

# Deep learning-based downscaling of tropospheric nitrogen dioxide using ground-level and satellite observations

Manzhu Yu <sup>a,\*</sup>, Qian Liu <sup>b</sup>

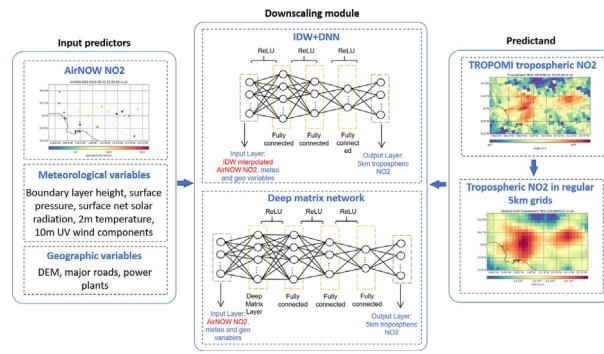
<sup>a</sup> Department of Geography, Institute of Computational and Data Sciences, Pennsylvania State University, PA, USA

<sup>b</sup> NSF Spatiotemporal Innovation Center, Department of Geography and GeoInformation Science, George Mason University, USA

## HIGHLIGHTS

- Deep learning methods are developed to downscale tropospheric nitrogen dioxide ( $\text{NO}_2$ ).
- Predictors include surface  $\text{NO}_2$  and other meteorological and geographical variables.
- Tropospheric  $\text{NO}_2$  can be estimated at the sub-urban scale on an hourly basis.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 23 October 2020

Received in revised form 4 January 2021

Accepted 7 January 2021

Available online 6 February 2021

Editor: Jianmin Chen

### Keywords:

Nitrogen dioxide  
Spatial downscaling  
Spatial interpolation  
Deep learning  
TROPOMI  
AirNOW

## ABSTRACT

Air quality is one of the major issues within an urban area that affect people's living environment and health conditions. Existing observations are not adequate to provide a spatiotemporally comprehensive air quality information for vulnerable populations to plan ahead. Launched in 2017, TROPOspheric Monitoring Instrument (TROPOMI) provides a high spatial resolution (~5 km) tropospheric air quality measurement that captures the spatial variability of air pollution, but still limited by its daily overpass in the temporal dimension and relatively short historical records. Integrating with the hourly available AirNOW observations by ground-level discrete stations, we proposed and compared two deep learning methods that learn the relationship between the ground-level nitrogen dioxide ( $\text{NO}_2$ ) observation from AirNOW and the tropospheric  $\text{NO}_2$  column density from TROPOMI to downscale the daily  $\text{NO}_2$  to an hourly resolution. The input predictors include the locations of AirNOW stations, AirNOW  $\text{NO}_2$  observations, boundary layer height, other meteorological status, elevation, major roads, and power plants. The learned relationship can be used to produce  $\text{NO}_2$  emission estimates at the sub-urban scale on an hourly basis. The two models include 1) an integrated method between inverse weighted distance and a feed forward neural network (IDW + DNN), and 2) a deep matrix network (DMN) that maps the discrete AirNOW observations directly to the distribution of TROPOMI observations. We further compared the accuracies of both models using different configurations of input predictors and validated their average Root Mean Squared Error (RMSE), average Mean Absolute Error (MAE) and the spatial distribution of errors. Results show that DMN generates more reliable  $\text{NO}_2$  estimates and captures a better spatial distribution of  $\text{NO}_2$  concentrations than the IDW + DNN model.

© 2021 Elsevier B.V. All rights reserved.

\* Corresponding author.

E-mail address: [mqy5198@psu.edu](mailto:mqy5198@psu.edu) (M. Yu).

## 1. Introduction

Air pollution continues to be a widespread issue despite government efforts to decrease the amount of pollution in the air since the 1970s. Urban areas are especially at risk for poor air quality and the associated health effects. According to the American Lung Association, approximately 45.8% of Americans live in counties with unhealthy air (American Lung Association, 2020). The timely and precise production and dissemination of air quality information (e.g., PM<sub>2.5</sub>, and PM<sub>10</sub>, and ozone) with a high-spatiotemporal resolution to urban citizens would be of great importance for them to make daily activity decisions for protecting their health and saving their lives eventually. Nitrogen oxides are a category of gases regulated by the United States Environmental Protection Agency (US EPA)—nitrogen dioxide being among the most important. Nitrogen dioxide (NO<sub>2</sub>) is mainly produced from the consumption of fossil fuel. By far, the leading contributors to nitrogen dioxide emissions are power plants, cars and trucks and non-road equipment. Breathing in high levels of NO<sub>2</sub> can lead to respiratory problems. NO<sub>2</sub> can cause coughing and wheezing symptoms by irritating the lining of the lungs, and impairs the ability of human bodies to defend pulmonary infections. Due its high sensitivity, NO<sub>2</sub> is also an essential indicator of industrial production and can be utilized in the assessment of economic conditions (Duncan et al., 2016).

Reliable and comprehensive NO<sub>2</sub> emission estimates are needed to evaluate air quality mitigation strategies, estimate industrial production, and as input to models for simulating and forecasting air pollution. Ground-level observations of NO<sub>2</sub> are regularly measured by weather stations where air quality sensors are mounted, such as AirNOW, Purple Air, and IQAir. However, the discrete air quality observations are limited where no observed measurements are available. Satellite instruments, including Global Ozone Monitoring Experiment (GOME), Ozone Monitoring Instrument (OMI), and TROPOspheric Monitoring Instrument (TROPOMI), retrieve atmospheric trace gas concentrations in the atmosphere using spectroscopy. NO<sub>2</sub> column density can be determined by measuring the backscattered light, and tropospheric NO<sub>2</sub> column and stratospheric NO<sub>2</sub> column are separated using a data assimilation system (Veenkind et al., 2012). The advantage of satellite NO<sub>2</sub> observation is the capability of providing a comprehensive perspective on the spatial distribution of global emissions. However, TROPOMI's daily overpass limits the benefit of satellite NO<sub>2</sub> observation in the temporal dimension, whereas NO<sub>2</sub> values show a high daily variability (Blond et al., 2007). Accurate emission estimates remain clearly needed at the suburban scale on an hourly basis. Furthermore, high-resolution satellite observations for NO<sub>2</sub> column densities are with relatively short historical records, such as TROPOMI that is only available since 2018. Climatological analysis is usually done with lower spatial resolutions using sensors such as OMI which has a resolution of 0.25 degree (Liu et al., 2018). A reliable method for estimating NO<sub>2</sub> emissions with dataset that has a longer availability period is crucial for environmental analysis.

High resolution NO<sub>2</sub> emission forecasts can be produced by numerical simulations, such as the Community Multiscale Air Quality Modeling System (CMAQ, Uno et al., 2007) and the Weather Research and Forecasting (WRF) model coupled with Chemistry (WRF-CHEM, Ghude et al., 2013). Although the simulated NO<sub>2</sub> emissions correlate in a good agreement with satellite observations, high-resolution numerical simulations require time- and memory-consuming computations (Fuhrer et al., 2018), in addition, the high-resolution numerical weather prediction (NWP) data might not be available to all the public users (Baklanov et al., 2002). Spatiotemporal downscaling based on heterogeneous observations can provide an alternative approach to complement the spatiotemporal resolutions from different data sources. Existing downscaling methods include dynamical downscaling and statistical downscaling. Dynamical downscaling simulates using high-resolution physical local-area models based on low-resolution boundary conditions; however, it is computational demanding (Hong et al., 2017; Yahya et al., 2017; Wang et al., 2020). Statistical downscaling

trains linear or nonlinear statistical models to estimate high-resolution information, but the downscaled variable is generally the same as the low-resolution origin (Zhu et al., 2016; Ahmed et al., 2018; Oteros et al., 2019; Khan et al., 2019). In addition, most existing downscaling applications for climate and meteorological data are based on structured grid, while few have explored on unstructured grid, such as generating high resolution information based on observations from discrete weather stations.

To fill the aforementioned gaps and produce a high-spatiotemporal resolution NO<sub>2</sub> tropospheric column density product, this research proposes and compares two deep learning methods that learn the relationship between the ground-level NO<sub>2</sub> observation from AirNOW and the tropospheric NO<sub>2</sub> column density from TROPOMI. The input predictors include the locations of AirNOW stations, AirNOW NO<sub>2</sub> observations, boundary layer height, other meteorological status, elevation, major roads, and power plants. The learned relationship can be used to produce NO<sub>2</sub> emission estimates at the sub-urban scale on an hourly basis. The two methods include 1) an integrated method between inverse weighted distance and a feed forward neural network (IDW + DNN), and 2) a deep matrix network (DMN) that maps the discrete AirNOW observations directly to the distribution of TROPOMI observations. We compared the accuracy of both models in estimating tropospheric NO<sub>2</sub> in the larger Los Angeles area, analyzed the feature importance of the input predictors, and examined the spatial distribution of prediction errors. The proposed methods and results can also be utilized on long-term climatic and environmental analysis with high spatiotemporal resolutions by inputting historical record of model predictors.

## 2. Related work

### 2.1. Spatial interpolation of airborne pollutants

Spatial interpolation is one of the most widely used methods to estimate the air pollution distribution where no observed measurements are available. A variety of spatial interpolation methods utilizes nonlocal geometric similarities to construct high-resolution images (Zhu et al., 2016), analyze the spatiotemporal variograms to conduct spatiotemporal kriging-based interpolation (Ahmed et al., 2018; Oteros et al., 2019), or examined the adjacent slope to perform the interpolation (Khan et al., 2019).

Zhu et al. (2016) proposed a robust interpolation scheme by using the nonlocal geometric similarities to construct the high-resolution image. In their method, the minimum mean square error (MMSE)-based interpolation weighting coefficients are generated by solving a regularized least squares problem that is built upon a number of dual-reference patches drawn from the given low-resolution image and regularized by the directional gradients of these patches. The accuracy is higher than common interpolation methods such as Bicubic, new edge-directed interpolation (NEDI), soft-decision adaptive interpolation (SAI), and regularized local linear regression (RLLR). Ahmed et al. (2018) analyzed the spatiotemporal variability of air pollutants, SO<sub>2</sub>, NO<sub>2</sub>, and PM<sub>10</sub>, in Egypt from the air quality monitoring network, and employed the spatiotemporal kriging to interpolate the monthly averages. Nori-Sarma et al. (2020) monitored, modeled and interpolated nitrogen dioxide (NO<sub>2</sub>) in Mysore, a rapidly urbanizing city in India using two distinct models: land use regression (LUR) approach and universal kriging methods. They concluded that the influence of pollution factors (e.g., point sources) and highly localized characteristics of the urban environment (e.g., proximity to religious points of interest) are the major contributors to ambient air pollution levels.

As reviewed above, spatial interpolation methods are used to interpolate a particular variable from discrete known points or lower-resolution grids to higher-resolution map distributions. However, this research combines the advantage of NO<sub>2</sub> measurements from two different sources (i.e., ground-level and tropospheric) and extend the

downscaling application to the temporal domain, thus simply interpolating the ground-level NO<sub>2</sub> does not translate directly to tropospheric NO<sub>2</sub>. Therefore, we propose to use deep learning methods to learn the relationship between the two NO<sub>2</sub> measurements and predict the higher-resolution map distribution of tropospheric NO<sub>2</sub>.

## 2.2. Dynamical downscaling of airborne pollutants

Dynamical downscaling uses high-resolution physical local-area models to dynamically extrapolate the effects of large-scale climate processes to regional or local scales of interest. Dynamical downscaling relies on explicit representations of physical principals, e.g., the laws of thermodynamics and fluid mechanics, thus they can be sensitive to large-scale biases. Hong et al. (2017) presented a two-way coupled model using dynamical downscaling approach with advanced chemistry and aerosol treatments for high-resolution regional simulations (downscaled from 0.9° × 1.25° to 36 km × 36 km). Results show good predictability of PM<sub>2.5</sub> in winter and O<sub>3</sub> in summer in terms of statistical performance and spatial distributions. Yahya et al. (2017) downscaled a localized version of the Community Earth System Model using Weather Research and Forecasting Model with Chemistry (WRF/Chem) from 0.9° × 1.25° to 36 km × 36 km, and improved the model performance for most surface meteorological variables (except for precipitation) and air quality variables including O<sub>3</sub> and PM<sub>2.5</sub>. Wang et al. (2020) examined four different dynamical downscaling methods to increase the spatial resolution of SO<sub>2</sub> and NO<sub>x</sub> emissions generated by GEOS-Chem from 2° × 2.5°. The dynamical downscaling methods integrated high-resolution emission inventory, nighttime light observations, and TROPOMI NO<sub>2</sub> observations to downscale coarse-resolution concentrations. However, dynamical downscaling can be computational demanding, with a four-times increase of horizontal resolution, the theoretical required computation will increase by about 4 (latitude) × 4 (longitude) × 4 (time steps) = 64 times (Xie et al., 2010). In addition, dynamical downscaling required input data with a structured grid, whereas few existing works have handled data with an unstructured grid. Our proposed methods have higher computational efficiencies and lower requirements for input dataset, which benefit the users with non-climate-model background.

## 2.3. Machine learning based downscaling of airborne pollutants

A variety of machine learning methods have been developed to downscale airborne pollutants by using random forests (Liu et al., 2018), support vector regression (Berrocal et al., 2020), feed forward neural networks (Di et al., 2016), and generalized additive models (Xiao et al., 2017). These methods are not restricted to downscaling datasets with regular grids, but ranges from a variety of input possibilities.

Liu et al. (2018) downscaled the satellite derived gridded PM<sub>2.5</sub> datasets (typically at 0.1° of spatial resolution) to a refined spatial resolution (0.01°) using the combination of random forests and regression kriging. The study demonstrated the effectiveness of integrating long-term environmental variables into the model, including brightness of nighttime lights, elevation, and normalized difference vegetation index. Di et al. (2016) trained a feed forward convolutional neural network using discrete PM<sub>2.5</sub> daily monitoring data and other environmental information (normalized difference vegetation index (NDVI), surface reflectance, absorbing aerosol index, and meteoroidal fields) for the continental United States from 2000 to 2012. The trained neural network was then used to predict daily PM<sub>2.5</sub> at 1 km × 1 km grid cells, but this predicted daily output is not validated regarding its accuracy. Xiao et al. (2017) used a GAM to impute MAIAC AOD over the Yangtze River Delta of China, which achieved an average R<sup>2</sup> of 0.77 (ranging from 0.48 to 0.97 in model fitting) and an R<sup>2</sup> of 0.44 in validation with AERONET AOD. In addition, they included more covariates than what we used in this paper, such as cloud fraction, normalized difference

vegetation index and CMAQ simulations. However, CMAQ simulations are not always publicly and readily available for the long time series imputation in certain regions. Li et al. (2020) trained an autoencoder based residual network to fill in missing data in the high-resolution (1 km daily) Multiangle Implementation of Atmospheric Correction (MAIAC) Aerosol Optical Depth (AOD) dataset. The process of filling missing data in a high-resolution dataset using coarse-resolution dataset is essentially a downscaling process. In this study, the coarse-resolution dataset used as input variables in the neural network are the meteorology variables, daily mean AOD, coordinates, and elevation.

The above-mentioned studies using a variety of modeling approaches had lower performance results or lower rates of missing data. To the best of our knowledge, this is the first study that employs advanced deep learning techniques for robust downscaling of tropospheric air quality to generate a high spatiotemporal resolution dataset.

## 3. Data

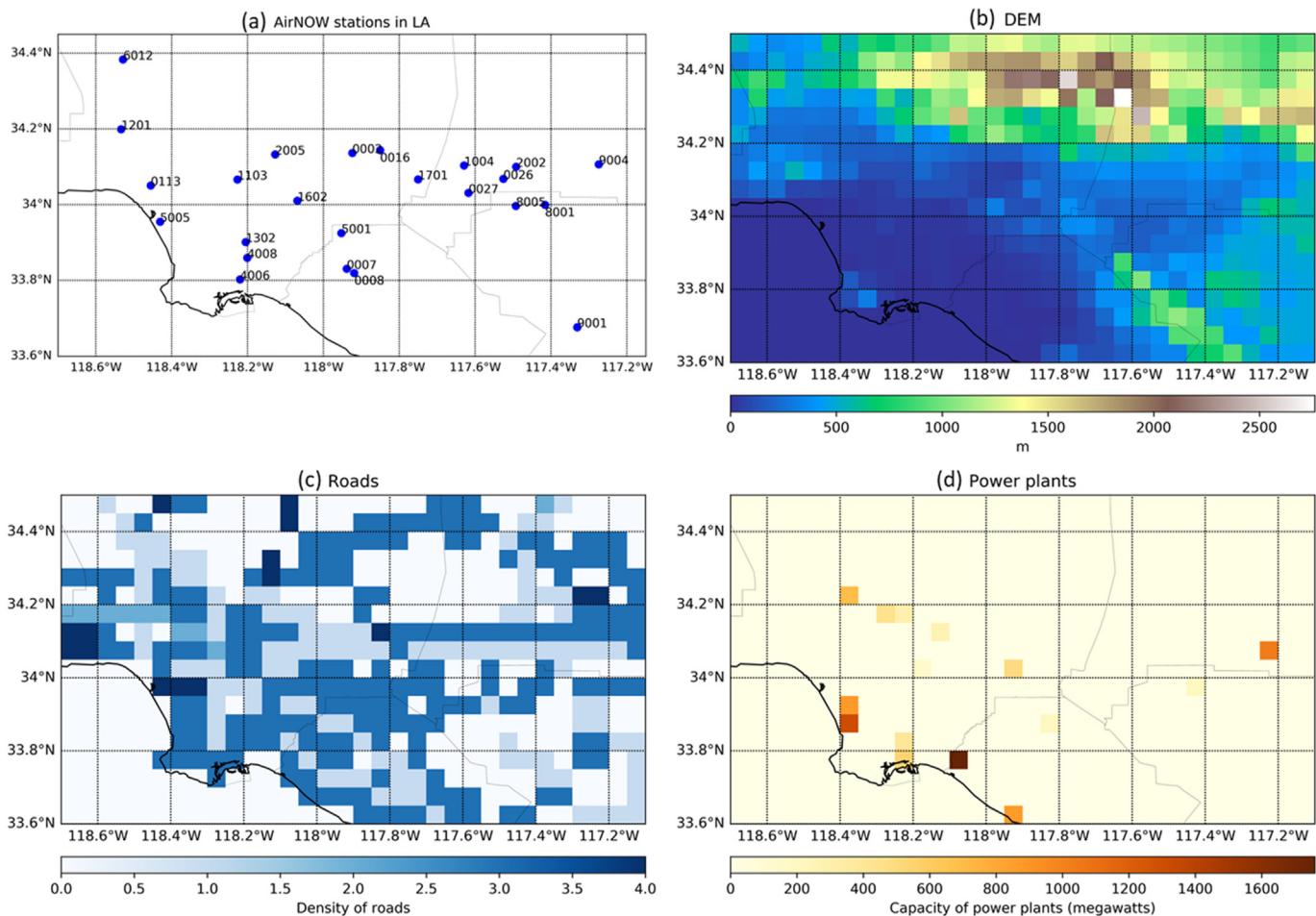
For model training and evaluation, the predictors are NO<sub>2</sub> observed by ground-level stations, station location (longitude, latitude), boundary layer height, surface pressure, surface net solar radiation, 2 m temperature, 10 m UV wind components, and the predictand is the 5 km tropospheric NO<sub>2</sub>. Specifically, the ground-level NO<sub>2</sub> observations are from EPA AirNOW, the surface meteorological variables are from ERA-Interim, and the tropospheric NO<sub>2</sub> is from the TROPOMI's daily overpass from May 2018 to August 2019. The study area is the larger LA area (118.7°W to 117.1°W, 33.6°N to 34.5°N), and is comprised of urban and suburban areas surrounding LA.

### 3.1. TROPOMI

The TROPOspheric Monitoring Instrument (TROPOMI) is a newly launched (October 2017) instrument on the Sentinel-5 Precursor (S5P) mission that provides observations for atmospheric conditions on air quality, climate and the ozone layer (Veefkind et al., 2012). The TROPOMI Level 2 data products provide the information related to cloud, aerosol, and air quality. In this study, we used the tropospheric NO<sub>2</sub> vertical column densities (VCDs) in the unit of molecules cm<sup>-2</sup> (molec cm<sup>-2</sup>), which is one of the variables provided in the Level 2 product. The NO<sub>2</sub> retrieval algorithm used for TROPOMI NO<sub>2</sub> product was adapted from the algorithm developed for the Dutch OMI NO<sub>2</sub> (DOMINO) within the European Union's Quality Assurance for Essential Climate Variables (QA4ECV) project (Boersma et al., 2011; van Geffen et al., 2019). The algorithm utilized the retrieval of the total NO<sub>2</sub> slant column density from the Level-1b product, and separated the total NO<sub>2</sub> slant column density into a stratospheric and a troposphere part based on a data assimilation system of the TM5-MP chemical transport model. Then the troposphere slant column density was transformed into the tropospheric vertical column density using a look-up table of altitude-dependent air-mass factors and the daily vertical distribution of NO<sub>2</sub> from TM5-MP (van Geffen et al., 2019). Regarding spatial resolution the average pixel size of NO<sub>2</sub> data is 3.5 × 7 km<sup>2</sup> and changed to 3.5 × 5.5 km<sup>2</sup> since August 6, 2019. For TROPOMI, quality-control has been performed before the analysis. A flag, namely quality assurance value (qa\_value), for each ground pixel indicates the status and quality of the retrieval result, ranging from 0 (no output) to 1 (all is well). We selected the valid pixel, for which the qa\_value was above 0.75, to exclude part of the scenes covered by snow/ice, errors and problematic retrievals. We use tropospheric NO<sub>2</sub> columns with cloud radiance fractions less than 0.3. The overpass of TROPOMI in the larger LA area is around 21:00 UTC every day (14:00 local time).

### 3.2. AirNOW

The US Environmental Protection Agency's (EPA) nationwide, voluntary program, AirNow ([www.airnow.gov](http://www.airnow.gov)), provides real-time air



**Fig. 1.** (a) AirNOW NO<sub>2</sub> stations in LA, (b) DEM, (c) Count of major roads, (d) Power plant capacity.

quality data and forecasts to protect public health across the United States, Canada, and parts of Mexico. AirNow receives real-time ozone and PM<sub>2.5</sub> data from over 2500 monitors and collects air quality forecasts for more than 500 cities. In this study, we used the variable NO<sub>2</sub>, which is computed as NOx-NO in the unit of parts per billion (ppb). There are 24 available NO<sub>2</sub> sensors within the study area (Fig. 1a). Time of the measurement is reported in GMT and corresponds to the beginning of the sampling period. We selected tropospheric columns that have a TROPOMI pixel centered within 5 km of the AirNOW station and are measured under clear-sky situations. The AirNOW data are selected within half an hour for local time, which covered the satellite overpass time to ensure meaningful comparisons with the satellite measured values.

### 3.3. Meteorological variables

To assist building the connection between ground-level NO<sub>2</sub> and tropospheric NO<sub>2</sub>, meteorological variables from ERA-Interim are used. ERA-Interim is a climate reanalysis dataset, covering the period from 1979 to August 31, 2019. The ERA-Interim dataset contains atmospheric and surface parameters, including 6-hourly atmospheric fields on model levels, pressure levels, potential temperature and potential vorticity; and 3-hourly surface fields and daily vertical integrals (Berrisford et al., 2011). The selected variables from ERA-Interim include boundary layer height, surface pressure, surface net solar radiation, 2 m temperature, and 10 m UV wind components. In this study, we use the hourly forecast dataset with a horizontal resolution of 0.125°, re-gridded from the original dataset.

### 3.4. Geographic variables

In this study, geographic variables were also used as input predictors, including digital elevation model (DEM), locations of major roads, and capacity of power plants. The reason for including these geographic variables is that the main source of NO<sub>2</sub> emissions are from road traffic, power plants, and off-road equipment, and the aggregation of NO<sub>2</sub> emissions along the vertical dimension is correlated to the elevation of a particular impacted area (Fenn et al., 2003). The digital elevation dataset used for the LA area is the subset of Global 30 Arc-Second Elevation (GTOPO30) data. GTOPO30<sup>1</sup> has a spatial resolution of approximately 1 km and is aggregated in average to the 5 km regular grids used in this study (Fig. 1b). The major roads are downloaded from the 2019 TIGER/Line Geodatabases of Census Bureau.<sup>2</sup> Only primary roads are used among all classes, and the number of roads is counted for each grid in the designated 5 km regular grids (Fig. 1c). The operable power plants are downloaded from the US Energy Information Administration,<sup>3</sup> and the capacity of power plants is aggregated for each grid in the 5 km regular grids (Fig. 1d).

## 4. Method

Fig. 2 illustrates the workflow of our downscaling methods. Before training, the TROPOMI tropospheric NO<sub>2</sub> is preprocessed into regular

<sup>1</sup> Global 30 Arc-Second Elevation (GTOPO30) Digital Object Identifier (DOI) number: /10.5066/F7DF6PQS

<sup>2</sup> <https://www2.census.gov/geo/tiger/TIGER2019/ROADS/>

<sup>3</sup> [https://www.eia.gov/maps/layer\\_info-m.php](https://www.eia.gov/maps/layer_info-m.php)

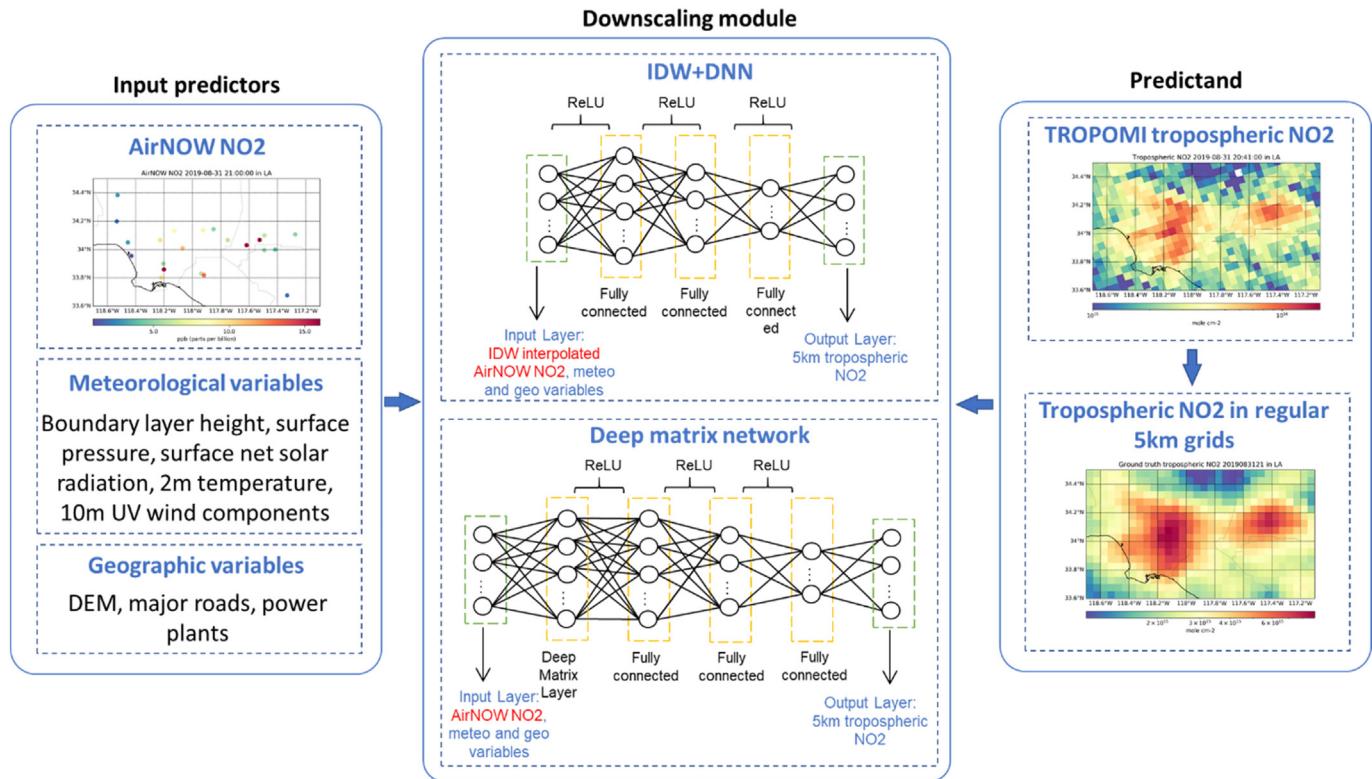


Fig. 2. Workflow of NO<sub>2</sub> downscaling.

5 km grids (Section 4.1). In addition, a preliminary statistical analysis is conducted to examine the correlation between TROPOMI tropospheric NO<sub>2</sub> and AirNOW NO<sub>2</sub> based on spatiotemporal collocation of the two observations (Section 4.2). The two comparing models that we developed and explored are introduced in Sections 4.3 and 4.4. And the training methodology is described in Section 4.5.

#### 4.1. TROPOMI preprocessing

The TROPOMI tropospheric NO<sub>2</sub> data is defined on irregular grids dependent on the satellite swath, and its grid resolution changed from  $3.5 \times 7 \text{ km}^2$  to  $3.5 \times 5.5 \text{ km}^2$  since August 6, 2019. As the proposed training process requires the input data to have a consistent data structure, data needs to be resampled on a regular grid structure. Since resampling using spatial interpolation can smooth out and even remove relevant structures, the initial data is resampled appropriately by searching for the nearest grid points within a distance that incrementally increases until at least one nearest point is found. If there are multiple points within a certain distance range, a mean value is calculated. This approach preserves the spatial adjacency of grid nodes for a large proportion of the nodes, which is important to facilitate proper learning of spatial correlations (Fig. 3).

#### 4.2. Correlation between TROPOMI and AirNOW NO<sub>2</sub> observations

Previous studies discovered that an empirical relationship exists between surface NO<sub>2</sub>, boundary layer height, and tropospheric NO<sub>2</sub> column (Dieudonné et al., 2013; Lorente et al., 2019), especially in urban conditions. The empirical relationship is a multivariate linear Eq. (1):

$$\text{Tropospheric NO}_2 \text{ column} = K(a_1 \times h(C-a_2) + a_3(C-a_4)) \quad (1)$$

where K is the constant factor that converts  $1\mu\text{g}/\text{m}^3$  in a deep boundary layer of 1 km into a column of  $1.31 \times 10^{15} \text{ molec.cm}^{-2}$ , h is the boundary layer height from ECMWF, and C is the surface NO<sub>2</sub> concentration. The

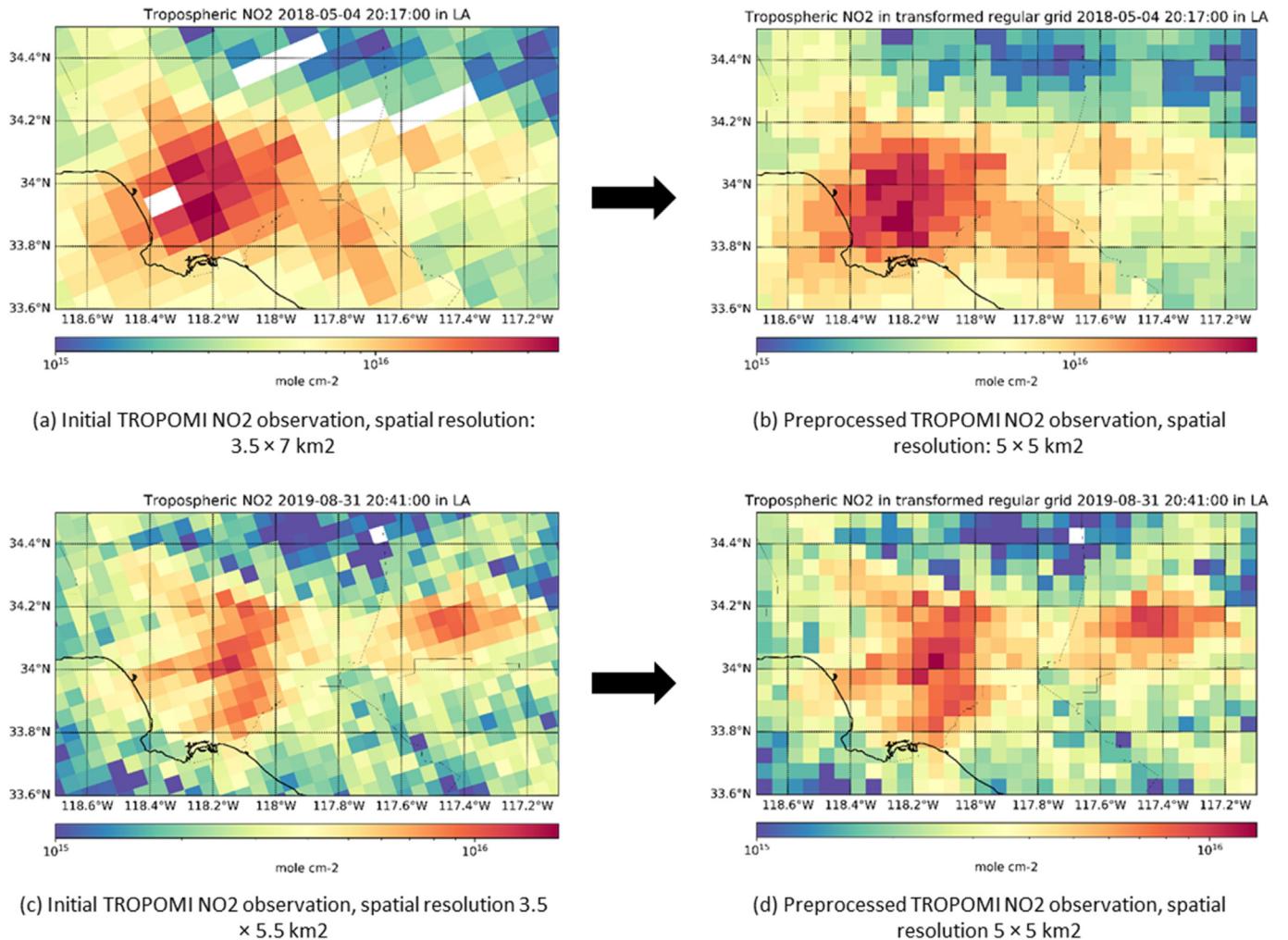
scaling factors ( $a_1 - a_4$ ) are determined by fitting the tropospheric NO<sub>2</sub> columns against NO<sub>2</sub> surface measurements for different boundary layer height classes. The NO<sub>2</sub> columns scale progressively with increasing boundary layer height.

Based on the empirical relationship, we compared the TROPOMI tropospheric vertical column of nitrogen dioxide against in situ (AirNOW) NO<sub>2</sub> measurements taken on the 24 stations within the study area. The hourly AirNOW NO<sub>2</sub> measurements closest in time (within 30 mins) to the TROPOMI overpass were selected, and the closest TROPOMI grids to AirNOW stations (within 5 km) were selected. The comparison is conducted on each station individually. Taking one station (id: 060371201) as an example, the timeseries of AirNOW measured NO<sub>2</sub> and the TROPOMI tropospheric NO<sub>2</sub> column show a good agreement regarding the temporal pattern, which generally increases in winter and decreases in summer (Fig. 4 a and b). After transforming the ground-level NO<sub>2</sub> to tropospheric NO<sub>2</sub> based on Eq. (1), the transformed AirNOW tropospheric NO<sub>2</sub> is also showing a relatively high correlation with the TROPOMI tropospheric NO<sub>2</sub>, with a slope value of 0.9999 (Fig. 4c).

Most stations show good correlations ( $0.3 < R^2 < 0.7$ ) between AirNOW ground-level NO<sub>2</sub> and TROPOMI's tropospheric NO<sub>2</sub> column in the nearest overpass, and the Root Mean Square Error (RMSE) remains less than  $5.2 \times 10^{15} \text{ molec.cm}^{-2}$  (Fig. 4d). The  $R^2$  value is generally lower than the ones reported by Lorente et al. (2019), which was 0.88 but it covered only 28 samples. Significant differences between the ground-based NO<sub>2</sub> observations and satellite observed NO<sub>2</sub> column densities are most observed in and around downtown LA (Station-4008:  $R^2 = 0.28$ , RMSE = 5.29) and downtown San Bernardino (Station-0027:  $R^2 = 0.28$ , RMSE = 3.17).

#### 4.3. IDW + DNN

In order to convert the station based AirNOW NO<sub>2</sub> to the distribution maps of NO<sub>2</sub> concentrations, a spatial interpolation was conducted; and in this study, we used the Inverse Distance Weighted (IDW)



**Fig. 3.** TROPOMI NO<sub>2</sub> resampled to 5 × 5 km<sup>2</sup> regular grids.

interpolation method as an exploration. IDW is an interpolation method that computes the score of query points based on the scores of their k-nearest neighbors, weighted by the inverse of their distances. As each query point is evaluated using the same number of data points, this method allows for strong gradient changes in regions of high sample density while imposing smoothness in data sparse regions. Note that different interpolation methods and specifying parameters may return different output values. For example, Kriging is an advanced geostatistical procedure that generates an estimated surface based on the investigation of the spatial behavior represented by the known values using variograms. However, Kriging did not produce a result better than IDW, because in our study each time step has only 24 station-based measurements from AirNOW, which is difficult to fit a valid variogram.

In IDW, the calculation of an unknown value at a point  $x$  is based on the k-nearest neighbors as the following Eq. (2):

$$u(x) = \frac{\sum_{i=1}^k w_i(x) u_i}{\sum_{i=1}^k w_i(x)}, \text{ if } d(x, x_i) \neq 0 \quad (2)$$

where the weights for each neighboring point  $w_i(x)$  is calculated as:

$$w_i(x) = \frac{1}{d(x, x_i)^p} \quad (3)$$

Here, Point  $x$  is the unknown point, Point  $x_i$  is a known neighboring point,  $u_i$  is the known value of the interpolating variable for the known

point  $x_i$ ,  $d$  is the distance between the known point  $x_i$  and the unknown point  $x$ , and  $p$  is a positive real number, called the power parameter. So, weight decreases as distance increases from the interpolated points. Noted that we adopt the Euclidian distance metrics, a power parameter of 2, and 6 nearest neighboring points.

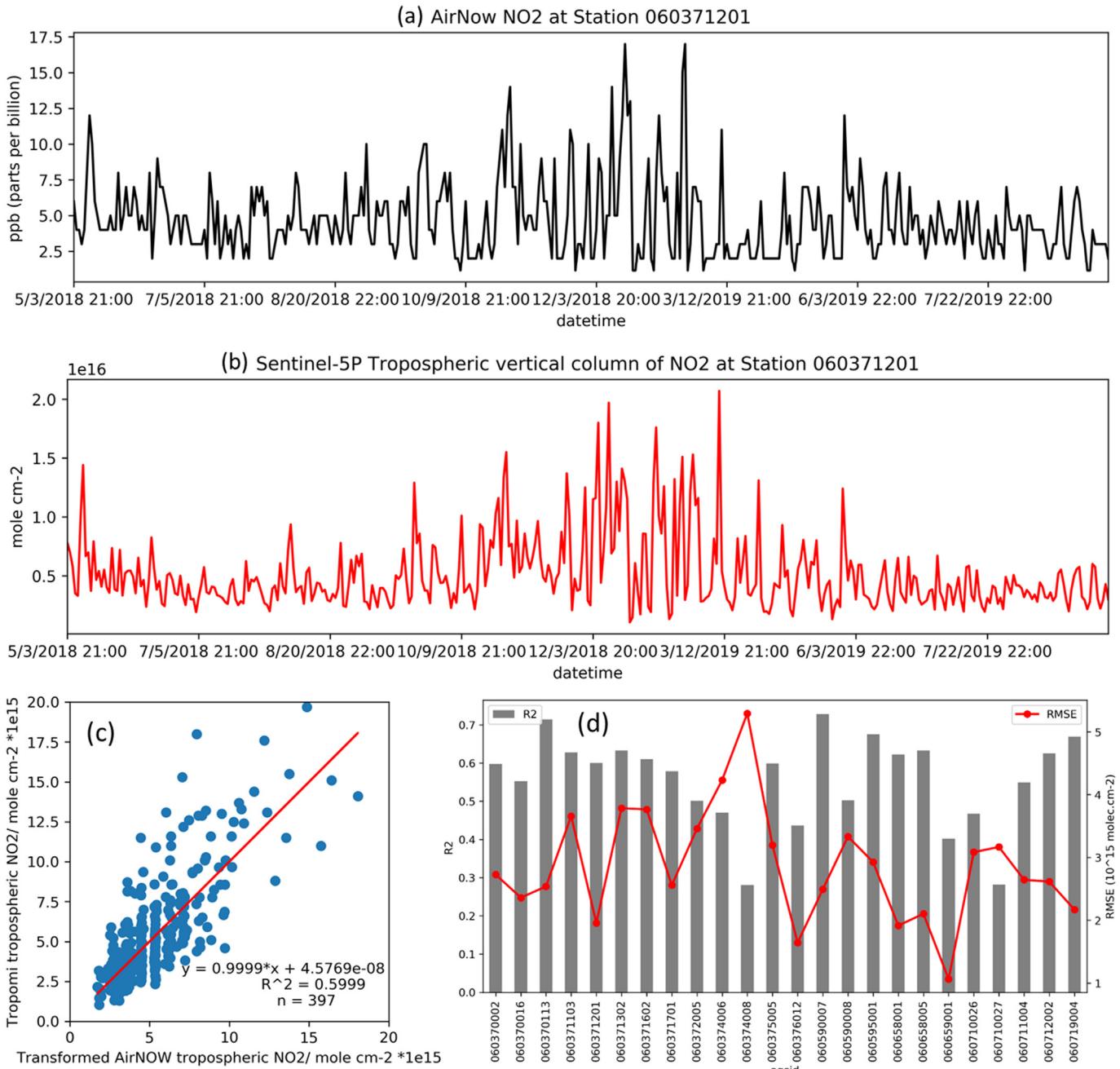
The IDW interpolated NO<sub>2</sub> maps are then trained pixel-wise as the input of a deep neural network (DNN). The network was a deep neural network composed of three fully connected layers, also known as multilayer perceptrons (MLP) (Goodfellow et al., 2016). The input layer represented the input features for each regular grid, which include the location, the IDW interpolated AirNOW NO<sub>2</sub> value, and the meteorological features from ERA-Interim. The output layer was a single unit representing the predicted yield. The DNN model can be simplified as:

$$y = Wx + b \quad (4)$$

which learns the non-linear relationships between flattened predictor vectors and flattened predicted vectors. The number of hidden layers and hidden neurons were two important hyper-parameters of the network defined through an empirical process in which the performance of various network architectures, selected based on the domain knowledge, were evaluated.

#### 4.4. Deep Matrix Networks (DMN)

In our proposed architecture, there are four layers in the deep matrix networks (DMN), where the first layer is a Deep Matrix Layer, and the



**Fig. 4.** Timeseries of (a) AirNOW measured  $\text{NO}_2$  at Station 060371201 and (b) the corresponding TROPOMI overpass. (c) Scatter plot and the results of the transformed AirNOW tropospheric  $\text{NO}_2$  based on Eq. (1) and the TROPOMI observation. (d)  $R^2$  and RMSE between AirNOW ground-level  $\text{NO}_2$  and TROPOMI's tropospheric  $\text{NO}_2$  column for all 24 stations within the study area.

rest are fully connected layers (Fig. 5). Let  $X^l \in R^{N_i \times N_f}$  be the input matrix where  $N_i$  is the number of input  $\text{NO}_2$  stations and  $N_f$  is the number of input predictors. The deep matrix layer  $l_1$  is connected to the first fully connected layer ( $l_2$ ) by

$$X^{l2} = \sigma(W^{l1}X^{l1} + b^{l1}) \quad (5)$$

here  $W^{l1} \in R^{N_o \times N_i \times N_f}$  is the weights to be learned in the deep matrix layer, where  $N_o$  is the number of output  $\text{NO}_2$  grids;  $b^{l1} \in R^{N_f}$  represents the biases for each feature; and  $\sigma(\cdot)$  is the activation function acting on each element of matrix.

The computation within the deep matrix layer is described as follows. For each feature, a weight matrix is initiated, and a matrix

multiplication is conducted using the weights and the input matrix for the specific feature. Each matrix multiplication result is then appended with an initiated bias matrix. The results of the matrix multiplication plus bias for all features are then concatenated as one output matrix. The output shape of the deep matrix layer is  $X^{l2} \in R^{N_o \times N_f}$ . The deep matrix layer is followed by three fully connected layers that are similar to the configurations of the IDW + DNN, where the output layer is  $y \in R^{N_o}$ . We use the Rectified Linear Unit (ReLU) as the activation function at each hidden layer, including the deep matrix layer and the fully connected layers. For training the parameters of weight matrices on each layer, we use back propagation to update the model parameters with batches.

The difference between IDW + DNN and the DMN is that IDW + DNN learns the relationship between input predictors and

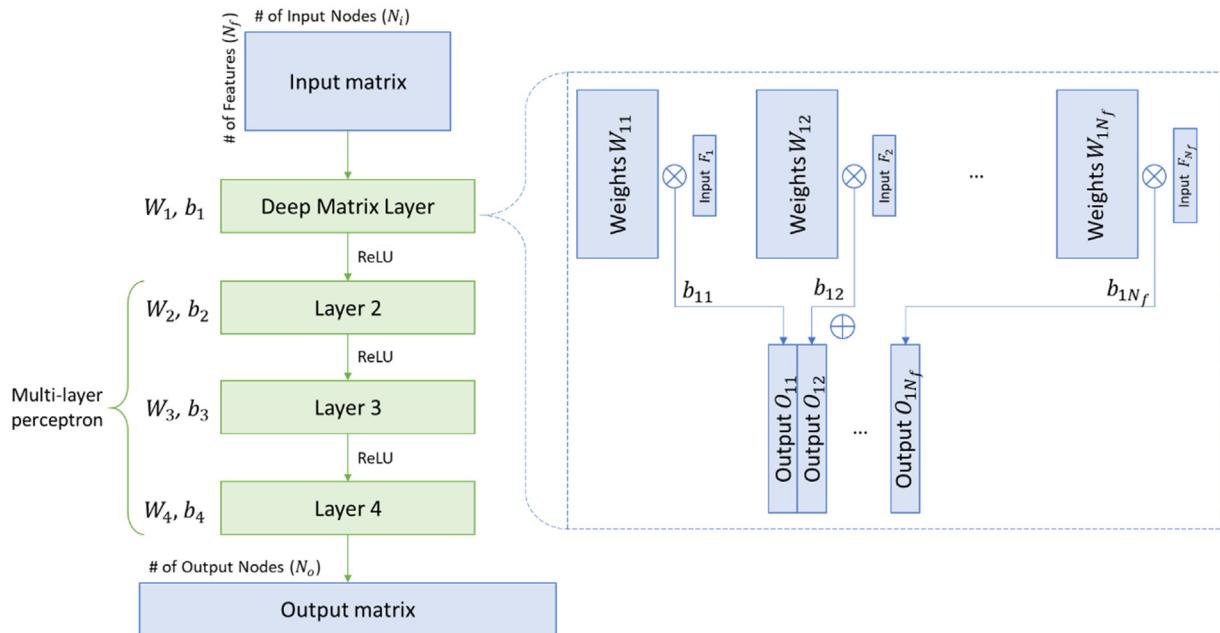


Fig. 5. The architecture of DMN.

output predictand for each high-resolution pixel, whereas DMN learns this relationship for the direct mapping between the discrete stations and the high-resolution matrix. The direct mapping does not involve the interpolation errors resulted from IDW. Based on this characteristic, DMN can be applied to any tasks that directly interpolate values from discrete points to map distributions.

#### 4.5. Training methodology

The training data ranges from May 1, 2018 to April 30, 2019, and the testing data ranges from May 1, 2019 to August 30, 2019. The input predictors and output predictand of the training dataset are normalized per feature using the MinMaxScaler tool from scikit-learn (Pedregosa et al., 2011), and the fitted scalers are then applied to the testing dataset for evaluation. For measuring error magnitude between predictions and true values, we use the Mean Squared Error (MSE) as the loss function, and RMSE and MAE as the evaluation functions. Given that  $y_i$  and  $\hat{y}_i$  represent the target NO<sub>2</sub> values and predicted NO<sub>2</sub> values at node  $i$ , with  $i \in [i, \dots, n]$  indexing the nodes of the high-resolution grid, the MSE, RMSE, and MAE are calculated as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (8)$$

Both models have been realized and evaluated in PyTorch (Paszke et al., 2019). Optimization is performed using the stochastic gradient descent with an initial learning rate of  $10^{-3}$ , which is further tuned by reducing the value till the validation loss does not decay within 5 epochs. To guarantee a proper convergence of the models, we train for 1500 epochs in each of the cross-validation experiments, and both training and validation losses showed only minor variations.

## 5. Experiments

To compare different model architectures with respect to downscaling performance, we consider sample-wise deviations between target predictands and model predictions and investigate the extent to which the predictions depend on particular predictors. To examine the importance of different types of predictors, the models are trained with four different predictor configurations, including the station based NO<sub>2</sub> and location only; providing boundary layer height predictor; providing more meteorological predictors, or the full set of parameters. The predictor settings are detailed in Table 1 and the performance is compared in Section 5.1. Individual feature importance is analyzed by perturbing a specific feature and calculating the RMSE change, and the results are reported in Section 5.4. Distinctions of this strategy for different model comparisons arise for IDW + DNN model. In the case of IDW + DNN, we use the interpolated AirNOW NO<sub>2</sub> in the target 5 km regular grids as input, whereas in other models, we use the discrete AirNOW NO<sub>2</sub> as node input.

#### 5.1. Model comparison

The spatially averaged RMSE and MAE on the validation data are illustrated in Table 2, confirming that both model architecture and predictor selection have a considerable effect on model performance. The

**Table 1**  
Predictor configurations for model trainings with varying combinations of predictors.

	Predictors (X)	Predictand (Y)
Experiment 1	Latitude, longitude, AirNOW NO <sub>2</sub>	TROPOMI NO <sub>2</sub>
Experiment 2	Latitude, longitude, AirNOW NO <sub>2</sub> , boundary layer height	TROPOMI NO <sub>2</sub>
Experiment 3	Latitude, longitude, AirNOW NO <sub>2</sub> , boundary layer height, surface pressure, surface net solar radiation, 2 m temperature, 10 m UV wind components	TROPOMI NO <sub>2</sub>
Experiment 4	Latitude, longitude, AirNOW NO <sub>2</sub> , boundary layer height, surface pressure, surface net solar radiation, 2 m temperature, 10 m UV wind components, DEM, major roads, power plants	TROPOMI NO <sub>2</sub>

**Table 2**

Comparison of validation accuracies for model variants with varying combinations of input predictors.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
IDW + DNN	RMSE: 2.3449	RMSE: 1.9252	RMSE: 1.4784	RMSE: 1.4134
	MAE: 1.7121	MAE: 1.3727	MAE: 1.0453	MAE: 1.0191
DMN	RMSE: 1.8837	RMSE: 1.4813	RMSE: 1.2651	RMSE: 1.2101
	MAE: 1.3587	MAE: 1.0225	MAE: 0.8662	MAE: 0.8245

weaker model is IDW + DNN, showing larger overall errors. DMN takes advantage of the local parameterization and achieves better results than the IDW + DNN. DMN not only reduces the RMSE and MAE but can also take advantage of additional predictors more effectively. IDW + DNN responds with an increased tendency of overfitting, whereas DMN achieves a reduction in deviation score when supplied with more predictors. By comparing the RMSE across the four experiment settings, Experiment 1 is suffering from the highest error rate, while adding input predictors of surface meteorological status, elevation, and built infrastructures (Experiment 4) improves the predictability by ~40%. Notably that the accuracy improvement brought by adding boundary layer height is ~18% and ~21% for IDW + DNN and DMN respectively, indicating that boundary layer height is an important factor that assist the understanding of the non-linear relationship between ground-level NO<sub>2</sub> and tropospheric NO<sub>2</sub>.

### 5.2. Analysis of downscaled NO<sub>2</sub> hotspots

For reasons of conciseness, we limit the comparison to outputs of the best-performing configuration – Experiment 4, for both models. The first example is given for August 31, 2019, 21:00 UTC. Fig. 6 (a) and (b) show the original AirNOW observed ground-level NO<sub>2</sub> and the TROPOMI tropospheric NO<sub>2</sub> column density. Differences in NO<sub>2</sub> values indicate that the discrete station based NO<sub>2</sub> values fail to properly capture the local variability in the northern rural places of the study area. The results of IDW + DNN and DMN predictions are shown in Fig. 6 (c) and (d), respectively. IDW + DNN tends to not reconstruct the spatial pattern of high NO<sub>2</sub> values, especially when there is a significant difference between AirNOW and TROPOMI NO<sub>2</sub>. DMN, in contrast, uses both ground-level NO<sub>2</sub> observation and global information about the meteorological status and elevation, is able to better replicate the tropospheric NO<sub>2</sub>. Especially over the elevated area in the northern part of the study area, IDW + DNN introduced more errors that lead to overestimation in the prediction.

The second example is for June 4, 2019 at 22:00 UTC, when a high level of NO<sub>2</sub> value is observed by AirNOW, especially in the urban region of San Bernardino county (Fig. 6 e-h). Contrary to the concrete ground-level NO<sub>2</sub> observations, TROPOMI exhibits another high NO<sub>2</sub> concentrated cluster in Los Angeles County, especially in the urban region and along the major roads. IDW + DNN predicts the correct spatial pattern of tropospheric NO<sub>2</sub>, but predicted values deviate more from the ground truth than the DMN prediction. The deviations are more obvious in the highly elevated area in the northern and southern regions of the study area, where NO<sub>2</sub> values are overestimated by the IDW + DNN model.

### 5.3. Spatiotemporal distribution of prediction errors

To examine the spatial distribution of downscaling errors, we calculated the magnitude-specific deviation measures that is aggregated over the testing temporal range. The systematic deviation magnitude provides a measure for how much the respective models over or underestimate NO<sub>2</sub> values.

Fig. 7 shows the spatial distribution of RMSE and magnitude difference between the prediction and true value using the IDW + DNN method across all four model variants with varying combinations of

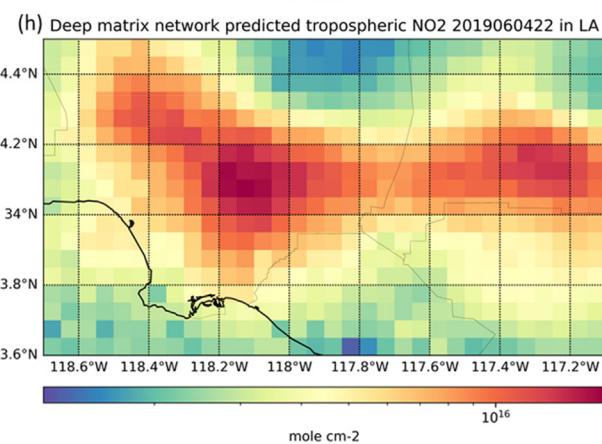
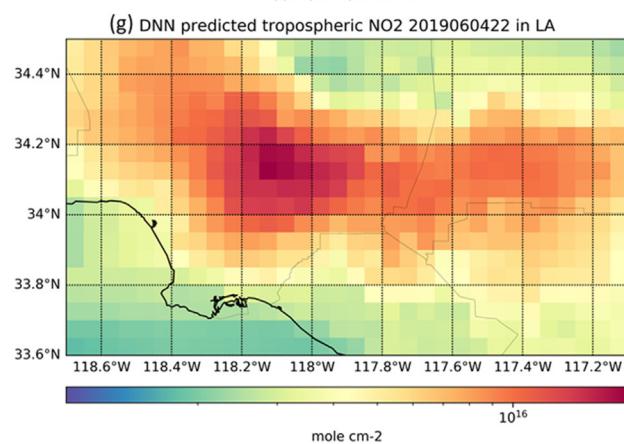
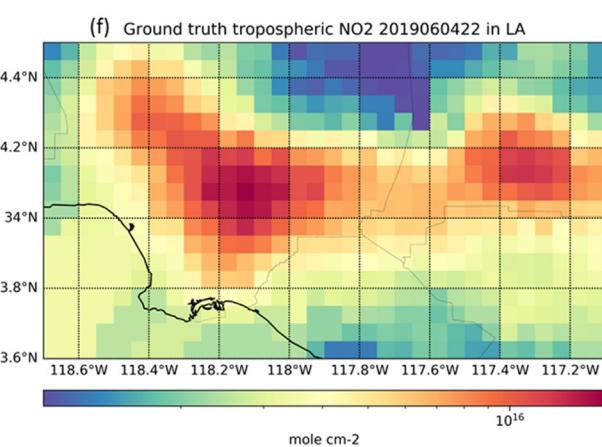
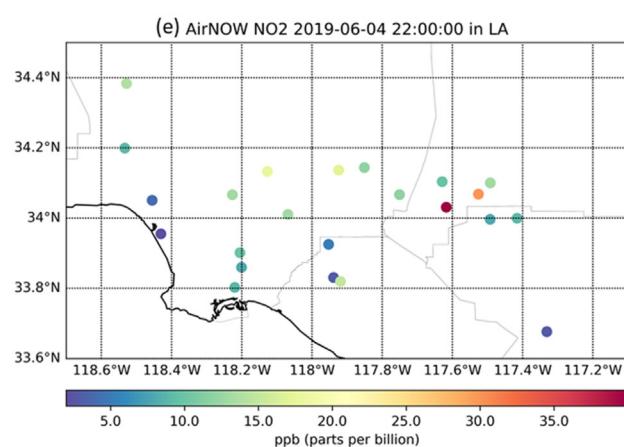
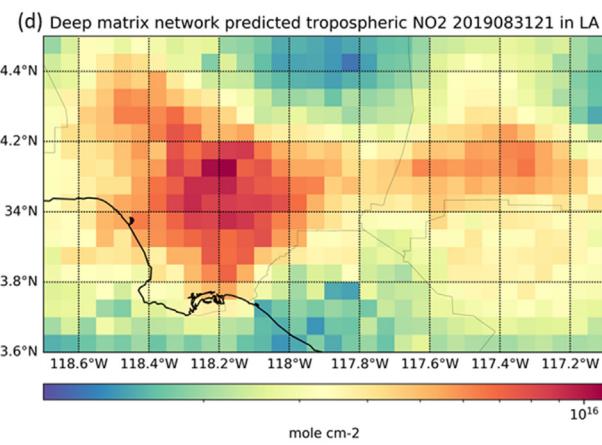
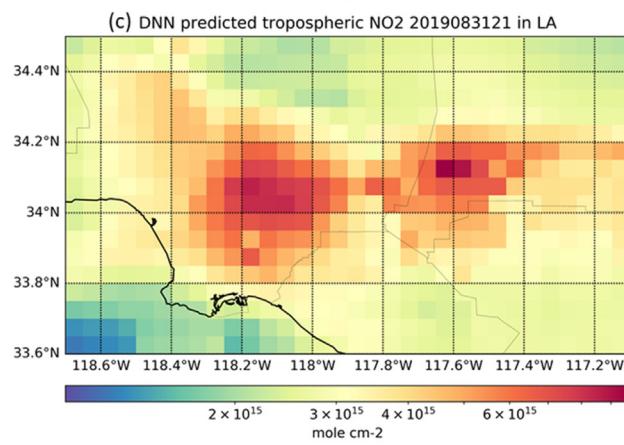
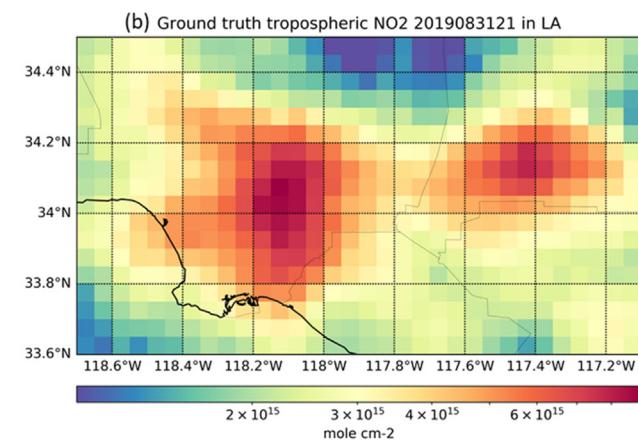
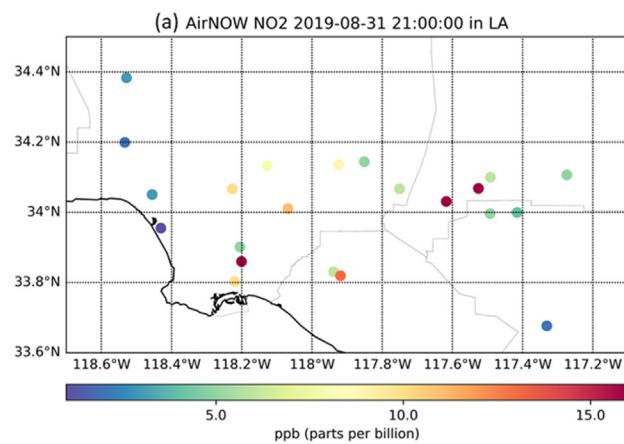
input predictors. Based on the RMSE distributions, it can be observed that the predicted tropospheric NO<sub>2</sub> in the urban regions of Los Angeles County and San Bernardino County are not well captured. Based on the magnitude difference distributions, overestimation of tropospheric NO<sub>2</sub> is observed for most grids in the IDW + DNN predictions. The high overestimation values are clustered in the urban regions of Los Angeles County and San Bernardino County. The stripe patterns shown in the spatial distributions of RMSE error and magnitude difference are generally caused by the errors induced by IDW interpolation. It is also observed that there are two neighboring grids in LA downtown that generally have a large gap in the values of ground-level NO<sub>2</sub>, leading to outlying overestimation and underestimation for the two grids respectively. Adding surface meteorological status helped corrected a proportion of the overestimation in the urban areas, whereas adding DEM and built infrastructures helped corrected the overestimation on the outlying point of AirNOW NO<sub>2</sub> observation.

For the DMN method, the RMSE distributions show similar high-error spatial regions as the IDW + DNN predictions. The DMN shows a larger percentage of grids suffering from overestimation, but the magnitude difference is generally smaller than the ones from IDW + DNN. The spatial clustering effect is also smoother than the one from IDW + DNN, with a clearer spatial distribution that the urban areas have higher overestimation of tropospheric NO<sub>2</sub>. Notably that there is no such outlying points of overestimation and underestimation in LA downtown in the predictions generated by the DMN, indicating that the DMN is a more spatially generalized model than IDW + DNN. The DMN is able to better replicate extreme transitions in magnitude, occurring on small spatial scales, which results in smaller RMSEs and magnitude differences.

To examine the temporal distribution of downscaling errors, we calculated the RMSE and magnitude difference (i.e., bias) between the predictions and the ground truth for each time step in the testing data for Experiment 4, which has the best configuration of input predictors (Fig. 8). The temporal distribution of RMSE and bias errors for both IDW + DNN and DMN showed similar patterns, but DMN outperforms throughout the testing period. The RMSE of IDW + DNN ranges between 1.1 and 5.2 whereas the RMSE of DMN ranges between 0.3 and 3.5. The highest error for both IDW + DNN and DMN occurred on June 10, 2019. Examining across the testing period, DNN suffered from higher overestimating errors in June 2019 than the other months. This behavior can be explained by the correlation between the surface NO<sub>2</sub> from AirNOW and the tropospheric NO<sub>2</sub> from TROPOMI, where the monthly mean R<sup>2</sup> between these two datasets was 0.5 for May 2019, 0.49 for June 2019, 0.53 for July 2019, and 0.56 for Aug 2019.

### 5.4. Analysis of feature importance

For the model configuration, which was trained on the full set of predictors (Experiment 4), we also investigate the importance of predictors by perturbing the model inputs from the validation data set by randomly shuffling single predictors, and then measure the change in the predictor error that is caused by the perturbation. Fig. 9 illustrates the change of RMSE when perturbing a specific predictor in the two models. Each perturbation is explored in ten experiments and the box plot shows the median value and interquartile range of RMSE changes. In good agreement with expectations, AirNOW NO<sub>2</sub> has the largest effect on model performance for both model architectures in the comparison, indicating that the models in fact use mainly the information on ground-level NO<sub>2</sub> for downscaling. IDW + DNN also leverages latitude, U-wind, DEM, and power plant as major features. The capability of power plants is more recognized by IDW + DNN than major roads. The combination of U wind component at 10 m and latitude composes the wind impacts from both longitude and latitude dimensions. The DMN model utilizes Temperature at 2 m and DEM as input predictors with higher impact to the model's RMSE change, while the other input predictors show less impacts. Surprisingly, boundary layer height does



not necessarily improve the performance of the model. One possible assumption is that the boundary layer height variable might correlate with other meteorological variables from ERA-Interim.

## 6. Discussion

The meteorological variables used in this study are from ERA-Interim with  $0.125^\circ$  spatial resolution. ERA-Interim is only available till August 31, 2019 and has been superseded by the ERA5 reanalysis. Currently, ERA-5 has a resolution of  $0.25^\circ$ , and no other higher-resolution re-gridded datasets are available. We explored changing ERA-Interim to ERA-5 with the same temporal range to test the model performance, and the average RMSE decreased from 1–2 to 3–4 (unit:  $10^{15}$  molec. $\text{cm}^{-2}$ ). The decreased RMSEs indicated that the spatial patterns of the meteorological variables are too generalized for the deep learning models to learn. With the prospect that ECMWF will upgrade its reanalysis datasets to 5 km by 2025 (Pappenberger and Hewson, 2017), we can adapt the proposed models accordingly. ERA-5 has data available from 1979 to within five days of real-time, which will be much longer than the training and testing period of this study.

The trained model can leverage the AirNOW observations (hourly), the meteorological variables from the climate reanalysis dataset (hourly), and the other geographic variables (static) to predict the hourly tropospheric  $\text{NO}_2$  in 5 km regular grids. The predicted results can provide a fine-temporal-scale tropospheric  $\text{NO}_2$  trend for the larger LA area. In future research, the model will be adapted in other regional or continental areas since the data sources we selected are also available on a larger scale. In addition, the model can be used to downscale other air pollutants, such as  $\text{SO}_2$ , ozone, or  $\text{CO}$ , that are available as TROPOMI Level 2 products.

The current predictand is TROPOMI tropospheric  $\text{NO}_2$ , which is the satellite observation that has the highest possible resolution to date. NASA is launching a geostationary instrument, Tropospheric Emissions: Monitoring Pollution (TEMPO), to measure the same air pollutants as the ones in TROPOMI hourly during daytime in 2.1 km–4.7 km spatial resolution (Zoogman et al., 2017). The proposed models can be updated with the new satellite observations. With a high spatiotemporal resolution, our results will facilitate the study between air quality and health issues, and improve the understanding of the dynamic evolution of airborne pollutants. The high-resolution estimates will be capable of exploring the daytime diurnal evolution of trace gases and chemistry that influence air quality conditions.

Comparing to tropospheric  $\text{NO}_2$ , surface  $\text{NO}_2$  measurements and predictions have a more significant impact to human health. Although this paper presents a downscaling procedure from surface  $\text{NO}_2$  to tropospheric  $\text{NO}_2$ , but the DMN method can be used for various downscaling tasks where the low-resolution input is spatially sparse. The reason we did not use surface  $\text{NO}_2$  as the predictand is the unavailability of a long-term high-resolution surface  $\text{NO}_2$  measurements. One of our future research is to further downscale the 5 km tropospheric  $\text{NO}_2$  to surface  $\text{NO}_2$  at a street-level scale by integrating the ground-level discrete measurements (e.g., AirNOW), the satellite observations (e.g., TROPOMI), and high-resolution numerical predictions (e.g., WRF-CMAQ).

Despite the contributions of this paper and other related research, there are several aspects we can continue to explore and improve the downscaling accuracy for the future work. First, we will replace the ERA-Interim data with the latest ERA-5 with possibly higher resolution to enable the proposed downscaling methodology with longer applicable time and better performance. Second, we will explore the seasonal sensitivities of the models, and compare the model performance in

different topographies. Thirdly, we will explore the adaptabilities of the models on other airborne pollutants. And fourth, we will continue investigating methods and exploring new datasets to produce high-resolution  $\text{NO}_2$  at a surface level.

## 7. Conclusions

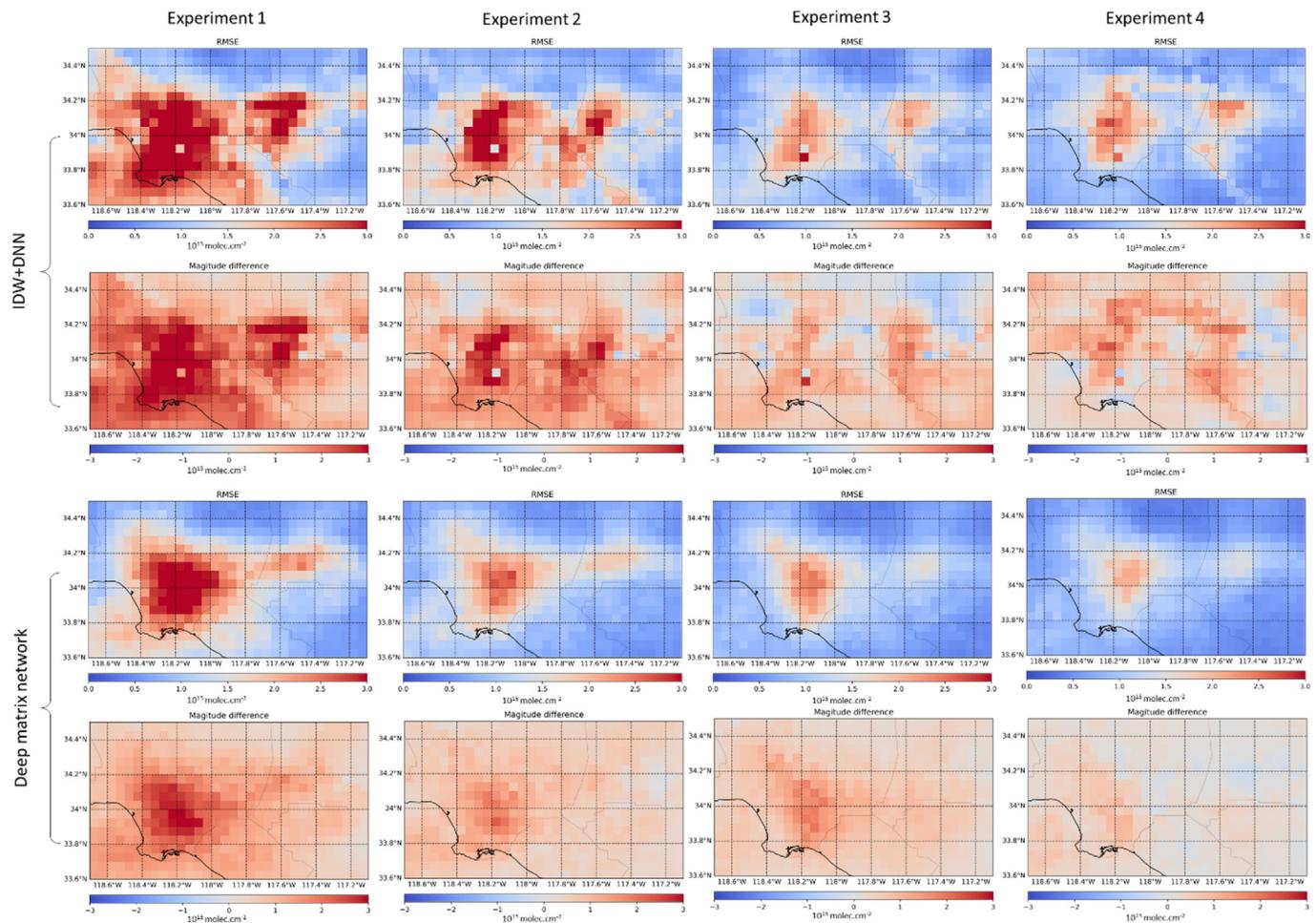
In this study, we proposed, compared, and evaluated two deep learning methods for downscaling discrete ground-level  $\text{NO}_2$  observations to estimate tropospheric  $\text{NO}_2$  column density. The two specific methods are 1) an integrated method between inverse weighted distance and a feed forward neural network (IDW + DNN), and 2) a deep matrix network (DMN) that maps the discrete AirNOW observations directly to the distribution of TROPOMI observations. We investigated the network performance using the larger LA area as a case study. Experiments showed that the prediction accuracy of the DMN is higher than what can be achieved with IDW + DNN. We attribute this to the distortion resulted from the initial IDW interpolation from the discrete AirNOW stations, whereas the DMN benefit from direct mapping schemes.

The input predictors of both proposed models include the locations of AirNOW stations, AirNOW  $\text{NO}_2$  observations, boundary layer height and other meteorological status, and geographic variables such as elevation, major roads, and power plants. We examined the model performance by incrementally adding more predictors by group, and adding all input predictors improves the accuracy by ~40% comparing to the minimum set of predictors (ground-level  $\text{NO}_2$  measurements and the corresponding station locations). Besides ground-level  $\text{NO}_2$  and the corresponding station locations, adding boundary layer height as an addition predictor improved the model accuracy ~18% and ~21% for IDW + DNN and DMN respectively. We also examined the feature importance by perturbing each input predictor. Results showed that the most important input predictor for both models is the ground-level  $\text{NO}_2$  measurements. In the perturbation experiment, unexpectedly, boundary layer height changes the RMSE only to a certain extent. This is different from the previous studies showing boundary layer height as an important linkage between ground-level  $\text{NO}_2$  and tropospheric  $\text{NO}_2$ . The reason of this difference is partly because the boundary layer height and the other meteorological variables are all from ERA-Interim, and variables might correlate with each other. Even if boundary layer height is perturbed in the experiment, other ERA-Interim meteorological variables can compensate with similar spatial patterns.

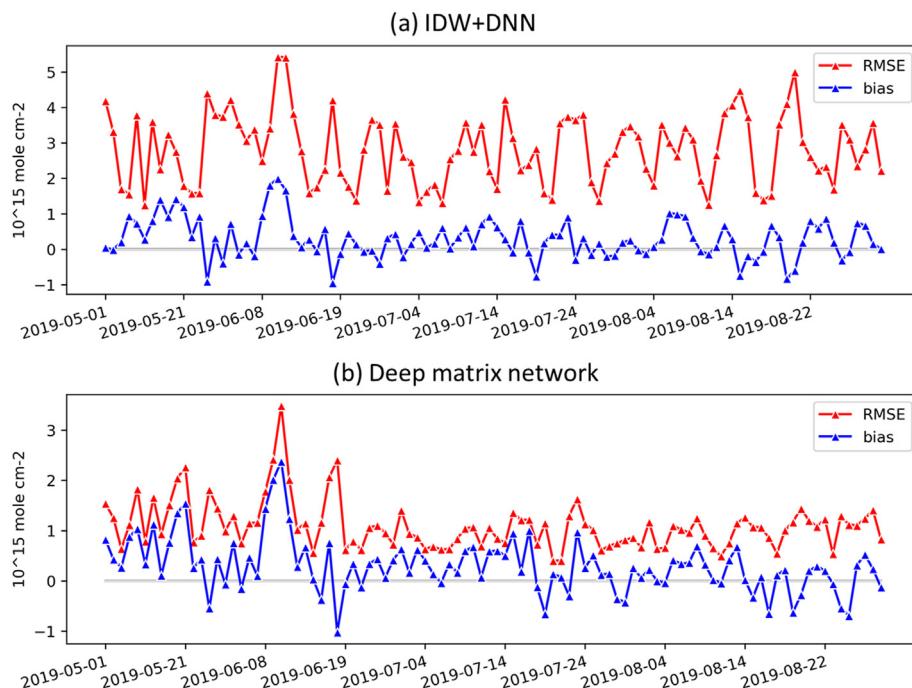
Both methods tended to overestimate tropospheric  $\text{NO}_2$  column density in most part of the study area, and adding geographic variables helped reducing the overestimation. We noticed some artifacts from IDW + DNN where two nearby grids in LA downtown showing outlying overestimation and underestimation, respectively. These two outlying values are inherited from the AirNOW observation. Adding surface meteorological status helped corrected a proportion of the overestimation in the urban areas, whereas adding DEM and built infrastructures helped corrected the overestimation on the outlying points of AirNOW  $\text{NO}_2$  observation. On the contrary, the DMN showed no such outlying nearby points of overestimation and underestimation in the predictions, indicating that the DMN is a more spatially generalized model than IDW + DNN. Although more pixels ended up with overestimation, the DMN resulted in a more smoothed spatial clustering effect, and the overestimation is much lower than the one in IDW + DNN.

The DMN model can be envisioned for operational purposes. With global climate reanalysis datasets and satellite air pollutant observations upgrading to a higher spatial resolution, this proposed method

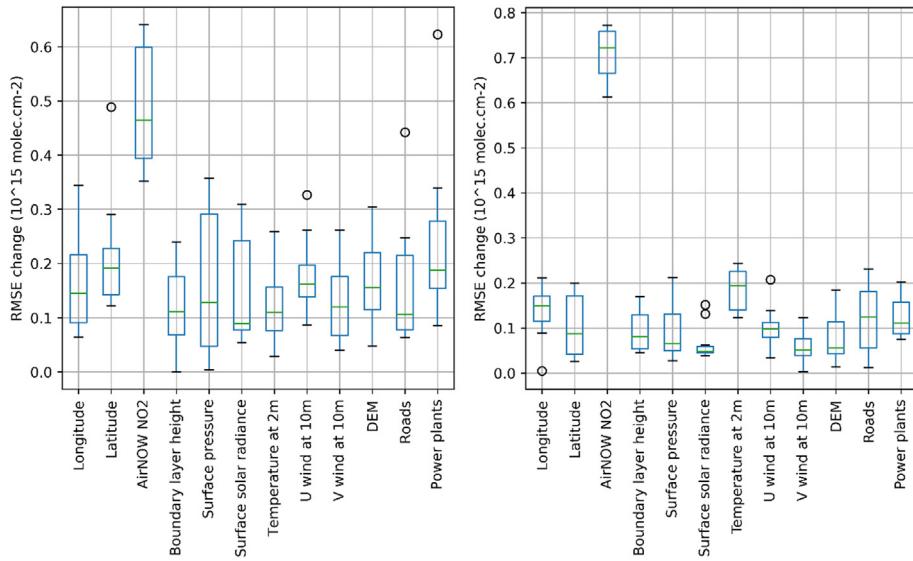
**Fig. 6.** (a-d) AirNOW observed ground-level  $\text{NO}_2$ , TROPOMI tropospheric  $\text{NO}_2$ , and model predictions for August 31, 2019, 21:00 UTC. Error of the IDW + DNN is RMSE = 4.1052, MAE = 3.5764; error of the deep neural network is RMSE = 1.7440, MAE = 1.1954. (e-h) AirNOW observed ground-level  $\text{NO}_2$ , TROPOMI tropospheric  $\text{NO}_2$ , and model predictions for June 4, 2019 at 22:00 UTC. Error of the IDW + DNN is RMSE = 4.5807, MAE = 3.5011; error of the deep neural network is RMSE = 1.46618, MAE = 1.1856. Unit:  $10^{15}$  molecules/ $\text{cm}^2$ .



**Fig. 7.** RMSE and magnitude difference for the four experiment configurations of input predictors.



**Fig. 8.** Timeseries of RMSE and bias in Experiment 4.



**Fig. 9.** Relative change in RMSE for both IDW + DNN (left) and the DMN (right), when provided with perturbed predictor data. Circle indicate outlying values.

can be used to downscale discrete air pollutant measurements to high resolution over any pre-determined spatial extent within the satellite observed area.

#### CRediT authorship contribution statement

**Manzhu Yu:** Data curation, Conceptualization, Methodology, Code implementation, Experiment, Result analysis, Paper Writing.

**Qian Liu:** Conceptualization, Result analysis, Paper review and editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

The authors would like to acknowledge the anonymous reviewers for their insightful comments. The authors also acknowledge Jeremy Diaz and Taylor Blackman on recommending the inclusion of land use and road networks to the predictors.

#### References

- Ahmed, S.O., Mazloum, R., Abou-Ali, H., 2018. Spatiotemporal interpolation of air pollutants in the Greater Cairo and the Delta, Egypt. *Environ. Res.* 160, 27–34.
- American Lung Association, 2020. State of the air report. Accessed from <https://www.stateoftheair.org/assets/SOTA-2020.pdf>.
- Baklanov, A., Rasmussen, A., Fay, B., Berge, E., Finardi, S., 2002. Potential and shortcomings of numerical weather prediction models in providing meteorological data for urban air pollution forecasting. *Water, Air and Soil Pollution: Focus* 2 (5–6), 43–60.
- Berrisford, P., Källberg, P., Kobayashi, S., Dee, D., Uppala, S., Simmons, A.J., Poli, P., Sato, H., 2011. Atmospheric conservation properties in ERA-Interim. *Q. J. R. Meteorol. Soc.* 137 (659), 1381–1399.
- Berrocal, V.J., Guan, Y., Muyskens, A., Wang, H., Reich, B.J., Mulholland, J.A. and Chang, H.H., 2020. A comparison of statistical and machine learning methods for creating national daily maps of ambient PM2.5 concentration. *Atmospheric Environment*, 222, p.117130.
- Blond, N., Boersma, K.F., Eskes, H.J., van der A, R.J., Van Roozendael, M., De Smedt, I., Bergametti, G. and Vautard, R., 2007. Intercomparison of SCIAMACHY nitrogen dioxide observations, in situ measurements and air quality modeling results over Western Europe. *Journal of Geophysical Research: Atmospheres*, 112(D10).
- Boersma, K.F., Eskes, H.J., Dirksen, R.J., Van Der R.J., A, Veefkind, J.P., Stammes, P., Huijnen, V., Kleipool, Q.L., Sneep, M., Claas, J., Leitão, J., 2011. An improved tropospheric NO<sub>2</sub> column retrieval algorithm for the ozone monitoring instrument. *Atmospheric Measurement Techniques* 4 (9), 1905–1928.
- Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing PM<sub>2.5</sub> exposures with high spatiotemporal resolution across the continental United States. *Environmental science & technology* 50 (9), 4712–4721.
- Dieudonné, E., Ravetta, F., Pelon, J., Goutail, F., Pommereau, J.P., 2013. Linking NO<sub>2</sub> surface concentration and integrated content in the urban developed atmospheric boundary layer. *Geophys. Res. Lett.* 40 (6), 1247–1251.
- Duncan, B.N., Lamsal, L.N., Thompson, A.M., Yoshida, Y., Lu, Z., Streets, D.G., Hurwitz, M.M., Pickering, K.E., 2016. A space-based, high-resolution view of notable changes in urban NO<sub>x</sub> pollution around the world (2005–2014). *Journal of Geophysical Research: Atmospheres* 121 (2), 976–996.
- Fenn, M.E., Haeuber, R., Tonneisen, G.S., Baron, J.S., Grossman-Clarke, S., Hope, D., Jaffe, D.A., Copeland, S., Geiser, L., Rueth, H.M., Sickman, J.O., 2003. Nitrogen emissions, deposition, and monitoring in the western United States. *BioScience* 53 (4), 391–403.
- Führer, O., Chadha, T., Hoefer, T., Kwasniewski, G., Lapillonne, X., Leutwyler, D., Lüthi, D., Osuna, C., Schär, C., Schulthess, T.C., Vogt, H., 2018. Near-global climate simulation at 1 km resolution: establishing a performance baseline on 4888 GPUs with COSMO 5.0. *Geosci. Model Dev.* 11 (4), 1665–1681.
- van Geffen, J. H. G. M., Eskes, H. J., Boersma, K. F., Maasakkers, J. D., and Veefkind, J. P., 2019. TROPOMI ATBD of the total and tropospheric NO<sub>2</sub> data products, Tech. Rep. S5P-KNMI-L2-0005-RP, Royal Netherlands Meteorological Institute (KNMI), <https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-NO2-data-products>, CI-7430-ATBD, issue 1.4.0.
- Ghude, S.D., Pfister, G.G., Jena, C., Van Der R.J., A, Emmons, L.K., Kumar, R., 2013. Satellite constraints on nitrogen oxide (NO<sub>x</sub>) emissions from India based on OMI observations and WRF-Chem simulations. *Geophys. Res. Lett.* 40 (2), 423–428.
- Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep learning. (Vol. 1, No. 2). MIT press, Cambridge.
- Hong, C., Zhang, Q., Zhang, Y., Tang, Y., Tong, D. and He, K., 2017. Multi-year downscaling application of two-way coupled WRF v3. 4 and CMAQ v5. 0.2 over east Asia for regional climate and air quality modeling: model evaluation and aerosol direct effects. *Geoscientific Model Development*, 10(6).
- Khan, S., Lee, D.H., Khan, M.A., Gilal, A.R., Mujtaba, G., 2019. Efficient edge-based image interpolation method using neighboring slope information. *IEEE Access* 7, 133539–133548.
- Li, L., Franklin, M., Girguis, M., Lurmann, F., Wu, J., Pavlovic, N., Breton, C., Gilliland, F., Habre, R., 2020. Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling. *Remote Sens. Environ.* 237, 111584.
- Liu, Y., Cao, G., Zhao, N., Mulligan, K., Ye, X., 2018. Improve ground-level PM<sub>2.5</sub> concentration mapping using a random forests-based geostatistical approach. *Environ. Pollut.* 235, 272–282.
- Lorente, A., Boersma, K.F., Eskes, H.J., Veefkind, J.P., Van Geffen, J.H.G.M., de Zeeuw, M.B., van der Gon, H.D., Beirle, S., Krol, M.C., 2019. Quantification of nitrogen oxides emissions from build-up of pollution over Paris with TROPOMI. *Sci. Rep.* 9 (1), 1–10.
- Nori-Sarma, A., Thimmulappa, R.K., Venkataramana, G.V., Fauzie, A.K., Dey, S.K., Venkareddy, L.K., Berman, J., Lane, K.J., Fong, K.C., Warren, J.L., Bell, M.L., 2020. Low-cost NO<sub>2</sub> Monitoring and Predictions of Urban Exposure Using Universal Kriging and Land-Use Regression Modelling in Mysore, India. *Atmospheric Environment*, p. 117395.
- Oteros, J., Bergmann, K.C., Menzel, A., Damialis, A., Traudl-Hoffmann, C., Schmidt-Weber, C.B., Buters, J., 2019. Spatial interpolation of current airborne pollen concentrations where no monitoring exists. *Atmos. Environ.* 199, 435–442.
- Pappenberger, F., and Hewson, T. 2017. ECMWF plans & product development. Accessed from: <https://www.ecmwf.int/sites/default/files/elibrary/2017/17273-ecmwf-products-development.pdf>.

- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. and Desmaison, A., 2019. Pytorch: an imperative style, high-performance deep learning library. In *Advances in neural information processing systems* (pp. 8026–8037).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Uno, I., He, Y., Ohara, T., Yamaji, K., Kurokawa, J.I., Katayama, M., Wang, Z., Noguchi, K., Hayashida, S., Richter, A., Burrows, J.P., 2007. Systematic Analysis of Interannual and Seasonal Variations of Model-simulated Tropospheric NO<sub>2</sub> in Asia and Comparison with GOME-satellite Data.
- Veefkind, J.P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H.J., De Haan, J.F., Kleipool, Q., Van Weele, M., 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 Sens. Environ.* 120, 70–83.
- Wang, Y., Wang, J., Zhou, M., Henze, D.K., Ge, C., Wang, W., 2020. Inverse modeling of SO<sub>2</sub> and NO<sub>x</sub> emissions over China using multisensor satellite data—part 2: downscaling techniques for air quality analysis and forecasts. *Atmos. Chem. Phys.* 20 (11), 6651–6670.
- Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily PM<sub>2.5</sub> estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* 199, 437–446.
- Xie, J., Yang, C., Zhou, B., Huang, Q., 2010. High-performance computing for the simulation of dust storms. *Comput. Environ. Urban. Syst.* 34 (4), 278–290.
- Yahya, K., Wang, K., Campbell, P., Chen, Y., Glotfelty, T., He, J., Pirhalla, M., Zhang, Y., 2017. Decadal application of WRF/Chem for regional air quality and climate modeling over the US under the representative concentration pathways scenarios. Part 1: model evaluation and impact of downscaling. *Atmos. Environ.* 152, 562–583.
- Zhu, S., Zeng, B., Zeng, L., Gabbouj, M., 2016. Image interpolation based on non-local geometric similarities and directional gradients. *IEEE transactions on Multimedia* 18 (9), 1707–1719.
- Zoogman, P., Liu, X., Suleiman, R.M., Pennington, W.F., Flittner, D.E., Al-Saadi, J.A., Hilton, B.B., Nicks, D.K., Newchurch, M.J., Carr, J.L., Janz, S.J., 2017. Tropospheric emissions: monitoring of pollution (TEMPO). *J. Quant. Spectrosc. Radiat. Transf.* 186, 17–39.