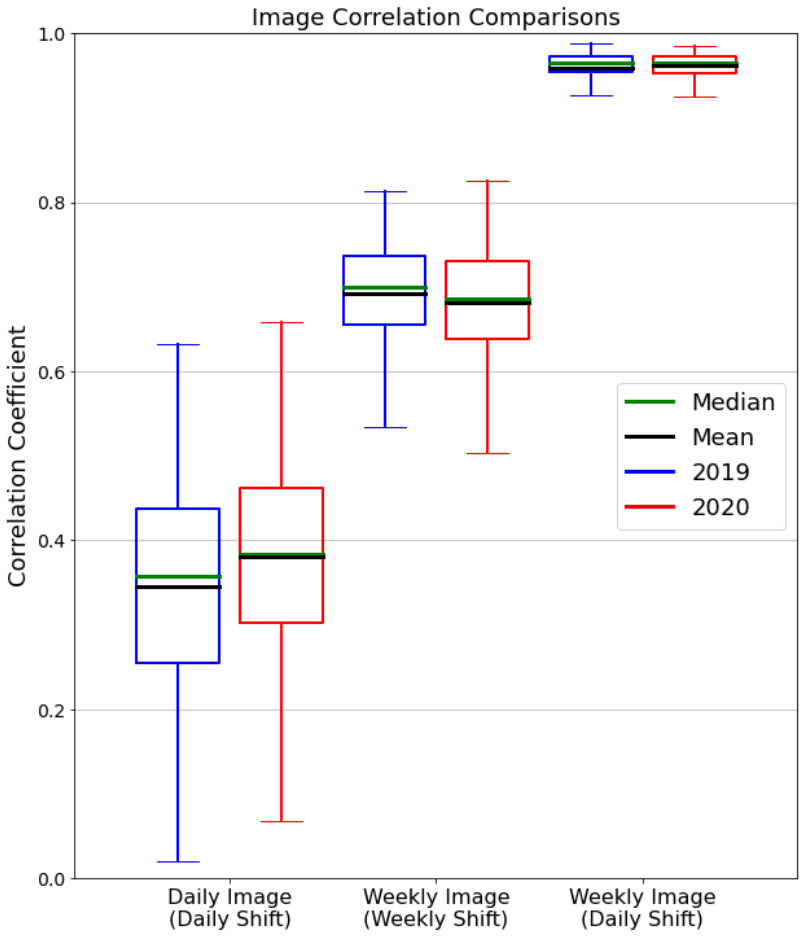

Figure F20: Boxplot showing the correlation coefficient between pixels within the TROPOMI tropospheric column NO_2 images with a daily shift, the weekly average with weekly shift, and the weekly average with a daily shift. The main section of the boxplot presents the interquartile range between the 25th and 75th percentiles. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range.

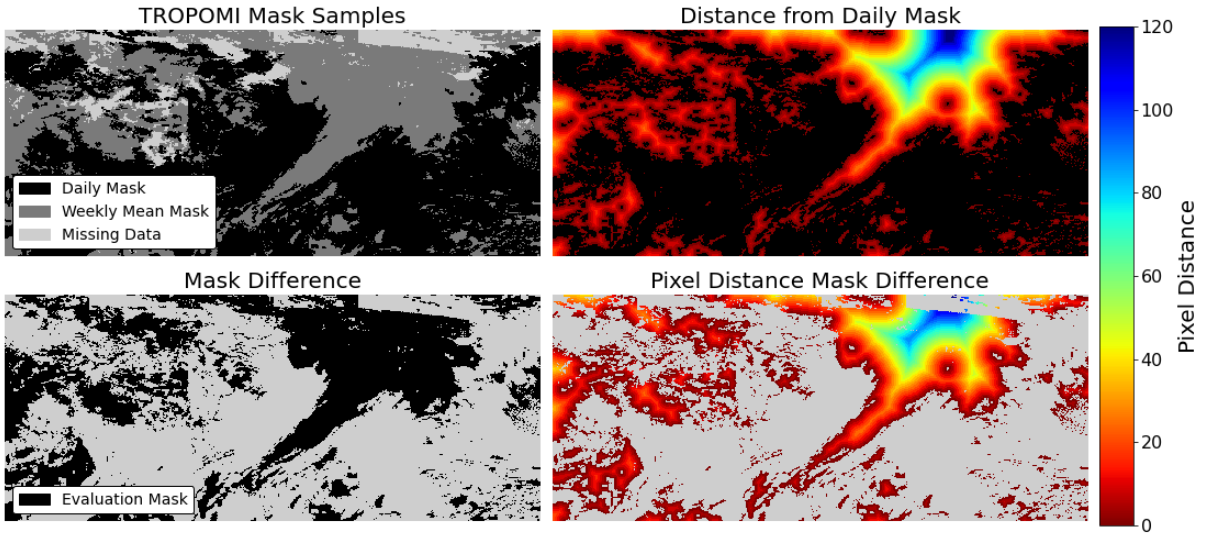

Figure F21: Demonstration of the method for evaluating the pixel distance. Daily and weekly mask differences are used for evaluating the performance of the models for imputing missing TROPOMI data. Pixel distances are estimated based on the euclidean distance from the available daily mask. For evaluating imputation performance, the evaluation mask is overlaid on the pixel distance from the hourly mask, which represents the pixel distances.

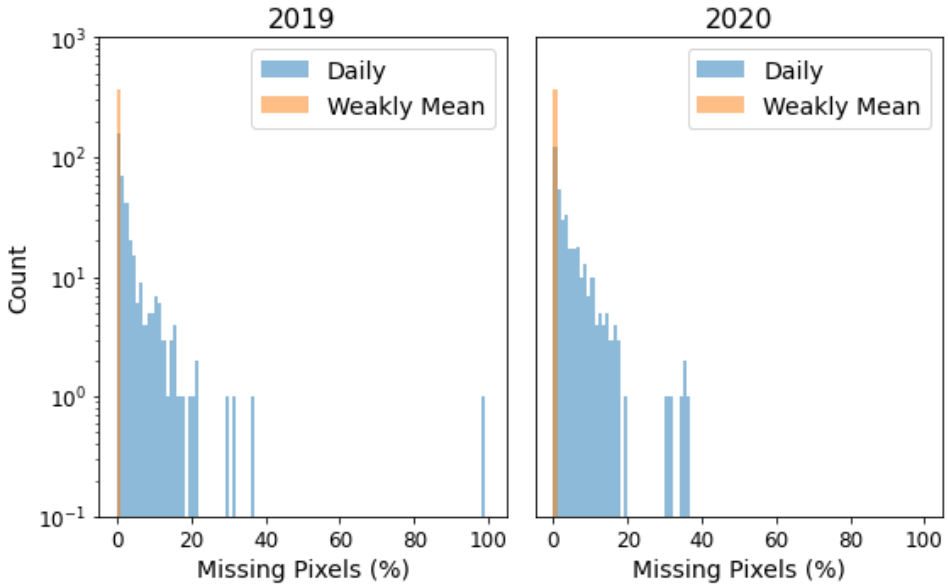

Figure F22: Histograms showing the distribution of percent of missing data within the TROPOMI images for 2019 (left) and 2020 (right).

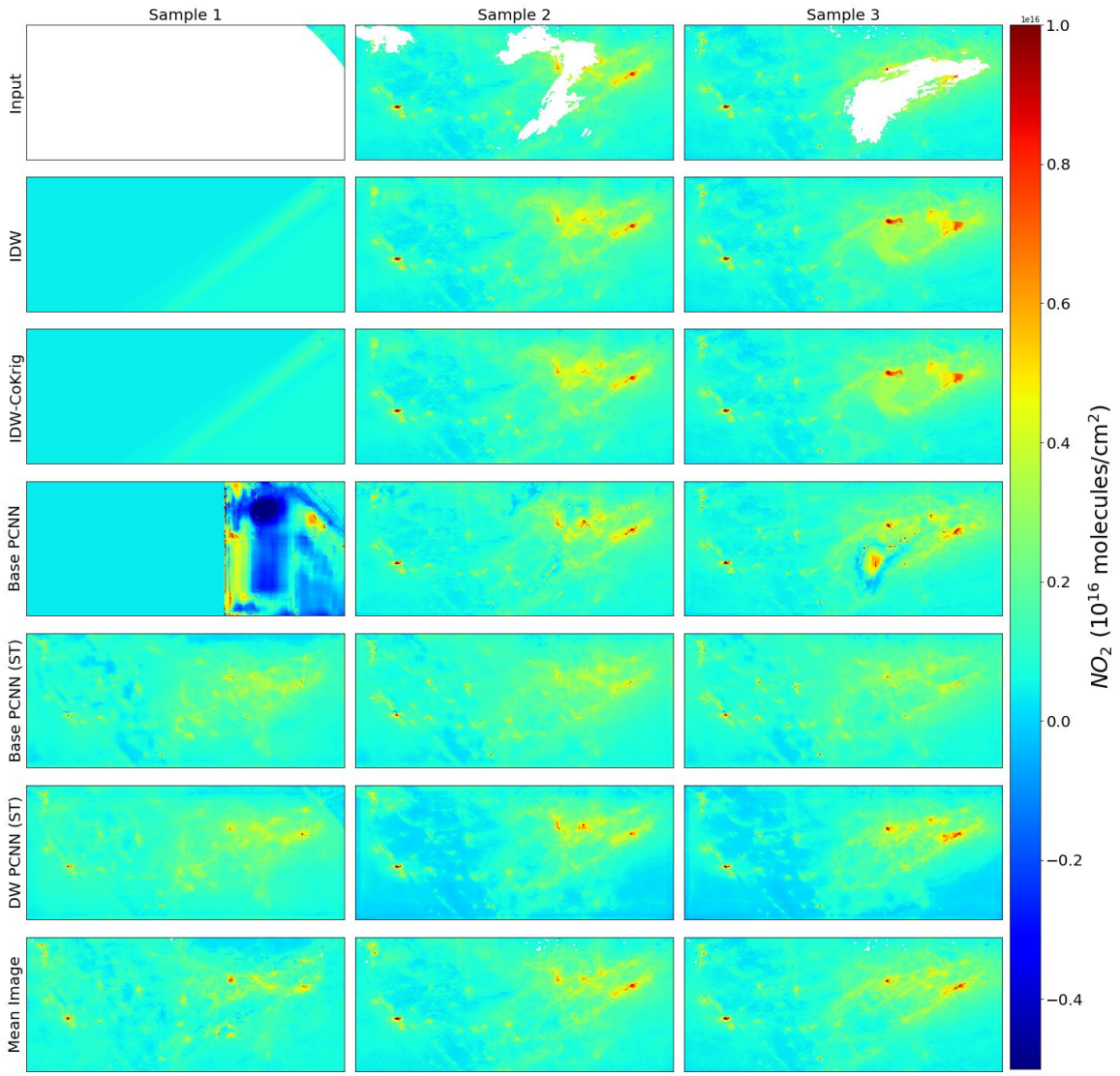

Figure F23: Sample cases of models imputing weekly averaged images with daily masks applied from TROPOMI 2019.

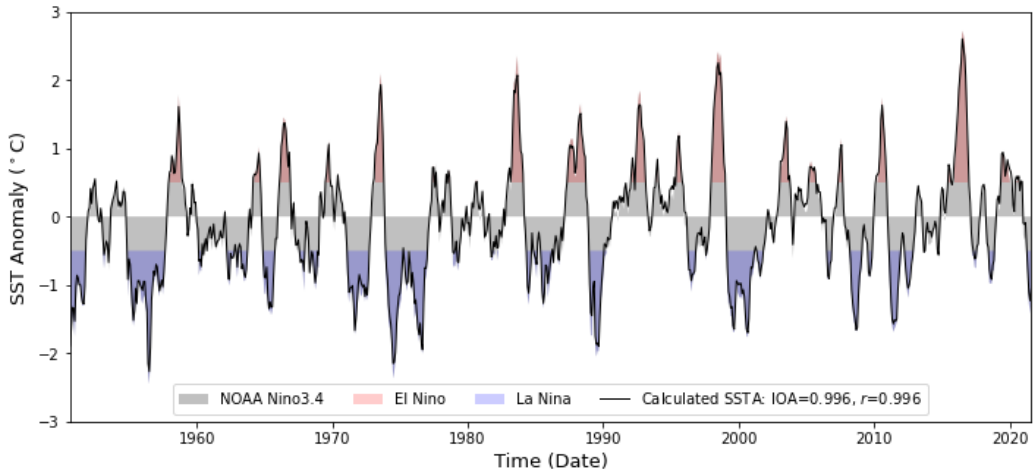

Figure F24: Timeseries comparing the NOAA Nino3.4 index (NOAA Nino3.4) to the calculated Sea Surface Temperature (SST) Anomaly from the NOAA Extended Reconstructed Sea Surface Temperature V5 (ERSSTv5) spatial SST dataset (Calculated SSTA). The El Nino and La Nina phenomena are shaded in red and blue, respectively.

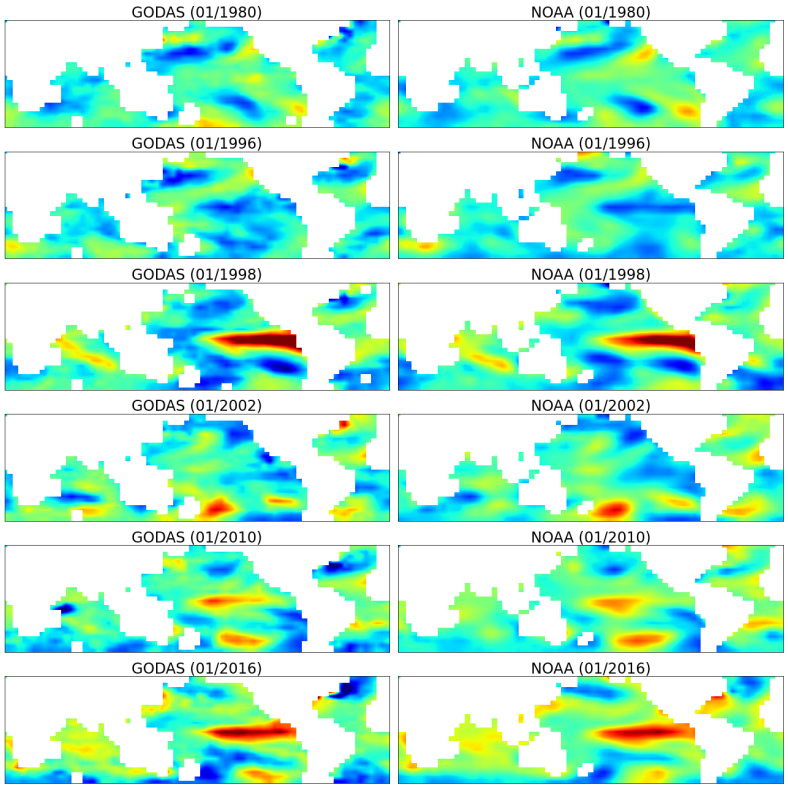

Figure F25: Spatial comparisons of NOAA spatial Sea Surface Temperature Anomaly (SSTA) and computed spatial SSTA from GODAS spatial datasets.

BIBLIOGRAPHY

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265-283).

Angosto, J. M., Moreno-Grau, S., Bayo, J., & Elvira-Rendueles, B. (2005). Multiple regression models for predicting total daily pollen concentration in Cartagena. *Grana*, *44*(2), 108-114.

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE transactions on medical imaging*, *35*(5), 1207-1216.

Bae, B., Kim, H., Lim, H., Liu, Y., Han, L. D., & Freeze, P. B. (2018). Missing data imputation for traffic flow speed using spatio-temporal cokriging. *Transportation Research Part C: Emerging Technologies*, *88*, 124-139.

BANKEN, R., & COMTOIS, P. (1990). Concentration du pollen de l'herbe à poux et prévalence de la rhinite allergique dans deux municipalités des Laurentides. *Union médicale du Canada*, *119*(4), 178-181.

Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Transactions on Energy Conversion*, 21(1), 273-284.

Bartková-Ščevková, J. (2003). The influence of temperature, relative humidity and rainfall on the occurrence of pollen allergens (Betula, Poaceae, Ambrosia artemisiifolia) in the atmosphere of Bratislava (Slovakia). *International Journal of Biometeorology*, *48*(1), 1-5.

Bastl, K., Berger, U., & Kmenta, M. (2017). Evaluation of pollen apps forecasts: the need for quality control in an eHealth service. *Journal of medical Internet research*, *19*(5), e152.

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1-4). Springer, Berlin, Heidelberg.

Bengio, Y. (2009). *Learning deep architectures for AI*. Now Publishers Inc.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

Bertalmio, M., Sapiro, G., Caselles, V., & Ballester, C. (2000, July). Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (pp. 417-424).

Bertalmio, M., Bertozzi, A. L., & Sapiro, G. (2001, December). Navier-stokes, fluid dynamics, and image and video inpainting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Vol. 1, pp. I-I). IEEE.

Bjorck, J., Gomes, C., Selman, B., & Weinberger, K. Q. (2018). Understanding batch normalization. *arXiv preprint arXiv:1806.02375*.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., ... Saide, P. E. (2015). Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmospheric chemistry and physics*, 15(10), 5325-5358.

Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The lancet*, *360*(9341), 1233-1242.

Bugeau, A., Bertalmío, M., Caselles, V., & Sapiro, G. (2010). A comprehensive framework for image inpainting. *IEEE transactions on image processing*, *19*(10), 2634-2645.

Bunge, L., & Clarke, A. J. (2009). A verified estimation of the El Niño index Niño-3.4 since 1877. *Journal of Climate*, *22*(14), 3979-3992.

Byun, D., & Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied mechanics reviews*, *59*(2), 51-77.

Campbell-Lendrum, D., & Prüss-Ustün, A. (2019). Climate change, air pollution and noncommunicable diseases. *Bulletin of the World Health Organization*, *97*(2), 160.

Carton, J. A., & Giese, B. S. (2008). A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Monthly weather review*, *136*(8), 2999-3017.

Carvalho, J. R. P. D., Nakai, A. M., & Monteiro, J. E. (2016). Spatio-Temporal modeling of data imputation for daily rainfall series in Homogeneous Zones. *Revista Brasileira de Meteorologia*, *31*, 196-201.

Chai, T., Kim, H. C., Lee, P., Tong, D., Pan, L., Tang, Y., ... & Stajner, I. (2013). Evaluation of the United States National Air Quality Forecast Capability experimental real-time predictions in 2010 using Air Quality System ozone and NO 2 measurements. *Geoscientific Model Development*, *6*(5), 1831-1850.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, *7*(3), 1247-1250.

Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E., & Huang, D. (2004). Predictability of El Niño over the past 148 years. *Nature*, *428*(6984), 733-736.

Choi, J. K., Park, Y. J., Ahn, J. H., Lim, H. S., Eom, J., & Ryu, J. H. (2012). GOCI, the world's first geostationary ocean color observation satellite, for the monitoring of temporal variability in coastal water turbidity. *Journal of Geophysical Research: Oceans*, *117*(C9).

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y. J., Jeong, U., ... & Song, C. K. (2016). GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during the DRAGON-NE Asia 2012 campaign.

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y. J., Holben, B., ... & Song, C. H. (2018). GOCI Yonsei aerosol retrieval version 2 products: an improved algorithm and error analysis with uncertainty estimation from 5-year validation over East Asia. *Atmospheric Measurement Techniques*, *11*(1), 385-408.

Choi, M., Lim, H., Kim, J., Lee, S., Eck, T. F., Holben, B. N., ... & Liu, H. (2019). Validation, comparison, and integration of GOCI, AHI, MODIS, MISR, and VIIRS aerosol optical depth over East Asia during the 2016 KORUS-AQ campaign. *Atmospheric Measurement Techniques*, *12*(8), 4619-4641.

Choi, Y., Kim, J., Eldering, A., Osterman, G., Yung, Y. L., Gu, Y., & Liou, K. N. (2009). Lightning and anthropogenic NOx sources over the United States and the western North Atlantic Ocean: Impact on OLR and radiative effects. *Geophysical Research Letters*, *36*(17).

Chollet, F. (2015). Keras. https://keras.io

Chollet, F. (2017). Keras (2015).

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

Chollet, F. (2018). Keras: The python deep learning library. *ascl*, ascl-1806.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Cressie, N. (1990). The origins of kriging. *Mathematical geology*, *22*(3), 239-252.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing*, *7*(3–4), 197-387.

de Weger, L. A., Beerthuizen, T., Hiemstra, P. S., & Sont, J. K. (2014). Development and validation of a 5-day-ahead hay fever forecast for patients with grass-pollen-induced allergic rhinitis. *International journal of biometeorology*, *58*(6), 1047-1055.

Dey, R., & Salem, F. M. (2017, August). Gate-variants of gated recurrent unit (GRU) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)* (pp. 1597-1600). IEEE.

Diaz, H. F., Hoerling, M. P., & Eischeid, J. K. (2001). ENSO variability, teleconnections and climate change. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *21*(15), 1845-1862.

Emberlin, J. (1994). The effects of patterns in climate and pollen abundance on allergy. *Allergy*, *49*, 15-20.

Ervens, B. T. B. W. R., Turpin, B. J., & Weber, R. J. (2011). Secondary organic aerosol formation in cloud droplets and aqueous particles (aqSOA): a review of laboratory, field and model studies. *Atmospheric Chemistry & Physics Discussions*, *11*(8).

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

Eslami, E., Salman, A. K., Choi, Y., Sayeed, A., & Lops, Y. (2019b). A data ensemble approach for real-time air quality forecasting using extremely randomized trees and deep neural networks. *Neural Computing and Applications*, 1-17.

Eyth, A., & Vukovich, J. (2016). Technical Support Document (TSD) preparation of emissions inventories for the version 6.3, 2011 emissions modeling platform. *U S Environmental Protection Agency, Office of Air Quality Planning and Standards*.

Eyth, A., Vukovich, J., Farkas, C. (2016). Technical Support Document (TSD): Preparation of Emissions Inventories for the Version 7.1 - 2016 *North American Emissions Modeling Platform*.

Fairley, D., & Batchelder, G. L. (1986). A study of oak-pollen production and phenology in northern California: prediction of annual variation in pollen counts based on geographic and meterologic factors. *Journal of Allergy and Clinical Immunology*, *78*(2), 300-307.

Fawzi, A., Samulowitz, H., Turaga, D., & Frossard, P. (2016, September). Adaptive data augmentation for image classification. In *2016 IEEE international conference on image processing (ICIP)* (pp. 3688-3692). Ieee.

Fedotova, E. V., & Luferov, V. S. (2019, November). Implementation of artificial intelligence for analysis of long-term climate variability. In *IOP Conference Series: Earth and Environmental Science* (Vol. 386, No. 1, p. 012054). IOP Publishing.

Fisher, N. I., Lewis, T., & Embleton, B. J. (1993). *Statistical analysis of spherical data*. Cambridge university press.

Fu, R., Zhang, Z., & Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* (pp. 324-328). IEEE.

García-Mozo, H., Yaezel, L., Oteros, J., & Galán, C. (2014). Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. *Science of the Total Environment*, *473*, 103-109.

Gerber, F., de Jong, R., Schaepman, M. E., Schaepman-Strub, G., & Furrer, R. (2018). Predicting missing values in spatio-temporal remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(5), 2841-2853.

Ghahremanloo, M., Choi, Y., Sayeed, A., Salman, A. K., Pan, S., & Amani, M. (2021a). Estimating Daily High-Resolution PM2. 5 Concentrations over Texas: Machine Learning Approach. *Atmospheric Environment*, 118209.

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

Ghahremanloo, M., Lops, Y., Choi, Y., & Yeganeh, B. (2021c). Deep Learning Estimation of Daily Ground-Level NO_2 Concentrations from Remote Sensing Data. *Journal of Geophysical Research: Atmospheres*. 126, e2021JD034925. https://doi.org/10.1029/2021jd034925

Giese, B. S., & Ray, S. (2011). El Niño variability in simple ocean data assimilation (SODA), 1871–2008. *Journal of Geophysical Research: Oceans*, *116*(C2).

Gould, J., Roemmich, D., Wijffels, S., Freeland, H., Ignaszewsky, M., Jianping, X., ... & Riser, S. (2004). Argo profiling floats bring new era of in situ ocean observations. *Eos, Transactions American Geophysical Union*, *85*(19), 185-191.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354-377. doi:10.1016/j.patcog.2017.10.013.

Guanter, L., Aben, I., Tol, P., Krijger, J. M., Hollstein, A., Köhler, P., ... & Landgraf, J. (2015). Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. *Atmospheric Measurement Techniques*, *8*(3), 1337-1352.

Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y. C., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, *40*(30), 5880-5892.

Gupta, M., Kodamana, H., & Sandeep, S. (2020). Prediction of enso beyond spring predictability barrier using deep convolutional lstm networks. *IEEE Geoscience and Remote Sensing Letters*.

Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568-572.

Hase, H., Masumoto, Y., Kuroda, Y., & Mizuno, K. (2008). Semiannual variability in temperature and salinity observed by Triangle Trans-Ocean Buoy Network (TRITON) buoys in the eastern tropical Indian Ocean. *Journal of Geophysical Research: Oceans*, *113*(C1).

Hashempour, Y., Nasseri, M., Mohseni-Bandpei, A., Motesaddi, S., & Eslamizadeh, M. (2020). Assessing vulnerability to climate change for total organic carbon in a system of drinking water supply. *Sustainable Cities and Society*, *53*, 101904.

He, D., Lin, P., Liu, H., Ding, L., & Jiang, J. (2019, August). Dlenso: A deep learning enso forecasting model. In *Pacific Rim International Conference on Artificial Intelligence* (pp. 12-23). Springer, Cham.

Helbig, N., Vogel, B., Vogel, H., & Fiedler, F. (2004). Numerical modelling of pollen dispersion on the regional scale. *Aerobiologia*, *20*(1), 3-19.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., ... & Wright, J. S. (2019). From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmospheric Chemistry and Physics*, *19*(5), 3097-3124.

Hogrefe, C., Pouliot, G., Wong, D., Torian, A., Roselle, S., Pleim, J., & Mathur, R. (2015). Annual application and evaluation of the online coupled WRF–CMAQ system over North America under AQMEII phase 2. *Atmospheric Environment*, *115*, 683-694.

Holben, B. N., Eck, T. F., Slutsker, I. A., Tanre, D., Buis, J. P., Setzer, A., ... & Smirnov, A. (1998). AERONET—A federated instrument network and data archive for aerosol characterization. *Remote sensing of environment*, *66*(1), 1-16.

Holben, B. N., Tanre, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., ... & Kaufman, Y. J. (2001). An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. *Journal of Geophysical Research: Atmospheres*, *106*(D11), 12067-12097.

Huang, B., Xue, Y., & Behringer, D. W. (2008). Impacts of Argo salinity in NCEP global ocean data assimilation system: the tropical Indian Ocean. *Journal of Geophysical Research: Oceans*, *113*(C8).

Huang, B., Xue, Y., Zhang, D., Kumar, A., & McPhaden, M. J. (2010). The NCEP GODAS ocean analysis of the tropical Pacific mixed layer heat budget on seasonal to interannual time scales. *Journal of Climate*, *23*(18), 4901-4925.

Ishii, M., & Kimoto, M. (2009). Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *Journal of Oceanography*, *65*(3), 287-299.

Jäger, S. (1989). Trends in the frequency of different pollen types in Vienna from 1976 to 1989. *Aerobiologia*, *5*(1), 9-16.

Jeon, W., Choi, Y., Roy, A., Pan, S., Price, D., Hwang, M. K., ... & Oh, I. (2018). Investigation of primary factors affecting the variation of modeled oak pollen concentrations: a case study for Southeast Texas in 2010. *Asia-Pacific Journal of Atmospheric Sciences*, *54*(1), 33-41.

Jin, K. H., McCann, M. T., Froustey, E., & Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, *26*(9), 4509-4522.

Jones, M. W., Smith, A., Betts, R., Canadell, J. G., Prentice, I. C., & Le Quéré, C. (2020). Climate change increases risk of wildfires. *ScienceBrief Review*, *116*, 117.

Jung, J., Souri, A. H., Wong, D. C., Lee, S., Jeon, W., Kim, J., & Choi, Y. (2019). The Impact of the Direct Effect of Aerosols on Meteorology and Air Quality Using Aerosol Optical Depth Assimilation During the KORUS-AQ Campaign. *Journal of Geophysical Research: Atmospheres*, 124(14), 8303-8319.

Kang, H., Zhu, B., Zhu, C., de Leeuw, G., Hou, X., & Gao, J. (2019). Natural and anthropogenic contributions to long-term variations of SO2, NO2, CO, and AOD over East China. *Atmospheric Research*, *215*, 284-293.

Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics*, (4), 580-585.

Kim, K. H., Kabir, E., & Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environment international*, *74*, 136-143.

Kim, H. C., Kim, E., Bae, C., Cho, J. H., Kim, B. U., & Kim, S. (2017). Regional contributions to particulate matter concentration in the Seoul metropolitan area, South Korea: seasonal variation and sensitivity to meteorology and emissions inventory. *Atmospheric Chemistry & Physics*, *17*(17).

Kim, S. (2019). Jeju Island climate modeling using multiple linear regression analysis. *Journal of Analysis of Applied Mathematics*, *13*, 118-136.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kiotseridis, H., Cilio, C. M., Bjermer, L., Tunsäter, A., Jacobsson, H., & Dahl, Å. (2013). Grass pollen allergy in children and adolescents-symptoms, health related quality of life and the value of pollen prognosis. *Clinical and translational allergy*, *3*(1), 1-12.

Koulouri, E., Saarikoski, S., Theodosi, C., Markaki, Z., Gerasopoulos, E., Kouvarakis, G., ... & Mihalopoulos, N. (2008). Chemical composition and sources of fine and coarse aerosol particles in the Eastern Mediterranean. *Atmospheric Environment*, *42*(26), 6542-6550.

Kramer, O. (2013). K-nearest neighbors. In *Dimensionality reduction with unsupervised nearest neighbors* (pp. 13-23). Springer, Berlin, Heidelberg.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems* (pp. 1097-1105).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.

Kyriakidis, P. C., & Journel, A. G. (1999). Geostatistical space–time models: a review. *Mathematical geology*, *31*(6), 651-684.

Labe, Z. M., & Barnes, E. A. (2021). Detecting climate signals using explainable AI with single-forcing large ensembles. *Journal of Advances in Modeling Earth Systems*, e2021MS002464.

Lal, R. (2020). Managing soils for negative feedback to climate change and positive impact on food and nutritional security. *Soil Science and Plant Nutrition*, *66*(1), 1-9.

Lauret, P., Heymes, F., Aprin, L., & Johannet, A. (2016). Atmospheric dispersion modeling using Artificial Neural Network based cellular automata. *Environmental Modelling & Software*, *85*, 56-69.

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, *8*(1), 98-113.

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

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Lee, H., & Kwon, H. (2017). Going deeper with contextual CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, *26*(10), 4843-4855.

Lee, S., Song, C. H., Park, R. S., Park, M. E., Han, K. M., Kim, J., ... & Woo, J. H. (2016). GIST-PM-Asia v1: development of a numerical system to improve particulate matter forecasts in South Korea using geostationary satellite-retrieved aerosol optical data over Northeast Asia. *Geoscientific Model Development*, *9*(1).

Levetin, E., & Van de Water, P. K. (2003). Pollen count forecasting. *Immunology and Allergy Clinics*, *23*(3), 423-442.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014, December). Medical image classification with convolutional neural network. In *2014 13th international conference on control automation robotics & vision (ICARCV)* (pp. 844-848). IEEE.

Li, W., Wang, C., Wang, H., Chen, J., Yuan, C., Li, T., ... & Wang, B. (2014). Distribution of atmospheric particulate matter (PM) in rural field, rural village and urban areas of northern China. *Environmental Pollution*, *185*, 134-140.

Li, Y., Xie, W., & Li, H. (2017). Hyperspectral image reconstruction by deep convolutional neural network for classification. *Pattern Recognition*, *63*, 371-383.

Liu, G., Reda, F. A., Shih, K. J., Wang, T. C., Tao, A., & Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 85-100).

Liu, L., Solmon, F., Vautard, R., Hamaoui-Laguel, L., Torma, C. Z., & Giorgi, F. (2016). Ragweed pollen production and dispersion modelling within a regional climate system, calibration and application over Europe. *Biogeosciences*, *13*(9), 2769-2786.

Liu, Y., Chen, X., Huang, S., Tian, L., Lu, Y. A., Mei, Y., ... & Xiang, H. (2015). Association between air pollutants and cardiovascular disease mortality in Wuhan, China. *International Journal of Environmental Research and Public Health*, *12*(4), 3506-3516.

Lops, Y., Choi, Y., Eslami, E., & Sayeed, A. (2019). Real-time 7-day forecast of pollen counts using a deep convolutional neural network. *Neural Computing and Applications*, 1-10.

Lops, Y., Pouyaei, A., Choi, Y., Jung, J., Salman, A. K., & Sayeed, A. (2021). Application of a Partial Convolutional Neural Network for Estimating Geostationary Aerosol Optical Depth Data. *Geophysical Research Letters*, *48*(15), e2021GL093096.

Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & geosciences*, *34*(9), 1044-1055.

Ludewig, A. (2021). S5P Mission Performance Centre Level 1b Readme. Reference: S5P-MPC-KNMI-PRF-L1B, (2.0), 3.0.0.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013, June). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (Vol. 30, No. 1, p. 3).

Makra, L., Matyasovszky, I., Thibaudon, M., & Bonini, M. (2011). Forecasting ragweed pollen characteristics with nonparametric regression methods over the most polluted areas in Europe. *International journal of biometeorology*, *55*(3), 361-371.

Martin, M. D., Chamecki, M., & Brush, G. S. (2010). Anthesis synchronization and floral morphology determine diurnal patterns of ragweed pollen dispersal. *Agricultural and Forest Meteorology*, *150*(9), 1307-1317.

Martin, R. V. (2008). Satellite remote sensing of surface air quality. *Atmospheric environment*, *42*(34), 7823-7843.

McPhaden, M. J. (1995). The tropical atmosphere ocean array is completed. *Bulletin of the American Meteorological Society*, *76*(5), 739-741.

Mhawish, A., Kumar, M., Mishra, A. K., Srivastava, P. K., & Banerjee, T. (2018). Remote sensing of aerosols from space: retrieval of properties and applications. In *Remote Sensing of Aerosols, Clouds, and Precipitation* (pp. 45-83). Elsevier.

Mikhaylov, A., Moiseev, N., Aleshin, K., & Burkhardt, T. (2020). Global climate change and greenhouse effect. *Entrepreneurship and Sustainability Issues*, *7*(4), 2897.

Mikolov, T., Kombrink, S., Burget, L., Černocký, J., & Khudanpur, S. (2011, May). Extensions of recurrent neural network language model. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 5528-5531). IEEE. Mukherjee, A., & Agrawal, M. (2017). World air particulate matter: sources, distribution and health effects. *Environmental Chemistry Letters*, 15(2), 283-309.

Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P., & Gentine, P. (2021). Assessing the Potential of Deep Learning for Emulating Cloud Superparameterization in Climate Models With Real-Geography Boundary Conditions. *Journal of Advances in Modeling Earth Systems*, *13*(5), e2020MS002385.

Mu, B., Peng, C., Yuan, S., & Chen, L. (2019, July). ENSO forecasting over multiple time horizons using ConvLSTM network and rolling mechanism. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering*, *114*(4), 358-371.

Myszkowska, D. (2014). Poaceae pollen in the air depending on the thermal conditions. *International journal of biometeorology*, *58*(5), 975-986.

Nagi, J., Ducatelle, F., Di Caro, G. A., Cireşan, D., Meier, U., Giusti, A., ... & Gambardella, L. M. (2011, November). Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 342-347). IEEE.

NOAA. *Climate indices: Monthly atmospheric and ocean time series: NOAA Physical Sciences Laboratory*. PSL. Retrieved September 17, 2021, from https://psl.noaa.gov/data/climateindices/list/.

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2548-2563.

Orr, M. J. (1996). Introduction to radial basis function networks.

Park, R. S., Song, C. H., Han, K. M., Park, M. E., Lee, S. S., Kim, S. B., & Shimizu, A. (2011). A study on the aerosol optical properties over East Asia using a combination of CMAQ-simulated aerosol optical properties and remote-sensing data via a data assimilation technique. *Atmospheric Chemistry and Physics*, *11*(23), 12275-12296.

Park, S., Jeong, Y., & Kim, H. S. (2017, November). Multiresolution CNN for reverberant speech recognition. In *2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)* (pp. 1-4). IEEE.

Partee, S., Ellis, M., Rigazzi, A., Bachman, S., Marques, G., Shao, A., & Robbins, B. (2021). Using Machine Learning at Scale in HPC Simulations with SmartSim: An Application to Ocean Climate Modeling. *arXiv preprint arXiv:2104.09355*.

Pawankar, R., Canonica, G. W., Holgate, S. T., Lockey, R. F., & Blaiss, M. (2013). World Allergy Organisation (WAO) white book on allergy: update 2013. Milwaukee: *World Allergy Organization*.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & geosciences*, *30*(7), 683-691.

Peng, D., Zhou, T., Zhang, L., Zhang, W., & Chen, X. (2020). Observationally constrained projection of the reduced intensification of extreme climate events in Central Asia from 0.5° C less global warming. *Climate Dynamics*, *54*(1), 543-560.

Pitchford, M., Malm, W., Schichtel, B., Kumar, N., Lowenthal, D., & Hand, J. (2007). Revised algorithm for estimating light extinction from IMPROVE particle speciation data. *Journal of the Air & Waste Management Association*, *57*(11), 1326-1336.

Pleim, J. E., & Gilliam, R. (2009). An indirect data assimilation scheme for deep soil temperature in the Pleim–Xiu land surface model. *Journal of Applied Meteorology and Climatology*, *48*(7), 1362-1376.

Pleim, J. E., & Xiu, A. (2003). Development of a land surface model. Part II: Data assimilation. *Journal of Applied Meteorology*, *42*(12), 1811-1822.

Pouyaei, A., Choi, Y., Jung, J., Sadeghi, B., & Song, C. H. (2020). Concentration Trajectory Route of Air pollution with an Integrated Lagrangian model (C-TRAIL Model v1. 0) derived from the Community Multiscale Air Quality Model (CMAQ Model v5. 2). *Geoscientific Model Development*, *13*(8), 3489-3505.

Pouyaei, A., Sadeghi, B., Choi, Y., Jung, J., Souri, A. H., Zhao, C., & Song, C. H. (2021). Development and Implementation of a Physics-Based Convective Mixing Scheme in the Community Multiscale Air Quality Modeling Framework. *Journal of Advances in Modeling Earth Systems*, *13*(6), e2021MS002475.

Prank, M., Chapman, D. S., Bullock, J. M., Belmonte, J., Berger, U., Dahl, A., ... & Sofiev, M. (2013). An operational model for forecasting ragweed pollen release and dispersion in Europe. *Agricultural and forest meteorology*, *182*, 43-53.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*.

Rojo, J., Rapp, A., Lara, B., Fernández-González, F., & Pérez-Badia, R. (2015). Effect of land uses and wind direction on the contribution of local sources to airborne pollen. *Science of the Total Environment*, *538*, 672-682.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Rossi, R. E., Dungan, J. L., & Beck, L. R. (1994). Kriging in the shadows: geostatistical interpolation for remote sensing. *Remote Sensing of Environment*, *49*(1), 32-40.

Rulloni, V., Bustos, O., & Flesia, A. G. (2012). Large gap imputation in remote sensed imagery of the environment. *Computational Statistics & Data Analysis*, *56*(8), 2388-2403.

Ryu, J. H., Han, H. J., Cho, S., Park, Y. J., & Ahn, Y. H. (2012). Overview of geostationary ocean color imager (GOCI) and GOCI data processing system (GDPS). *Ocean Science Journal*, *47*(3), 223-233.

Sánchez, J. M. B., Lugilde, D. N., de Linares Fernández, C., de la Guardia, C. D., & Sánchez, F. A. (2007). Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. *Expert Systems with Applications*, *32*(4), 1218-1225.

Sánchez-Mesa, J. A., Galán, C., Martínez-Heras, J. A., & Hervás-Martínez, C. (2002). The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula. *Clinical & Experimental Allergy*, *32*(11), 1606-1612.

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

Sayeed, A., Lops, Y., Choi, Y., Jung, J., & Salman, A. K. (2021). Bias Correcting and Extending the PM Forecast by CMAQ up to 7 days using Deep Convolutional Neural Networks. *Atmospheric Environment*, 118376.

Scaramuzza, P., & Barsi, J. (2005). Landsat 7 scan line corrector-off gap-filled product development. In *Proceeding of Pecora* (Vol. 16, pp. 23-27).

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

Servain, J., Busalacchi, A. J., McPhaden, M. J., Moura, A. D., Reverdin, G., Vianna, M., & Zebiak, S. E. (1998). A pilot research moored array in the tropical Atlantic (PIRATA). *Bulletin of the American Meteorological Society*, *79*(10), 2019-2032.

Shah, A. S., Langrish, J. P., Nair, H., McAllister, D. A., Hunter, A. L., Donaldson, K., ... & Mills, N. L. (2013). Global association of air pollution and heart failure: a systematic review and meta-analysis. *The Lancet*, *382*(9897), 1039-1048.

Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., & Zhang, L. (2015). Missing information reconstruction of remote sensing data: A technical review. *IEEE Geoscience and Remote Sensing Magazine*, *3*(3), 61-85.

Smith, R. D., Dukowicz, J. K., & Malone, R. C. (1992). Parallel ocean general circulation modeling. *Physica D: Nonlinear Phenomena*, *60*(1-4), 38-61.

Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on nuclear science*, *44*(3), 1464-1468.

Soldevilla, C. G., González, P. C., Teno, P. A., & Vilches, E. D. (2007). Spanish Aerobiology Network (REA): management and quality manual. *Servicio de publicaciones de la Universidad de Córdoba*, 1-300.

Srindhuna, M., & Baburaj, M. (2020, June). Estimation of Missing Data in Remote Sensing Images Using t-SVD based tensor completion. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res*, 15(1), 1929-1958.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American meteorological Society*, *93*(4), 485-498.

Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of graphics tools*, *9*(1), 23-34.

Toro A, R., Córdova J, A., Canales, M., Morales S, R. G., Mardones P, P., & Leiva G, M. A. (2015). Trends and threshold exceedances analysis of airborne pollen concentrations in Metropolitan Santiago Chile. *PloS one*, *10*(5), e0123077.

Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp. 242-264). IGI global.

Traidl-Hoffmann, C., Kasche, A., Menzel, A., Jakob, T., Thiel, M., Ring, J., & Behrendt, H. (2003). Impact of pollen on human health: more than allergen carriers?. *International archives of allergy and immunology*, *131*(1), 1-13.

Trenberth, K. E., & Stepaniak, D. P. (2001). Indices of el Niño evolution. *Journal of climate*, *14*(8), 1697-1701.

Tsoi, A. C., & Back, A. D. (1994). Locally recurrent globally feedforward networks: a critical review of architectures. *IEEE Transactions on Neural Networks*, 5(2), 229-239.

Tzuc, O. M., Gamboa, O. R., Rosel, R. A., Poot, M. C., Edelman, H., Torres, M. J., & Bassam, A. (2021). Modeling of hygrothermal behavior for green facade's concrete wall exposed to nordic climate using artificial intelligence and global sensitivity analysis. *Journal of Building Engineering*, *33*, 101625.

Uranishi, K., Ikemori, F., Shimadera, H., Kondo, A., & Sugata, S. (2019). Impact of field biomass burning on local pollution and long-range transport of PM2. 5 in Northeast Asia. *Environmental Pollution*, *244*, 414-422.

Urraca, R., Huld, T., Gracia-Amillo, A., Martinez-de-Pison, F. J., Kaspar, F., & Sanz-Garcia, A. (2018). Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data. *Solar Energy*, *164*, 339-354.

USEPA. (2020). Particulate Matter (PM) Basics. Retrieved December, 2020, from https://www.epa.gov/pm-pollution/particulate-matter-pm-basics

U.S. Geological Survey (2020) USGS EROS Archive - Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30). Retrieved November, 2020, from https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., & Arellano Jr, A. F. (2006). Interannual variability in global biomass burning emissions from 1997 to 2004. *Atmospheric Chemistry and Physics*, *6*(11), 3423-3441.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2017, August). Deepsd: Generating high resolution climate change projections through single image super-resolution. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1663-1672).Vogel, H., Pauling, A., & Vogel, B. (2008). Numerical simulation of birch pollen dispersion with an operational weather forecast system. *Int. J. Biometeorol*, 52(8), 805-814. doi:10.1007/s00484-008-0174-3.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., ... & Levelt, P. F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations

of the atmospheric composition for climate, air quality and ozone layer applications. *Remote sensing of environment*, *120*, 70-83.

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M., ... & Winkler, H. (2020). TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations. *Atmospheric Measurement Techniques*, *13*(7), 3751-3767.

Vu, D. T., Yamada, T., & Ishidaira, H. (2018). Assessing the impact of sea level rise due to climate change on seawater intrusion in Mekong Delta, Vietnam. *Water Science and Technology*, *77*(6), 1632-1639.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011). The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, *4*(3), 625-641.

Wiedinmyer, C., Quayle, B., Geron, C., Belote, A., McKenzie, D., Zhang, X., ... & Wynne, K. K. (2006). Estimating emissions from fires in North America for air quality modeling. *Atmospheric Environment*, *40*(19), 3419-3432.

Wiedinmyer, C., Yokelson, R. J., & Gullett, B. K. (2014). Global emissions of trace gases, particulate matter, and hazardous air pollutants from open burning of domestic waste. *Environmental science & technology*, *48*(16), 9523-9530.

Wang, X., Huang, J., Ji, M., & Higuchi, K. (2008). Variability of East Asia dust events and their long-term trend. *Atmospheric Environment*, *42*(13), 3156-3165.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., ... & Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. *J. Geophys. Res: Oceans*, 90(C5), 8995-9005.

Wozniak, M. C., & Steiner, A. L. (2017). A prognostic pollen emissions model for climate models (PECM1. 0). *Geoscientific Model Development*, *10*(11), 4105-4127.

Wu, Y., Arapi, A., Huang, J., Gross, B., & Moshary, F. (2018). Intra-continental wildfire smoke transport and impact on local air quality observed by ground-based and satellite remote sensing in New York City. *Atmospheric Environment*, *187*, 266-281.

Wu, Z., Chen, J., Wang, Y., Zhu, Y., Liu, Y., Yao, B., ... & Hu, M. (2018). Interactions between water vapor and atmospheric aerosols have key roles in air quality and climate change. *National Science Review*, *5*(4), 452-454.

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

Yeo, I., Choi, Y., Lops, Y., & Sayeed, A. (2021). Efficient PM2. 5 forecasting using geographical correlation based on integrated deep learning algorithms. *Neural Computing and Applications*, 1-17.

Yienger, J. J., & Levy, H. (1995). Empirical model of global soil-biogenic NOχ emissions. *Journal of Geophysical Research: Atmospheres*, *100*(D6), 11447-11464.

Yu, C., Chen, L., Su, L., Fan, M., & Li, S. (2011, June). Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth. In *2011 19th International Conference on Geoinformatics* (pp. 1-6). IEEE.

Yuan, X., Kaplan, M. R., & Cane, M. A. (2018). The interconnected global climate system—A review of tropical–polar teleconnections. *Journal of Climate*, *31*(15), 5765-5792.

Zhang, C., Li, W., & Travis, D. (2007). Gaps-fill of SLC-off Landsat ETM+ satellite image using a geostatistical approach. *International Journal of Remote Sensing*, *28*(22), 5103-5122.

Zhang, C. Y., Chen, C. P., Gan, M., & Chen, L. (2015). Predictive deep Boltzmann machine for multiperiod wind speed forecasting. *IEEE Transactions on Sustainable Energy*, *6*(4), 1416-1425.

Zhang, J., Zhang, L. Y., Du, M., Zhang, W., Huang, X., Zhang, Y. Q., ... & Xiao, H. (2016). Indentifying the major air pollutants base on factor and cluster analysis, a case study in 74 Chinese cities. *Atmospheric Environment*, *144*, 37-46.

Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag*, 4(2), 22-40. doi:10.1109/mgrs.2016.2540798.

Zhang, Q., Geng, G., Wang, S., Richter, A., & He, K. (2012). Satellite remote sensing of changes in NOx emissions over China during 1996–2010. *Chinese Science Bulletin*, *57*(22), 2857-2864.

Zhang, Q., Yuan, Q., Zeng, C., Li, X., & Wei, Y. (2018). Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(8), 4274-4288.

Zhang, R., Duhl, T., Salam, M. T., House, J. M., Flagan, R. C., Avol, E. L., ... & VanReken, T. M. (2014). Development of a regional-scale pollen emission and transport modeling framework for investigating the impact of climate change on allergic airway disease. *Biogeosciences*, *11*(6), 1461-1478.

Zhang, X., Zou, Y., & Shi, W. (2017, August). Dilated convolution neural network with LeakyReLU for environmental sound classification. In *2017 22nd International Conference on Digital Signal Processing (DSP)* (pp. 1-5). IEEE.

Zhu, F., Ding, R., Lei, R., Cheng, H., Liu, J., Shen, C., ... & Cao, J. (2019). The short-term effects of air pollution on respiratory diseases and lung cancer mortality in Hefei: A time-series analysis. *Respiratory medicine*, *146*, 57-65.

Zink, K., Vogel, H., Vogel, B., Magyar, D., & Kottmeier, C. (2012). Modeling the dispersion of Ambrosia artemisiifolia L. pollen with the model system COSMO-ART. *International journal of biometeorology*, *56*(4), 669-680.