

## Article

# The Influence of Seasonal Climate Variability on Vehicle Exhaust PM<sub>2.5</sub> in the State of California: A Hybrid Approach based on Artificial Neural Network and Spatial Analysis

Fan Yu<sup>1</sup>, Amin Mohebbi<sup>2,\*</sup>, Shiqing Cai<sup>3</sup>, Simin Akbariyeh<sup>4</sup>, Brendan J. Russo<sup>5</sup>, and Edward J. Smaglik<sup>6</sup>,

<sup>1</sup> Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA; [fy55@nau.edu](mailto:fy55@nau.edu)

<sup>2</sup> Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA; [Amin.Mohebbi@nau.edu](mailto:Amin.Mohebbi@nau.edu)

<sup>3</sup> Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA; [sc2729@nau.edu](mailto:sc2729@nau.edu)

<sup>4</sup> Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA; [Simin.Akbariyeh@nau.edu](mailto:Simin.Akbariyeh@nau.edu)

<sup>5</sup> Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA; [Brendan.Russo@nau.edu](mailto:Brendan.Russo@nau.edu)

<sup>6</sup> Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA; [Edward.Smaglik@nau.edu](mailto:Edward.Smaglik@nau.edu)

\* Correspondence: [amin.mohebbi@nau.edu](mailto:amin.mohebbi@nau.edu); Tel.: +1-928-523-4769

**Abstract:** This study aims to develop a hybrid approach based on backpropagation Artificial Neural Network (ANN) and spatial analysis techniques to predict particulate matter of size 2.5  $\mu\text{m}$  (PM<sub>2.5</sub>) from vehicle exhaust emissions in the State of California using Aerosol Optical Depth (AOD) and several climatic indicators (relative humidity, temperature, precipitation, and wind speed). The PM<sub>2.5</sub> data were generated using Motor Vehicle Emission Simulator (MOVES), the measured climatic variables and AOD were obtained from the California Irrigation Management Information System (CIMIS), and NASA's Moderate Resolution Spectroradiometer (MODIS). The data were resampled to a seasonal format and downscaled over grids of 10 by 10 to 150 by 150, and precipitation was determined to be the most important independent variable. Coefficient of determination ( $R^2$ ), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) were used to assess the quality of the ANN prediction model. The model peaked at winter seasons with  $R^2 = 0.984$ , RMSE = 0.027, and MAPE = 25.311, whereas it had the lowest performance in summer with  $R^2 = 0.920$ , RMSE = 0.057, and MAPE = 65.214. These results indicate that the ANN model can accurately predict the PM<sub>2.5</sub> concentration and can be used to forecast future trends.

**Keywords:** Vehicle Exhaust PM<sub>2.5</sub>, MOVES, Artificial Neural Network, Spatial Analysis, Aerosol Optical Depth

## 1. Introduction

Nowadays, there is a unanimous scientific consensus that air quality degradation leads to adverse health effects [1]. One of the factors impacting the air quality index is suspended particulate matter, which is associated with numerous human diseases. Health problems from particulate matter can generally be classified as those caused by coarser particles targeting the lower respiratory system and those caused by finer particles targeting the lung's gas-exchange region. The coarser particles are classified as those with diameters smaller than  $10\ \mu\text{m}$  (PM<sub>10</sub>), whereas the finer particles are classified as those with diameters smaller than  $2.5\ \mu\text{m}$  (PM<sub>2.5</sub>). Both classes of particulate matter are highly dependent on climatic factors such as precipitation, relative humidity, wind speed, and air temperature, and are positively correlated with the aerosol optical depth (AOD) [2,3]. In particular, AOD, which can be retrieved from remote sensing products, is considered a good proxy for PM<sub>2.5</sub> [4].

Anthropogenic activities and natural emission are the primary sources of PM<sub>2.5</sub> in the Western United States. Natural emissions of PM<sub>2.5</sub> consist of wind-blown dust [5,6], biogenic non-methane volatile organic compounds (NMVOCs) [7], sea salt, wildfire, and lightning NO<sub>x</sub> emissions [8]. The sharp pulses in PM<sub>2.5</sub> time-series concentration during the summer season resulting from wildfires are also significant contributors to air pollution in this region [9]. Examples of anthropogenic activities are agricultural burning, construction and demolition activities, industrial processes, and mobile emissions. In general, the spatial distribution of anthropogenic PM<sub>2.5</sub> emissions is highly dependent on population density. Currently, motor vehicles substantially contribute to air pollution, accounting for 65% of total PM<sub>2.5</sub> emissions in urban areas in the United States [10]. Mobile emissions contain the fuel combustion exhaust from different vehicle types, which can be determined using the Motor Vehicle Emission Simulator (MOVES) [11]. MOVES is an emission modeling system that estimates emissions from on-road motor vehicles and most off-road equipment nationwide.

There are several statistical techniques for particulate matter distribution forecasting, and among these, the artificial neural network (ANN) is expected to have a better performance than classical linear and non-linear regression methods [1]. The reason for better performance is that ANN has a better adaptation capability in highly non-linear physical processes [1]. Moreover, a hybrid method based on ANN and spatial analysis is expected to result in a more accurate model for particulate matter prediction. Hence, the study aims to develop a working model of particulate matter dispersion based on AOD with climatic variables as auxiliary inputs. Three relatively recent years (2010, 2014, and 2018) were selected to calibrate, validate, and test the model, and the U.S. state of California was chosen as a case study because of the high levels of particulate matter in this region. Although the model was developed for California, it also had the potential to be utilized in other regions worldwide, given the widespread availability of the input data used in this research.

The remainder of the paper is organized as follows: First, a literature review is presented to summarize the results of similar studies regarding the prediction of PM<sub>2.5</sub> with a particular focus on ANN models. The materials and methods section primarily focuses on the spatial analysis of data and ANN's mathematical theory. In particular, this section covers model input and target variables, including PM<sub>2.5</sub> generated from MOVES, AOD extracted from NASA's Moderate Resolution Spectroradiometer (MODIS), and meteorological data acquired from California Irrigation Management Information System (CIMIS). This work's major contribution is an ANN model for PM<sub>2.5</sub>, which is calibrated and assessed in the results and discussion section. Lastly, recommendations are provided to further highlight the practical aspect of the model.

## 2. Literature Review

In this section, a condensed and focused literature review is presented. For an in-depth review, the reader is referred to [12].

## 2.1 Classical regression models for particulate matter

Motor vehicles contribute 10% to 65% of total PM<sub>2.5</sub> emissions in urban areas in the United States [10]. To assess air quality impacts of transportation projects and determine which projects should be identified as a Project of Air Quality Concern (POAQC), Reid et al. (2017) used both the MOVES and the Emission Factors model (EMFAC) to quantify PM<sub>10</sub> and PM<sub>2.5</sub> emissions in an analysis conducted for the year 2006. Their goal was to evaluate the impact of fleet turn-over and truck use percentages on future project-level emissions [13]. Comparing the two models, they concluded that tire wear and re-entrained road dust emissions from both models were roughly equal. However, MOVES exhaust emissions estimates for 2006 were about 70% higher than the EMFAC estimates. On the other hand, EMFAC brake wear estimates were 2.5 times higher than the MOVES estimates [13].

Multiple regression approaches are used to define the correlation between particulate matter, AOD, and meteorological products. Gupta and Christopher (2009) developed a simple Two-Variate Method (TVM), where AOD observed by MODIS was initially used to estimate surface-level PM<sub>2.5</sub> mass concentration. Later, they gradually added meteorological parameters to the built model to form a Multiple Variate Method (MVM). In the study, PM<sub>2.5</sub> mass concentration was obtained from the EPA, and satellite aerosol data were obtained from the MODIS [14]. The study used a Rapid Update Cycle (RUC) to obtain meteorology data, including air temperature at 2 m, surface relative humidity, wind speed at 10 m, and height of the planetary boundary layer [14]. The results showed that applying the MVM model improved the accuracy of the TVM model's hourly and daily average particulate matter 2.5 values by 13% and 17%, respectively [14].

## 2.2 ANN models for particulate matter

A neural network can learn particular skills by utilizing three layers: the input layer, hidden layer, and output layer [14]. The input layer receives input data, which in most cases are physical characteristics of a complex phenomenon. The output layer receives the desired output data and *typically* is limited to a single variable. There is a hidden layer between input and output layers that establishes a non-linear relationship between the input and the output data. This operation is accomplished by nodes (neurons) embedded in the hidden layer simulating the actual human brain function. These nodes receive the input, multiply it by weight, add a bias, and then pass it to the output. This procedure is repeated until the hidden layer output (called target data) converges to the initial output. The governing equations will be discussed in depth in the Materials and Methods section.

The complexity inherited in particulate matter distribution makes the ANN an ideal candidate. In the study conducted by Gupta and Christopher (2009), the ANN model improved particulate matter estimation accuracy by 23%. Compared with the TVM and MVM models, ANN was found to have a greater capacity to reduce the absolute percentage error [14]. A similar study was conducted by Zaman et al. (2017), who used Multiple Linear Regression (MLR) and ANN to conduct regression analysis on particulate matter, meteorology variables, and AOD. The meteorological data and AOD data were obtained from satellites. Since ground-based measurement data had a low spatial resolution, it was only used for validation. The coefficient of determination ( $R^2$ ) of the ANN approach was 0.71, and the Root Mean Square Error (RMSE) was 11.61  $\mu\text{gm}^{-3}$ , which was better than the MLR results ( $R^2 = 0.66$  and RMSE = 12.39  $\mu\text{gm}^{-3}$ ) [15]. Thus, the ANN model was designated superior to the linear regression model.

To deal with low spatial resolution problems encountered in [15], Enotoriuwa et al. (2018) used spatial interpolation. In their study, a 10 by 10 grid was generated over the study location [16]. The interpolated data were then downscaled to the grid points to improve the spatial resolution. After removing the grids with outliers, the total number of the remaining nodes for regression was 108 [16]. This approach resulted in  $R^2$  values of 0.65 and 0.58 for PM<sub>2.5</sub> and PM<sub>10</sub>, accordingly.

Other literature acknowledged include particulate matter studies frequently conducted in the Beijing-Tianjin-Hebei (BTH) area because particulate matter is a significant contributor to haze weather

and air quality in this region. Wu et al. (2012) used a backpropagation ANN algorithm to estimate particulate matter at the BTH area using AOD and other meteorological data. In this study, the number of nodes in the hidden layer varied from 6 to 15 [17]. In terms of the correlation between observed and estimated particulate matter concentrations, most models' correlation coefficient values were greater than 0.4 [17]. Another study conducted by Yi et al. (2019) in the same area used three models: backpropagation ANN, Support Vector Regression ( $\epsilon$ -SVR), and combined backpropagation ANN and  $\epsilon$ -SVR. The models were used to estimate the missing historical values of PM<sub>2.5</sub> [18]. ANN and  $\epsilon$ -SVR performed similarly in terms of the correlation coefficient. However, combining the two methods further improved the accuracy of the estimate. This method was suitable for the BTH area but could be extended to other locations [18].

The present study is developed based on the assumption that the particulate matter is positively correlated to the AOD levels. Additionally, it builds upon the studies mentioned above to develop a working model of particulate matter for California. In particular, monthly PM<sub>2.5</sub> data are generated [13], monthly observed meteorological data are acquired [14,19], spatial analysis is utilized to improve low data resolution [16], and eventually, the ANN approach is used to develop an accurate model for particulate matter spatial distribution based on the meteorological variables [15,17,18]. The work's novelty is an in-depth study of the impact of spatial resolution and the use of model-generated particulate matter. This paper's most notable contribution is studying the seasonal impact on the model generated particulate matter levels.

### 3. Study Area

The study area for this research was the state of California (Figure 1), which is located on the west coast of North America (Latitude: 36.7783° N, Longitude 119.4179° W). California faces the Pacific Ocean on the west and is bounded on the north by Oregon, on the east by Arizona and Nevada, and on the south by the Mexican state of Baja California. Northern and central California have a Mediterranean climate, whereas the southeast of California has an arid climate [20]. California's average temperature is around 25°C, and the average annual precipitation is approximately 457.2 mm. California is the most populous state with the largest economy in the United States. It consists of 58 counties and has two large megalopolises: one in southern California and one in the Bay Area. The number of registered automobiles in California was 14.6 million in 2017, making it a significant contributor to particulate emissions.

### 4. Materials and Methods

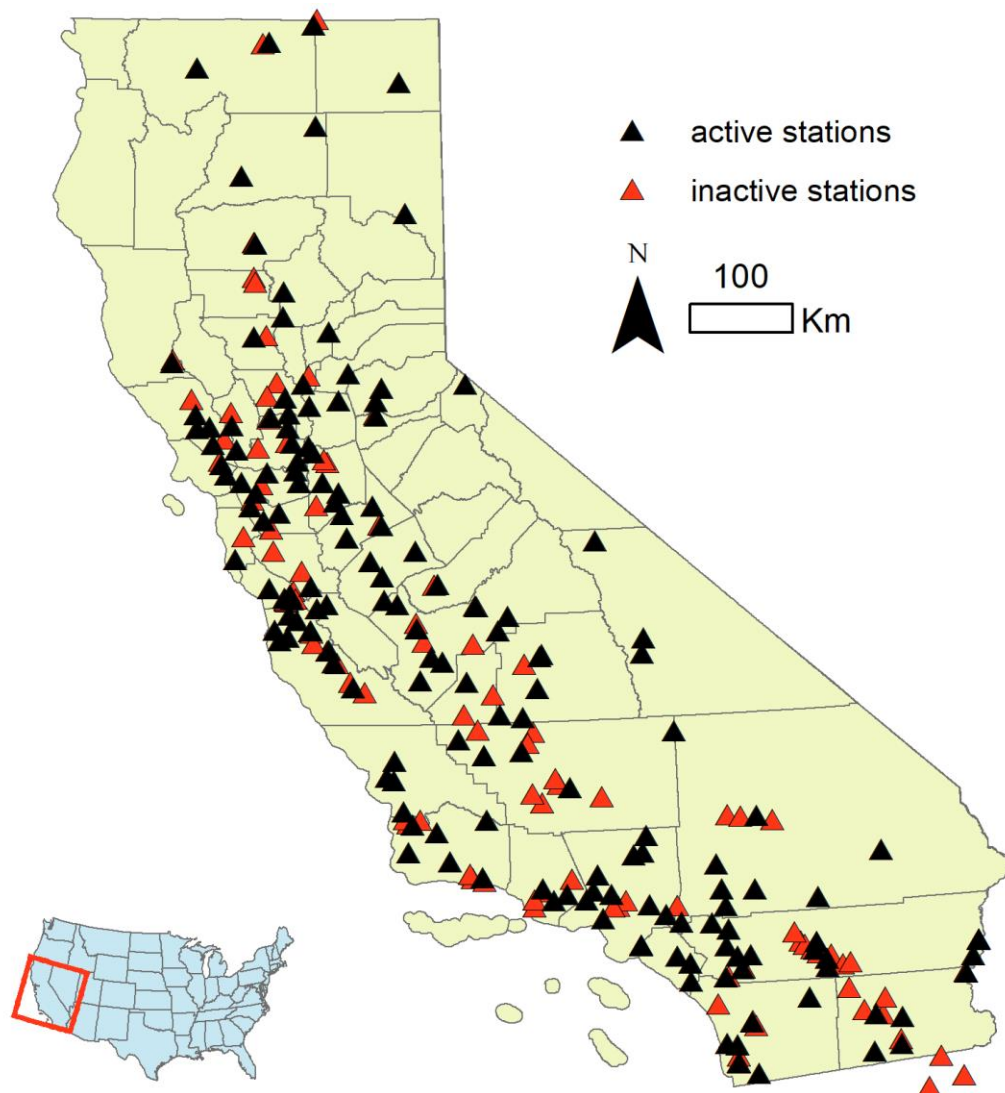
This study benefits from modeled (PM<sub>2.5</sub>), measuring station (temperature, wind speed, precipitation, and relative humidity), and observational satellite data (AOD). The measured and satellite data are available at different spatial and temporal resolutions; for example PM<sub>2.5</sub> and meteorological variables are available monthly while AODs are available daily. A seasonal time scale was chosen for this study, so data were scaled up to create seasonal averages. The Northern meteorological seasons are spring (March, April, and May), summer (June, July, and August), fall (September, October, and November), and winter (December, January, and February). The procedure for data processing has been visualized in Figure 2, in which the input data are fed into the spatial model for rescaling purposes.

Three recent years of 2010, 2014, and 2018 were selected as a case study to analyze the annual trends. As for the spatial measure, the data were resampled over uniform grids generated in Figure 3. These grids were initially created as squares with tiles of 10 by 10 (100 km by 100 km), 20 by 20, etc., through 150 by 150 (7 km by 7 km) and were overlaid on the state of California. Then, only the grids falling inside the California boundary were kept. The grids with tiles larger than 80 by 80 were not recognizable and therefore were not visualized in Figure 3. Once the data's temporal and spatial resolutions were adjusted, an ANN model was developed with the meteorological and AOD data as inputs and PM<sub>2.5</sub> mass as

target data. This is also shown in Figure 2, where the processed data are fed from the spatial model to the ANN model.

#### 4.1. Artificial Neural Network (ANN)

ANNs are the systems that imitate the mechanics of neural association of the human brain rather than implementing specific governing equations. The neural network has a robust non-linear mapping ability, which generates a more flexible functional relationship between the input and target data [21]. A feed-forward neural network was established to determine the non-linear relationship between input data and PM<sub>2.5</sub>. This ANN model consisted of an input layer, an output layer, and a hidden layer. Increasing the number of hidden layers can create a deep neural network as known as deep learning. The architecture of the feed-forward ANN, together with input and output parameters, are depicted in Figure 4. As shown in Figure 4, five input parameters, including precipitation, air temperature, relative humidity, wind speed, and AOD, were fed in the input layer to estimate the PM<sub>2.5</sub> mass in the output layer.



**Figure 1.** Weather monitoring stations in the study area



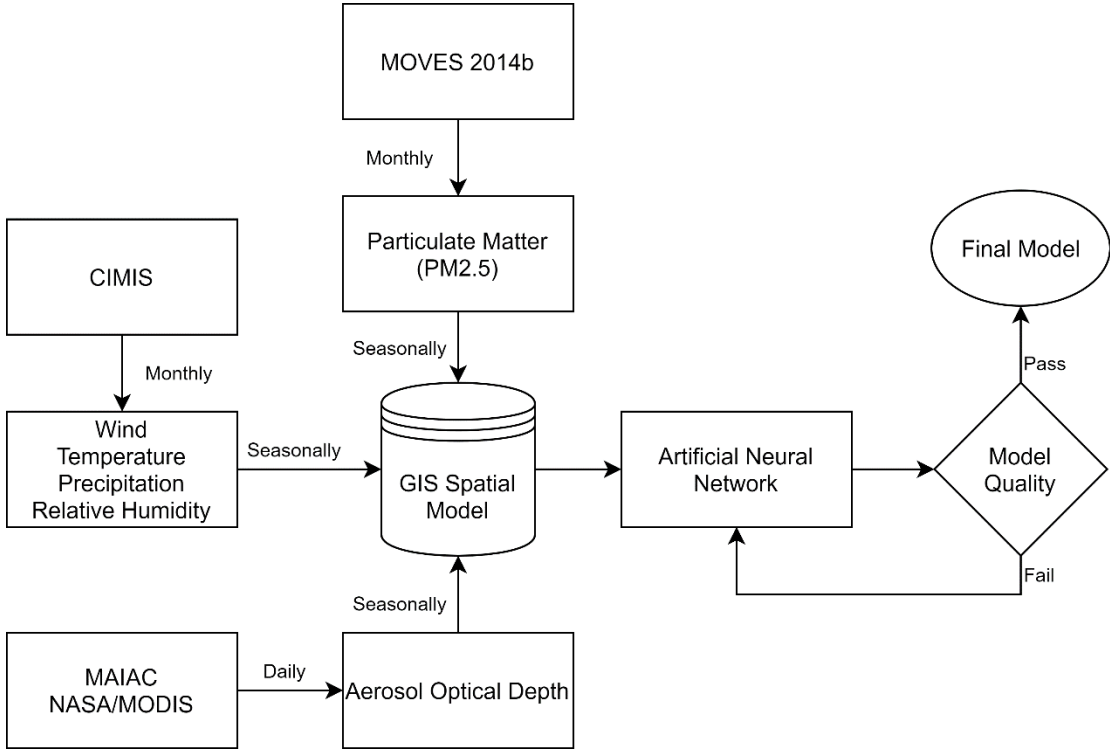


Figure 2. Project overview

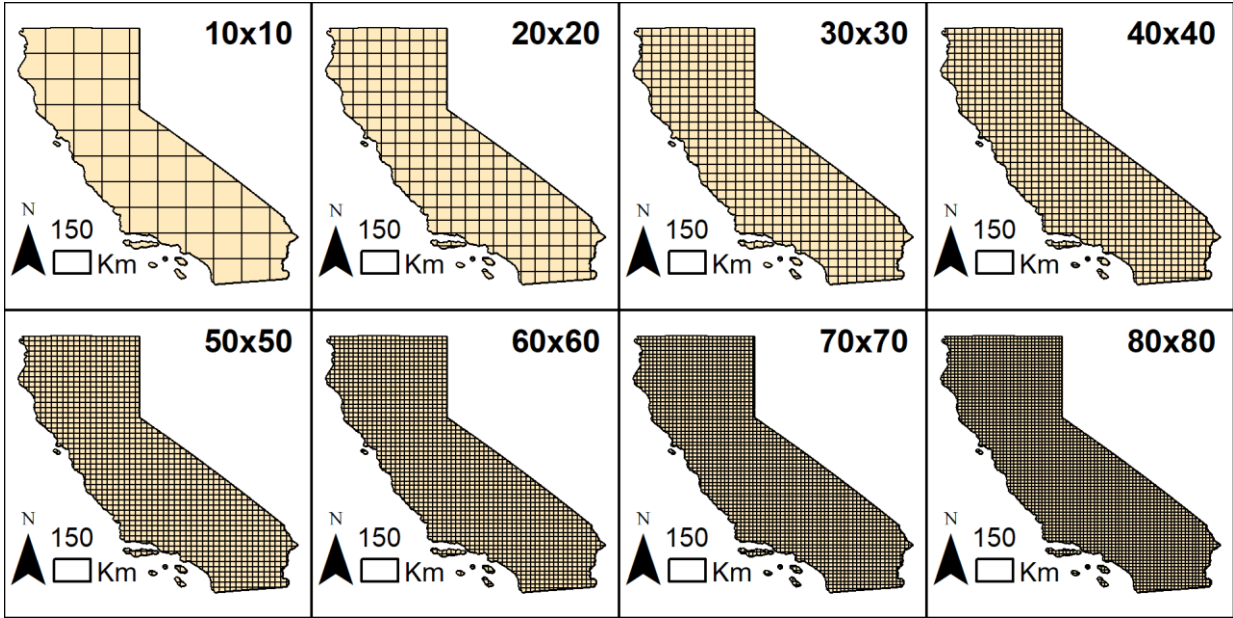
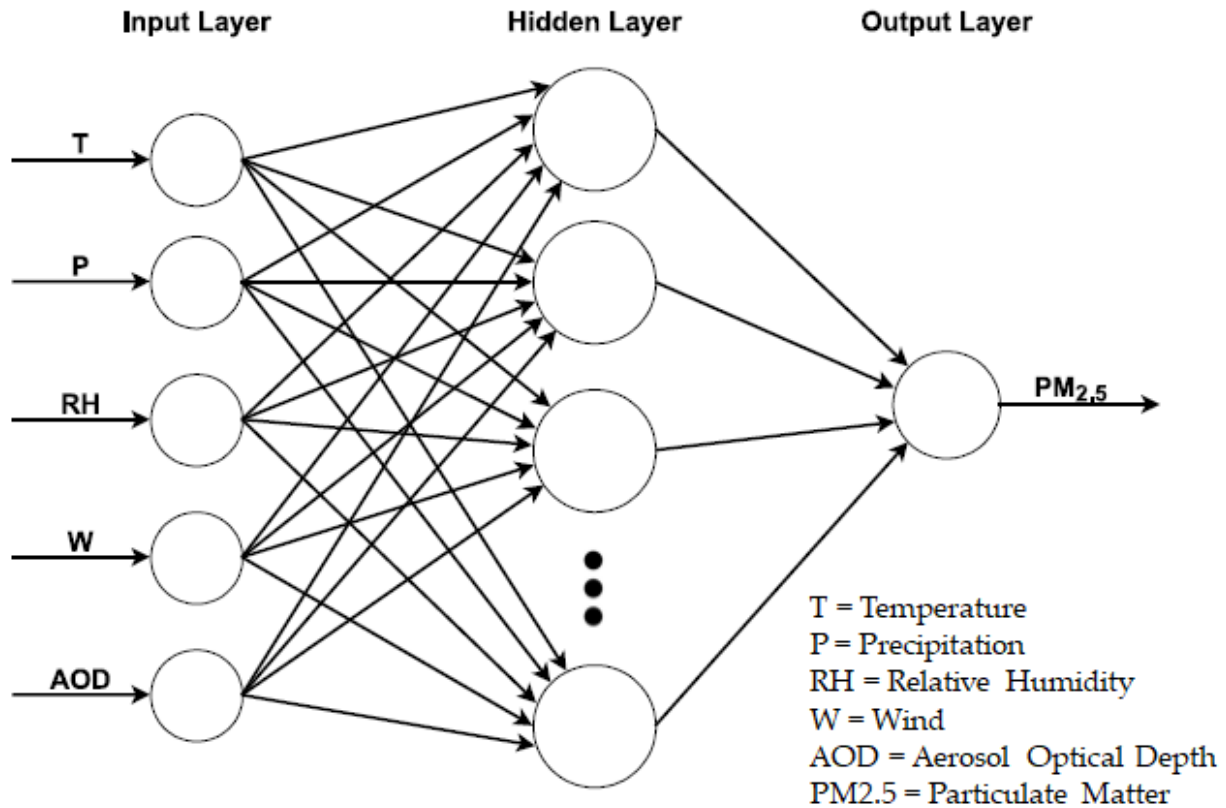


Figure 3. Spatial mesh in various resolutions generated for the state of California



**Figure 4.** Schematic view of a feed-forward artificial neural networks

Each layer may have multiple nodes/neurons that simulate the biological system by receiving signals from other neurons and transmitting to the next neurons. Each neuron is the output of the previous layer and input for the next layer. The line connections in Figure 4 represent the weights in ANN through which neurons interact [22]. The neurons in the hidden and output layer are calculated through the weight connection and net input function. The net input can be expressed as

$$n = \sum_{j=1}^R w_j p_j + b = \mathbf{W}\mathbf{p} + b \quad (1)$$

where  $p_j$  represents the input parameter, defined as the elements in the input vector  $\mathbf{P} = [p_1, p_2, \dots, p_R]$  with corresponding weights of  $\mathbf{W} = [w_1, w_2, \dots, w_j]$ . The input  $p_j$  is multiplied by the weight and added to bias ( $b$ ) to form the net input. Bias is an intercept to adjust the weighted sum of the net inputs to the neuron. The transfer function to process the net input to obtain the neuron output is calculated by

$$a = f(n) \quad (2)$$

with  $f$  as the activation function. In this study, the log-sigmoid activation function was adopted as

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

During the network training, the input parameters, number of neurons, and layers are preset. The network characteristics are stored in the form of weights and biases. This information, together with Equations 1-3, makes the output a function of weights and biases. In this study, the training was accomplished using Levenberg-Marquardt backpropagation, a hybrid algorithm combining backpropagation and Newton's method. The process of backpropagation involves adjusting the weights and biases in the network to minimize the error between the target and output values. This concept can

be turned into a cost function ( $\hat{F}$ ) expressed as the mean square error ( $e$ ) for the  $q^{th}$  input/output pair by

$$\hat{F} = \frac{1}{2} e_{-q}^T e_{-q} \quad (4)$$

The process of cost function minimization was performed using gradient descent. The gradient is the derivative of the cost function calculated with respect to the weights and biases and expressed as  $\frac{\partial \hat{F}}{\partial w}$  and  $\frac{\partial \hat{F}}{\partial b}$ , respectively. The process is finding a slope that leads the steepest descent to reach the local minimum of the function, i.e., the error between target and output is minimized. In the Levenberg-Marquardt algorithm, the Jacobian matrix was introduced during the computation of gradient descent. The Jacobian matrix requires calculating the derivatives of the network's errors instead of the mean square errors [23]. This can be computed as

$$\mathbf{g} = \mathbf{J}^T \mathbf{e} \quad (5)$$

where  $\mathbf{e}$  is the vector of network errors, and  $\mathbf{J}$  is the Jacobean matrix that contains first derivatives of the network's errors with respect to biases and weights [24]. The second derivative of the network's errors can be expressed as a Hessian matrix by

$$\mathbf{H} = \mathbf{J}^T \mathbf{J} \quad (6)$$

Applying the Gauss-Newton method, the Levenberg-Marquardt algorithm can be written as

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{H} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{e} \quad (7)$$

The parameter  $\mathbf{x}_k$  represents weights and biases at iteration step  $k$  with  $\mathbf{I}$  as an identity matrix, and  $\mu$  as a damping factor. When  $\mu$  increases, the function becomes closer to the gradient descent, slowly reaching the minimum error. As  $\mu$  decreases (getting closer to zero), the algorithm reverts into Newton's method. Newton's method is faster and more accurate near the error minima [24].

#### 4.2. Particulate matter

Contrary to the previous studies that utilize in-situ particulate matter measurements, this project benefits from the modeled particulate matter estimates. While acknowledging the availability of the directly measured data [25,26], it was not feasible to incorporate them into this study. This is because the study's focus was PM<sub>2.5</sub> from vehicle emissions, and to date, there has not been any direct measurement available for car emissions. The only available data were from Don Stedman's group at the University of Denver [27,28] but, the temporal and spatial resolutions of this data were not sufficient to serve the purpose of this study.

Because of this, particulate matter values were generated using the United States Environmental Protection Agency's Motor Vehicle Emissions Simulator (EPA MOVES). The MOVES model estimations of pollutant emissions from mobile sources are one of the two models available. It contains relevant vehicle emission inventory for the entire United States. The MOVES framework includes four different parts: the Total Activity Generator (TAG), the Source Bin Distribution Generator (SBDG), the Operating Mode Distribution Generator (OMDG), and the emission calculator [11]. TAG uses base year vehicle population and Vehicle Miles Traveled (VMT) to analyze yearly growth factors and categorizes them by road type, vehicle type, vehicle age, and time [29]. MOVES uses discrete binning when evaluating emissions, in which each type of traffic activity is binned into a single data point based on the fuel consumption of automobiles [29]. In the process of differentiating exhaust emissions, the Vehicle Specific Speed (VSP) was defined for light-duty vehicles and the Scaled Tractive Power (STP) for heavy-duty vehicles [30]. The VSP and STP values depend on the vehicle mass, dynamic parameters, speed, acceleration, and gravitational acceleration. The mathematical form of modeling equation for light-duty and heavy-duty vehicles was [31]



$$VSP \text{ or } STP = \left(\frac{A}{M}\right)v + \left(\frac{B}{M}\right)v^2 + \left(\frac{C}{M}\right)v^3 + \left(\frac{m}{M}\right)(a + g * \sin \theta)v \quad (8)$$

where  $A$  is coefficient of rolling resistance ( $\text{kW}\cdot\text{sm}^{-1}$ ),  $B$  is coefficient of rotational resistance ( $\text{kW}\cdot\text{s}^2\text{m}^{-2}$ ),  $C$  is coefficient of aerodynamic drag ( $\text{kW}\cdot\text{s}^3\text{m}^{-3}$ ),  $v$  is instantaneous vehicle speed ( $\text{ms}^{-1}$ ),  $a$  is instantaneous acceleration ( $\text{ms}^{-2}$ ),  $M$  is fixed mass factor for the source type (metric tons),  $m$  is vehicle mass (metric tons),  $\theta$  is road grade (fractional), and  $g$  is the gravitational acceleration ( $9.8 \text{ ms}^{-2}$ ). The  $m/M$  ratio of light-duty vehicles was assumed to be 1.0. The  $M$  in the heavy-duty vehicles formula was approximately the average running weight for all heavy-duty vehicles [30].

In creating the emission output using the MOVES 2014b model [11], the emission analysis scale was set to nationwide on-road vehicles. The defined time scale for calculation was monthly, and the geographic boundary was set to the state of California. The output of the model was the mass of particulate matter in kilograms per county. The mobile sources of emissions consisted of on-road vehicles of all fuel types. The particulate matter emission was generated for all road types with the 1990 Clean Air Act Amendments implemented, and the output data and the reported units were stored in the output database.

#### 4.3. Aerosol Optical Depth (AOD)

Aerosol consists of fine solid particles and liquid droplets in the atmosphere. It can absorb and scatter rays of light that pass through it. Often, the wavelength at  $0.55 \mu\text{m}$  is used for aerosol retrieval from the satellite images. AOD, a dimensionless positive value, is a measure of extinction of the solar irradiance at a specific wavelength. It represents the aerosols distributed within a column of air from the earth's surface to the top of the atmosphere. AOD magnitude range is from 0 to 1, where 0 represents an entirely transparent atmosphere, and 1 represents an entirely opaque atmosphere. The relationship between AOD and solar irradiance at a specific wavelength of  $\lambda$  can be described by the Beer-Lambert law by

$$I(\lambda) = I_0(\lambda) \exp[-\tau_a * m] \quad (9)$$

where  $I_0(\lambda)$  is solar irradiance of extra-terrestrial, which is considered as solar constant,  $I(\lambda)$  is solar irradiance at a wavelength of  $\lambda$  at the earth's surface,  $\tau_a$  is the atmospheric optical depth, and  $m$  is the optical air mass, which can be calculated as relative air column length of zenith angle  $\theta$  at that of zenith direction 1. The solar radiation diminishes due to scattering and absorption by atmospheric elements such as air molecules, aerosols, water vapor, carbon dioxide, and ozone. Therefore, to calculate the aerosol optical depth, it is required to eliminate the optical depth of air molecules, water vapor, and other atmospheric elements.

AOD values reflect the real-time particulate matter concentration in the atmosphere. The AOD data used in the analysis were extracted from the databases on the MODIS website. The MODIS is an essential remote sensor onboard NASA's satellites Terra and Aqua. Both satellites scan the entire surface of the earth for various data [32]. Typically, MODIS is used for acquiring information on aerosols, ocean color, cloud, and ozone. In this study, the daily AOD product used is from MCD19A2 version 6 of MODIS. These data are generated by an algorithm for Multi-Angle Implementation of Atmospheric Correction (MAIAC) [33]. Combining Terra and Aqua values can improve cloud detection accuracy, aerosol retrieval, and atmospheric correction.

The downloaded MCD19A2 Hierarchical Data Format (HDF) files include daily AOD gridded Level 2 data at 1-kilometer (km) pixel resolution [33]. MODIS terrestrial data are distributed in non-overlapping tiles in sinusoidal projection. The shape of the tiles is approximately square and measure 10 degrees by 10 degrees at the equator. The tile coordinate system starts at the upper left corner of the map and proceeds right (horizontally) and down (vertically). The horizontal axis is divided into 36 segments (0-35), and the vertical axis is divided into 18 segments (0-17) [34].

4.4. Meteorology

The daily meteorological data used in the study was collected from the CIMIS. The CIMIS department operates 259 weather stations, including 157 active (had data for the study period) and 102 inactive stations (did not have data for the study period). The locations of weather monitoring stations are depicted in Figure 1 as triangles. The sensors can collect weather data on a minute by minute basis. The hourly and daily data are calculated and stored in the data loggers and automatically transferred to the central computer for further analysis and correction. The processed data are stored in the central database server and is available on the CIMIS website [35].

The meteorological data used in this study for modeling were precipitation (P), relative humidity (RH), wind speed (W), and air temperature (T). The sensor for air temperature and relative humidity is Fenwal Thermistor/HUMICAP H-sensor. The range of detection is -35 to +50°C for the temperature and 0 to 100% for the relative humidity. The wind speed sensor is a three-cup anemometer utilizing a magnet-activated reed switch. The frequency of this anemometer is proportional to wind speed. The instrument’s height is 2.0 m and can measure the wind speed of 0 to 100 mph. A tipping bucket rain gauge with a magnetic reed switch is used to record the precipitation, and the accuracy of the sensor is within 1%.

5. Results and Discussion

5.1 Seasonal pattern of PM2.5

Figure S1 shows the MOVES estimated PM2.5 seasonal distribution in California counties. The color bar range and style were kept consistent for all the choropleth maps. According to Figure S1, the spatial variation of PM2.5 was generally more significant than its temporal variations. The highest PM2.5 concentration was observed near southern California, while the lowest values occurred in the north and eastern California regions. This is attributed to the fact that Los Angeles County has the highest and Alpine County the lowest numbers of registered vehicles in the state, based on California’s Department of Motor Vehicle report for the year 2019 [36]. Seasonally, summer received the highest maximum PM2.5, whereas winter received the lowest maximum PM2.5 (Table 1). On the other hand, winter received the highest minimum PM2.5, whereas summer received the lowest minimum PM2.5 (Table 1). This is attributed to the driving habits, motives, and attitudes of state residents [37].

Table 1. Maximum and minimum seasonal PM2.5 in the state of California for years 2010, 2014, and 2018

| Season/Year | Maximum PM2.5 (kg) |         |         | Minimum PM2.5 (kg) |         |         |
|-------------|--------------------|---------|---------|--------------------|---------|---------|
|             | 2010               | 2014    | 2018    | 2010               | 2014    | 2018    |
| Summer      | 3.0E+05            | 2.0E+05 | 1.6E+05 | 2.9E+02            | 1.9E+02 | 1.6E+02 |
| Spring      | 2.9E+05            | 1.9E+05 | 1.5E+05 | 3.0E+02            | 2.0E+02 | 1.6E+02 |
| Fall        | 2.8E+05            | 1.9E+05 | 1.5E+05 | 3.0E+02            | 1.9E+02 | 1.5E+02 |
| Winter      | 2.7E+05            | 1.8E+05 | 1.4E+05 | 3.2E+02            | 2.1E+02 | 1.6E+02 |

5.2 Seasonal pattern of AOD

In this study, the AOD files were obtained when the satellites traversed the study domain, with filenames *h08v04.hdf*, *h08v05.hdf*, and *h09v04.hdf*. They represent the three tiles covering the state of California. Table 2 shows the minimum and maximum latitude and longitude of these tiles. In the filenames, “h” stands for horizontal, and “v” stands for vertical. Post-processing was performed to generate seasonal data using the official NASA’s image processing software SeaDAS [38]. A Level-3 binning was used to create the seasonal averages from the downloaded daily AOD data. The generated data were cropped to the study region via the Well-Known Text (WKT) state boundary. The overlap

between the three files was averaged to create a smooth and uniform surface. Finally, the three tiles were joined to form a single raster dataset.

**Table 2.** Mosaic AOD files in Hierarchical Data Format (HDF) covering the state of California

| File Name | min longitude | max longitude | min latitude | max latitude |
|-----------|---------------|---------------|--------------|--------------|
| h08v04    | -155.5724     | -117.4758     | 40.0000      | 50.0000      |
| h08v05    | -130.5407     | -103.9134     | 30.0000      | 40.0000      |
| h09v04    | -140.0151     | -104.4217     | 40.0000      | 50.0000      |

A correlation study was conducted to validate the earlier assumption with both the emission data and aerosol data in hand. This study's central assumption was a strong positive correlation between the AOD and the particulate matter. This assumption was backed by studies cited in the Literature Review section but was further investigated here. To this end, the frequency curves of PM<sub>2.5</sub> and AOD were plotted on separate graphs (not shown here). The final frequency distribution in both cases exhibited a right-skewed Gumbel distribution, proving the existence of a correlation between the two variables.

The AOD raster datasets are visualized in Figure S2, where the concentration of AOD was found to be the highest in summer. These high levels were due to large natural emissions during dry seasons, especially the northern part of California, which was impacted by frequent massive wildfires in the summer of 2018, including the destructive Carr and the Mendocino Complex fire [39]. The lowest amounts were observed during winter, which is in contrast to the countries such as China with high reliance on coal-based energy.

Also, it should be noted that the AOD concentration reflects the amount of particulate matter in the air. However, it may not directly represent the near-surface anthropogenic emissions. The cloudy condition close to the eastern border during the spring and winter affects AOD levels. This was very significant in winter 2010, where 30% of AOD data were lost (Figure S2). These missing data could be recreated by interpolation methods such as kriging [40], but this required a validation study later, which was out of the scope of this research. Therefore, data with a threshold lower than 0.01 were not passed to the ANN model.

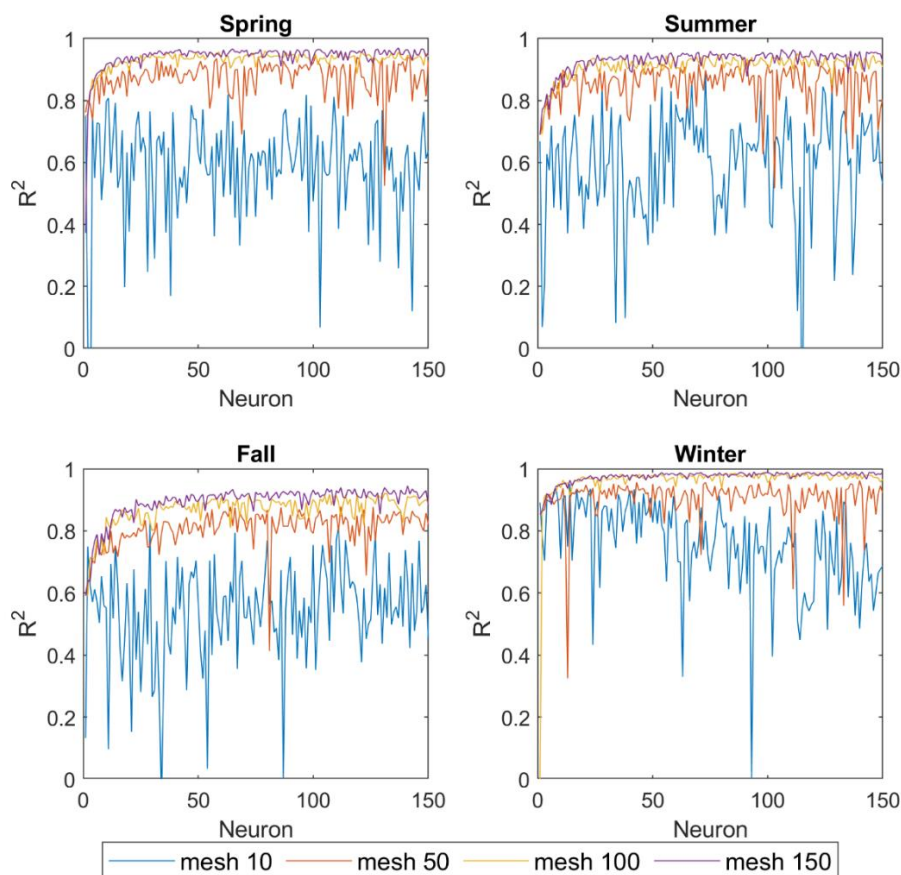
### 5.3 Seasonal pattern of climate

Meteorological data were acquired from CIMIS in a monthly timescale from each active station in the domain (Figure 1). Using the Inverse distance weighted (IDW) interpolation, pointwise data were converted into a surface. The IDW power parameter was kept the recommended value of two. Figures S3-S6 show seasonal variations of precipitation, relative humidity, wind speed, and temperature, generated from monthly measurements for the study period. Active stations are superimposed on the figures as small black triangles. The white spot at the northwest corner of the state indicates no data due to the stations' low density. Summer precipitation is significantly lower than the rest of the seasons. Winter shows the highest relative humidity, whereas the rest of the seasons show relatively similar magnitudes. The wind speed is higher in the spring and summer and at least one order of magnitude lower for the fall and winter. The variability of temperature is trivial, with the highest in summer and the lowest in winter.

It is worth mentioning that the meteorological variables selected here are not the only variables affecting the particulate matter distribution. Other variables such as wind direction, planetary boundary layer height, terrain height, vegetation fraction, perturbation pressure, and base state pressure also play a role [3]. However, this research aimed to limit the input data to the most available ones to reach broader applicability.

### 5.4 ANN model architecture

The artificial neural network model was built using the Levenberg-Marquardt backpropagation algorithm, as described in the methodology section. A feed-forward neural network with one hidden layer was coded in MATLAB version 9.4 to determine the relationship between input and target. Adding more hidden layers may result in a lower error, but a rigorous method to prevent overfitting was required. All the datasets, including input and output, were standardized to [0, 1] prior to model training [41]. For a reasonable PM<sub>2.5</sub> level prediction, the best grid size and the optimal number of neurons must be determined, which is crucial to avoid the ANN model under-training or over-fitting [19]. To this end and to study the model's spatial dependency, data were processed in different grid sizes ranging from 10 to 150 in each direction. Moreover, neurons ranging from 1 to 150 were examined in each layer. The coefficient of determination ( $R^2$ ) was used as an indicator to find the best number of neurons and grid size. Figure 5 presents the variation of  $R^2$  and the number of neurons with respect to different grid sizes. Only four representative grid sizes were selected for graphing to avoid overcrowding the figure. According to Figure 5, the grid size of 150 by 150 had the highest  $R^2$  value for all seasons, but any grid size above 100 was within a 1.8% margin of error. Generally, the larger grid size generates data at higher spatial resolution and improves modeling accuracy. Additionally, the  $R^2$  was nearly constant for the number of neurons larger than 50. The optimum number of neurons for each of the four seasons was determined as 130. Although increasing the number of neurons in layers could improve accuracy, it would reduce the computational efficiency. Compared with previous research, which had optimal neurons less than 10 [19], this project adopts a larger number of neurons. The reason why a large number of neurons did not cause over-fitting (based on the test data presented later) in this case is due to the large size of the input dataset.



**Figure 5.** The variation of  $R^2$  with respect to the number of neurons for grids 10, 50, 100 and 150

### 5.5 ANN model performance

After deciding on the optimal number of neurons and grid size, the ANN model was run with the input layer consisting of AOD and meteorological variables. The available datasets were randomly split into three categories of training (70%), validation (15%), and testing (15%) [24]. The performance criteria including coefficient of determination ( $R^2$ ), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) [22] were selected to measure the goodness of fit between the ANN results and the PM2.5 generated by MOVES.

The resulting values of  $R^2$ , RMSE, and MAPE for the training, validation, and testing periods are reported in Table 3. The  $R^2$  was considered as the primary index for evaluating model performance. According to Table 3, in the training period, the best model performance was achieved for winter with an average  $R^2 = 0.984$ , while more moderate performance was seen in other seasons (spring  $R^2 = 0.952$ , fall  $R^2 = 0.947$ , and summer  $R^2 = 0.920$ ). RMSE values closer to zero and lower percentages of MAPE indicate better agreement between the simulated and modeled data. According to a similar study conducted by Yi et al. (2019), MAPE values below 20% were considered good [18]. As shown in Table 3, winter has the lowest MAPE (25.31%) and RSME (0.027) among all seasons. In contrast, the summer MAPE values for the three years of modeling were all above 50%, with the lowest  $R^2$  of 0.899 in 2018. Overall, the MAPE values do not represent as good an agreement as the  $R^2$  values did. This could be due to the large variability of the data that even standardization was not able to resolve.

During the modeling process, ANN backpropagation is prone to over-fitting. When the neural network is so closely fitted to the training data, it is not reliable to make predictions for new data [42]. Validation and testing processes were considered to avoid ANN overfitting. The validation dataset was used for optimizing the model, which stopped the model whenever the error rate was degraded for six consecutive epochs. In addition, a testing dataset was used at the end of the training process to evaluate the overall performance of the model. The performance in validation and testing resembled the results in training. Comparing the  $R^2$  values in validation and testing periods, the RMSEs remained below 5% (Table 3). Successful completion of this step proved that the method could replicate standalone case studies. The question remained unanswered: is the model skill in repeating itself, i.e., does the method exhibit seasonality?

The seasonality of the model is investigated by creating four master datasets, each representing a season, e.g., the first master dataset had 2010, 2014, and 2018 spring seasons. To better visualize the model's performance, a comparison was performed between the ANN simulated, and the MOVES estimated PM2.5 values (Figure 6). In this figure, if all the data points were aligned on the 1:1 line (dashed red line), a perfect agreement between the simulated and the modeled values would have been accomplished. Also, a line of best fit was shown with a solid black line that did not represent any physical meaning but further highlighted the model bias. As indicated in Figure 6 and discussed before, the winter plot aligned better on the 1:1 line. Some dots deviated from the 1:1 line, significantly affecting RSME and MAPE. This is attributed to either data imperfections or a poor ANN architecture. Perhaps, the inclusion of another hidden layer could resolve this issue. The  $R^2$  values listed on top of each subplot were calculated from the master dataset. These values were lower than the ones calculated for each case study reported in Table 3. This decrease was expected because the model is adjusting itself to three years' worth of data rather than just one case at a time. Nevertheless, the decrease in  $R^2$  was in the range of 7 to 12% compared to the mean correlation highlighted in bold in Table 3. Further, the training, validation, and testing  $R^2$  were fairly comparable, confirming that the model was not overfitted.

The other factor affecting the model is the atmospheric pressure, which is positively correlated to particulate matter [14]. As the atmospheric pressure increases, the concentration of various pollutants also increases [14]. Low winds and poor ventilation can cause a higher concentration of air pollutants during high-pressure events. Thus, pressure may be a potential factor affecting the PM2.5 model. The inclusion of atmospheric pressure can be the subject of future research to improve the ANN model. This discussion calls for an analysis of how independent variables are contributing to the outcome.



The relative importance of each independent variable was quantified by conducting a sensitivity analysis on each master dataset [43]. The analysis was performed based on the combined testing, validation, and training datasets. It should be noted that this process was computationally expensive and labor-intensive because of the number and size of the independent variables. The result is presented in Table 4, where the importance was a value between 0 and 1, with higher values being more critical in the model performance. Based on these findings, the most important variable was precipitation for summer and spring, temperature for fall, and relative humidity for winter. Precipitation was also the second most important variable in fall and winter, which is why it was one of the emission control methods recommended in the Practical Applications section of this study. This finding is in agreement with the study conducted by Ouyang et al. 2015 [44].

**Table 3.** The ANN model quality assessment using R<sup>2</sup>, RMSE, and MAPE

| Season | Year | Train          |       |        | Validation     |       |        | Test           |       |        |
|--------|------|----------------|-------|--------|----------------|-------|--------|----------------|-------|--------|
|        |      | R <sup>2</sup> | RMSE  | MAPE   | R <sup>2</sup> | RMSE  | MAPE   | R <sup>2</sup> | RMSE  | MAPE   |
| Spring | 2010 | 0.948          | 0.048 | 62.96% | 0.941          | 0.052 | 70.88% | 0.949          | 0.047 | 67.31% |
|        | 2014 | 0.955          | 0.044 | 37.69% | 0.943          | 0.050 | 41.94% | 0.946          | 0.047 | 43.58% |
|        | 2018 | 0.952          | 0.046 | 41.99% | 0.939          | 0.051 | 48.94% | 0.938          | 0.053 | 44.29% |
|        | Mean | 0.952          | 0.046 | 47.55% | 0.941          | 0.051 | 53.92% | 0.944          | 0.049 | 51.73% |
| Summer | 2010 | 0.954          | 0.043 | 63.15% | 0.942          | 0.049 | 65.33% | 0.939          | 0.052 | 67.51% |
|        | 2014 | 0.905          | 0.063 | 72.69% | 0.892          | 0.067 | 78.65% | 0.870          | 0.071 | 79.38% |
|        | 2018 | 0.899          | 0.065 | 59.81% | 0.874          | 0.071 | 63.95% | 0.884          | 0.073 | 58.27% |
|        | Mean | 0.920          | 0.057 | 65.21% | 0.903          | 0.062 | 69.31% | 0.897          | 0.065 | 68.39% |
| Fall   | 2010 | 0.935          | 0.052 | 55.09% | 0.928          | 0.055 | 57.58% | 0.919          | 0.059 | 66.09% |
|        | 2014 | 0.947          | 0.047 | 42.77% | 0.945          | 0.047 | 45.12% | 0.930          | 0.053 | 46.15% |
|        | 2018 | 0.960          | 0.041 | 34.89% | 0.950          | 0.048 | 35.55% | 0.948          | 0.049 | 40.39% |
|        | Mean | 0.947          | 0.047 | 44.25% | 0.941          | 0.050 | 46.08% | 0.932          | 0.054 | 50.88% |
| Winter | 2010 | 0.991          | 0.021 | 22.65% | 0.981          | 0.032 | 26.20% | 0.985          | 0.024 | 25.46% |
|        | 2014 | 0.980          | 0.030 | 26.95% | 0.977          | 0.031 | 28.70% | 0.979          | 0.030 | 27.81% |
|        | 2018 | 0.980          | 0.030 | 26.34% | 0.963          | 0.042 | 29.34% | 0.967          | 0.039 | 32.03% |
|        | Mean | 0.984          | 0.027 | 25.31% | 0.974          | 0.035 | 28.08% | 0.977          | 0.031 | 28.44% |

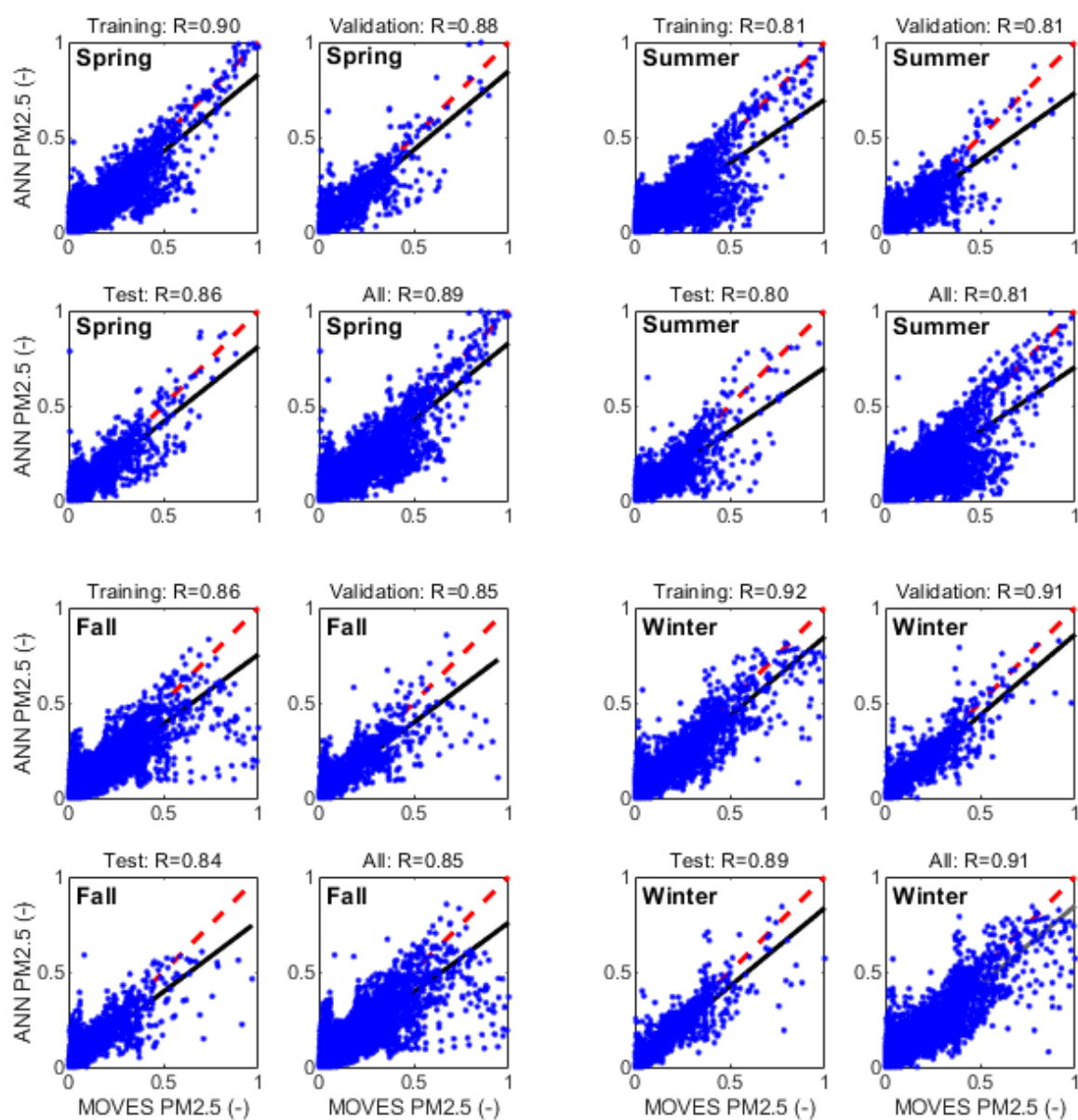


Figure 6. Seasonal performance of the ANN model

Table 4. Independent variable importance analysis

|     | Importance  |             |             |             |
|-----|-------------|-------------|-------------|-------------|
|     | Spring      | Summer      | Fall        | Winter      |
| AOD | 0.20        | 0.10        | 0.12        | 0.12        |
| P   | <b>0.33</b> | <b>0.36</b> | 0.23        | 0.24        |
| T   | 0.09        | 0.15        | <b>0.29</b> | 0.16        |
| RH  | 0.24        | 0.16        | 0.18        | <b>0.30</b> |
| W   | 0.14        | 0.23        | 0.17        | 0.18        |

## 6. Practical Applications

Particulate matter is a critical air pollutant limited by the Clean Air Act as well as the National Ambient Air Quality Standards (NAAQS). The specific laws and regulations related to emission limitations are crucial for protecting human health and public welfare. Currently, the code of federal regulations (40 CFR Part 86) uses Tier 2 and Interim Non-Tier 2 Full Useful Life Exhaust Mass Emission Standards to regulate PM<sub>2.5</sub> emissions for different vehicle types. Also, the California Air Resources Board (CARB) Regulations Title 13 covers air quality guidelines for motor vehicles. The model created by this research can assist decision-makers in predicting vehicle PM<sub>2.5</sub> emission, develop best management practices, and improve government regulations in controlling emissions.

### 6.1 Prediction

The models developed in this study can be used to estimate mobile PM<sub>2.5</sub> emissions based on climate indicators. The model's input parameters are AOD, temperature, wind speed, precipitation, and relative humidity, all of which are freely available from weather stations. The calibrated ANN model can be coded as a deployable web application capable of generating accurate PM<sub>2.5</sub> data. In addition, coupling this model with a climate model such as Weather Research and Forecasting (WRF) [45] will result in particulate matter future forecast. This is possible by using the trained model with the input data generated from a climate model instead of the in-situ and satellite data. The WRF chemistry module (WRF-Chem) [46] is a powerful tool to simulate aerosol in different wavelengths, which is the primary input to the ANN model. This forecast capability makes the model presented here more powerful. With proper trend analysis, more rigorous policies can be adopted to manage the particulate matter levels.

### 6.2 Best Management Practice

Best Management Practice (BMP) is a term most associated with water pollution and control systems. This terminology is adopted here to provide recommendations to practitioners. The method developed here can be used as an assessment tool to create multiple air quality forecast scenarios. The recommended scenarios are: i) impact of car Emission Control Systems, ii) impact of hybrid or electric vehicles, iii) impact of alternative fuels, iv) impact of public transportation, v) impact of precipitation. These scenarios are analogous with the idea behind the Clean Air Act of (1970) in which there is a significant attempt to reduce the air pollution from vehicle exhaust. Subsequently, it is required to train the ANN model based on new data generated by MOVES for each one of these scenarios. MOVES is customizable to different types of vehicles, fuels, and Clean Air Act regulations. The ANN model trained with these new adjustments will create alternative solutions for state decision-makers. New policies are introduced solely based on these solutions to achieve air pollution attainment. A prominent example of this is the ban on the sales of internal combustion engines in the state of California by 2035, as an attempt to reduce the carbon emission and fight the climate change [47].

## 7. Conclusions

Anthropogenic activities contribute majorly to the presence of PM<sub>2.5</sub> in the atmosphere, among which vehicle exhaust emissions are a large portion, especially in urban areas. This study aimed to develop an ANN model to analyze the relationship between PM<sub>2.5</sub> emissions from vehicle exhaust and the climatic variables for the state of California. The spatial data downscaling process was conducted by spatial interpolation using an equal-sized grid over the study domain. The following can be concluded based on this study:

- The pre-modeling evaluation indicated that increasing the grid density significantly improved the accuracy of modeling results.
- The Levenberg-Marquardt backpropagation algorithm with 130 neurons in each layer had the best performance.

- The maximum coefficient of determination was 0.991 for the winter of 2010, and the lowest coefficient of determination of 0.899 was for the summer of 2018, demonstrating the ANN model's capability in PM<sub>2.5</sub> predictions.
- Large variability in the dataset (having both very small and very big values) and the presence of some outliers affected the MAPE values.
- Winter had the best MAPE coefficient, where the possible reason was that winter had the most accurate estimate of AOD for vehicle emission particulate matter.
- Based on the sensitivity analysis, the most important variable among the independent variables was determined to be precipitation.

The analysis performed in this study had some inherent limitations. The MOVES can only estimate the PM<sub>2.5</sub> generated from mobile sources. The concentration of particulate matter was highly affected by the climate indicators in the atmosphere. However, nature emission is also a large source of PM<sub>2.5</sub> in California, and this can degrade the model performance during the summer season because of frequent wildfires.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1)

**Figure S1.** Seasonal PM<sub>2.5</sub> averages in the state of California for years 2010, 2014, and 2018 calculated by EPA MOVES

**Figure S2.** Seasonal AOD averages in the state of California for years 2010, 2014, and 2018 observed by MODIS satellite

**Figure S3.** Seasonal precipitation averages in the state of California for years 2010, 2014, and 2018 measured by CIMIS

**Figure S4.** Seasonal relative humidity averages in the state of California for years 2010, 2014, and 2018 measured by CIMIS

**Figure S5.** Seasonal wind speed magnitude averages in the state of California for years 2010, 2014, and 2018 measured by CIMIS

**Figure S6.** Seasonal temperature averages in the state of California for years 2010, 2014, and 2018 measured by CIMIS

**Author Contributions:** Conceptualization, A.M.; methodology, F.Y. and S.C.; software, F.Y. and S.C.; validation, F.Y.; formal analysis, A.M. and F.Y.; investigation, F.Y.; resources, F.Y.; data curation, F.Y.; writing—original draft preparation, F.Y., A.M., and S.A.; writing—review and editing, A.M., S.A., E.S., and B.R.; visualization, F.Y.; supervision, A.M.; project administration, A.M., E.S., and B.R.; funding acquisition, E.S. and B.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** The grant for this study was awarded by the United States Department of Transportation (USDOT) Pacific Southwest Region 9 University Transportation Center (UTC) to Northern Arizona University as a sub awardee.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Roy, S. Prediction of Particulate Matter Concentrations Using Artificial Neural Network. *Resour. Environ.* **2012**, *2*, 30–36, doi:10.5923/j.re.20120202.05.
2. Mohebbi, A.; Green, G.T.; Akbariyeh, S.; Yu, F.; Russo, B.J.; Smaglik, E.J. Development of Dust Storm Modeling for Use in Freeway Safety and Operations Management: An Arizona Case Study. *Transp. Res. Rec.* **2019**, 0361198119839978.
3. Mohebbi, A.; Yu, F.; Cai, S.; Akbariyeh, S.; Smaglik, E.J. Spatial study of particulate matter distribution, based on climatic indicators during major dust storms in the State of Arizona. *Front. Earth Sci.* **2020**, 1–18.
4. Grgurić, S.; Križan, J.; Gašparac, G.; AntoniĆ, O.; Špirić, Z.; Mamouri, R.; Christodoulou, A.; Nisantzi, A.; Agapiou, A.; Themistocleous, K.; et al. Relationship between MODIS based aerosol optical depth and PM10 over Croatia. *Open Geosci.* **2014**, *6*, 2–16.
5. Mohebbi, A.; Chang, H.I.; Hondula, D. WRF-Chem Model Simulations of Arizona Dust Storms. In Proceedings of the AGU Fall Meeting Abstracts; 2017.
6. Mohebbi, A.; Akbariyeh, S. Geographically Weighted Regression Analysis of Particulate Matter Levels during Major Dust Storms. In Proceedings of the Association of Environmental Engineering and Science Professors (AEESP); 2019.
7. Akbariyeh, S.; Patterson, B.M.; Kumar, M.; Li, Y. Quantification of vapor Intrusion pathways: An integration of modeling and site characterization. *Vadose Zo. J.* **2016**, *15*.
8. Wang, T.; Zhao, B.; Liou, K.N.; Gu, Y.; Jiang, Z.; Song, K.; Su, H.; Jerrett, M.; Zhu, Y. Mortality burdens in California due to air pollution attributable to local and nonlocal emissions. *Environ. Int.* **2019**, *133*, 1–9, doi:10.1016/j.envint.2019.105232.
9. Jaffe, D.; Hafner, W.; Chand, D.; Westerling, A.; Spracklen, D. Interannual variations in PM2.5 due to wildfires in the Western United States. *Environ. Sci. Technol.* **2008**, *42*, 2812–2818, doi:10.1021/es702755v.
10. McCarthy, M.C.; Eisinger, D.S.; Hafner, H.R.; Chinkin, L.R.; Roberts, P.T.; Black, K.N.; Clark, N.N.; McMurry, P.H.; Winer, A.M. Particulate matter: A strategic vision for transportation-related research. *Environ. Sci. Technol.* **2006**, *40*, 5593–5599, doi:10.1021/es062767i.
11. USEPA, U.S. Environmental Protection Agency (2014). *User Guide for MOVES2014. EPA report*; 2014;
12. PAUZI, H.M.; ABDULLAH, L. Airborne particulate matter research: a review of forecasting methods. *J. Sustain. Sci. Manag.* **2019**, *14*, 189–227.
13. Reid, S.; Bai, S.; Du, Y.; Craig, K.; Erdakos, G.; Baringer, L.; Eisinger, D.; McCarthy, M.; Landsberg, K. Emissions modeling with MOVES and EMFAC to assess the potential for a transportation project to create particulate matter hot spots. *Transp. Res. Rec.* **2016**, 2570, 12–20, doi:10.3141/2570-02.
14. Gupta, P.; Christopher, S.A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *J. Geophys. Res. Atmos.* **2009**, *114*.
15. Kamarul Zaman, N.A.F.; Kanniah, K.D.; Kaskaoutis, D.G. Estimating Particulate Matter using satellite based aerosol optical depth and meteorological variables in Malaysia. *Atmos. Res.* **2017**, *193*, 142–162, doi:10.1016/j.atmosres.2017.04.019.
16. Enotoriuwa, R.; Nwachukwu, E.; John, U. Spatial interpolation and land use regression modelling



- of air. In Proceedings of the The 18th International HSE Biennial Conference on the Oil & Gas Industry in Nigeria; Eko Hotels, Lagos, 2018; pp. 136–148.
17. Wu, Y.; Guo, J.; Zhang, X.; Tian, X.; Zhang, J.; Wang, Y.; Duan, J.; Li, X. Synergy of satellite and ground based observations in estimation of particulate matter in eastern China. *Sci. Total Environ.* **2012**, *433*, 20–30, doi:10.1016/j.scitotenv.2012.06.033.
  18. Yi, L.; Mengfan, T.; Kun, Y.; Yu, Z.; Xiaolu, Z.; Miao, Z.; Yan, S. Research on PM<sub>2.5</sub> estimation and prediction method and changing characteristics analysis under long temporal and large spatial scale - A case study in China typical regions. *Sci. Total Environ.* **2019**, *696*, 133983, doi:10.1016/j.scitotenv.2019.133983.
  19. Memarianfard, M.; Hatami, A.M. Artificial neural network forecast application for fine particulate matter concentration using meteorological data. *Glob. J. Environ. Sci. Manag.* **2017**, *3*, 333–340, doi:10.22034/gjesm.2017.03.03.010.
  20. Mohebbi, A.; Maruf, M.; Roubik, J.; Akbariyeh, S. Climate Change Impacts on Precipitation in Arid Regions: An Arizona Case Study. In Proceedings of the AGU Fall Meeting Abstracts; 2018.
  21. Wu, Z.; Fan, J.; Gao, Y.; Shang, H.; Song, H. Study on Prediction Model of Space-Time Distribution of Air Pollutants Based on Artificial Neural Network. *Environ. Eng. Manag. J.* **2019**, *18*.
  22. Vakili, M.; Sabbagh-Yazdi, S.R.; Khosrojerdi, S.; Kalhor, K. Evaluating the effect of particulate matter pollution on estimation of daily global solar radiation using artificial neural network modeling based on meteorological data. *J. Clean. Prod.* **2017**, *141*, 1275–1285.
  23. Beale, H.D.; Demuth, H.B.; Hagan, M.T. Neural network design. *Pws, Bost.* **1996**.
  24. Beale, M.H.; Hagan, M.T.; Demuth, H.B. Neural network toolbox. *User Guid. MathWorks* **2010**, *2*, 13–390.
  25. Allen, J.O.; Mayo, P.R.; Hughes, L.S.; Salmon, L.G.; Cass, G.R. Emissions of size-segregated aerosols from on-road vehicles in the Caldecott Tunnel. *Environ. Sci. Technol.* **2001**, *35*, 4189–4197.
  26. Li, X.; Dallmann, T.R.; May, A.A.; Stanier, C.O.; Grieshop, A.P.; Lipsky, E.M.; Robinson, A.L.; Presto, A.A. Size distribution of vehicle emitted primary particles measured in a traffic tunnel. *Atmos. Environ.* **2018**, *191*, 9–18.
  27. Bishop, G.A. Three decades of on-road mobile source emissions reductions in South Los Angeles. *J. Air Waste Manage. Assoc.* **2019**, *69*, 967–976.
  28. FEAT Fuel Efficiency Automobile Test Data Center Available online: [http://www.feat.biochem.du.edu/light\\_duty\\_vehicles.html](http://www.feat.biochem.du.edu/light_duty_vehicles.html) (accessed on Oct 12, 2020).
  29. Vallamsundar, S.; Lin, J. Overview of US EPA new generation emission model: MOVES. *Int. J. Transp. Urban Dev.* **2011**, *1*, 39.
  30. Guensler, R.; Liu, H.; Xu, Y.; Akanser, A.; Kim, D.; Hunter, M.P.; Rodgers, M.O. Energy consumption and emissions modeling of individual vehicles. *Transp. Res. Rec.* **2017**, *2627*, 93–102, doi:10.3141/2627-11.
  31. EPA *Population and Activity of On-road Vehicles in MOVES2014*; United States Environmental Protection Agency: Washington, DC, 2016;
  32. NASA Multi-Angle Implementation of Atmospheric Correction Available online: <https://modis-land.gsfc.nasa.gov/MAIAC.html> (accessed on Jan 20, 2020).
  33. Lyapustin, A.; Wang, Y. MCD19A2 MODIS/Terra+Aqua Land Aerosol Optical Depth Daily L2G

- Global 1km SIN Grid V006 [Data set]. Available online: <https://doi.org/10.5067/MODIS/MCD19A2.006> (accessed on Jan 20, 2020).
34. Wolfe, R.E.; Roy, D.P.; Vermote, E. MODIS land data storage, gridding, and compositing methodology: Level 2 grid. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1324–1338.
  35. CIMIS California Irrigation Management Information System Available online: <https://cimis.water.ca.gov/> (accessed on Jan 20, 2020).
  36. DMV California Department of Motor Vehicle, estimated vehicles registered by county for the period of January 1 through December 31, 2019 Available online: <https://www.dmv.ca.gov/portal/uploads/2020/06/2019-Estimated-Vehicles-Registered-by-County-1.pdf> (accessed on Sep 2, 2020).
  37. Ramos, É.M.S.; Bergstad, C.J.; Nässén, J. Understanding daily car use: Driving habits, motives, attitudes, and norms across trip purposes. *Transp. Res. part F traffic Psychol. Behav.* **2020**, *68*, 306–315.
  38. Baith, K.; Lindsay, R.; Fu, G.; McClain, C.R. Data analysis system developed for ocean color satellite sensors. *Eos, Trans. Am. Geophys. Union* **2001**, *82*, 202.
  39. Brown, J. Governor Brown Announces Federal Approval of Presidential Major Disaster Declaration for Shasta County. *Off. Gov.* 2018.
  40. Yang, J.; Hu, M. Filling the missing data gaps of daily MODIS AOD using spatiotemporal interpolation. *Sci. Total Environ.* **2018**, *633*, 677–683.
  41. LeCun, Y.A.; Bottou, L.; Orr, G.B.; Müller, K.-R. Efficient backprop. In *Neural networks: Tricks of the trade*; Springer, 2012; pp. 9–48.
  42. Skalski, P. Preventing Deep Neural Network from Overfitting Available online: <https://towardsdatascience.com/preventing-deep-neural-network-from-overfitting-953458db800a> (accessed on Sep 17, 2020).
  43. Vashist, R.; Garg, M.L. Computing the Significance of an Independent Variable Using Rough Set Theory and Neural Network. *Int. J. Res. Eng. Appl. Sci.* **2013**, *3*, 122–136.
  44. Ouyang, W.; Guo, B.; Cai, G.; Li, Q.; Han, S.; Liu, B.; Liu, X. The washing effect of precipitation on particulate matter and the pollution dynamics of rainwater in downtown Beijing. *Sci. Total Environ.* **2015**, *505*, 306–314.
  45. Wei Wang, Cindy Bruyère, Michael Duda, Jimmy Dudhia, Dave Gill, M.K.; Kelly Keene, Ming Chen, Hui-Chuan Lin, John Michalakes, Syed Rizvi, X.Z.; Judith Berner, S.H.; Fossell, K. *User's Guides for the Advanced Research WRF (ARW) Modeling System, Version 3.9*; Boulder, CO, 2018;
  46. Grell, G.A.; Peckham, S.E.; Schmitz, R.; McKeen, S.A.; Frost, G.; Skamarock, W.C.; Eder, B. Fully coupled “online” chemistry within the WRF model. *Atmos. Environ.* **2005**, *39*, 6957–6975.
  47. CA.gov Governor Newsom Announces California Will Phase Out Gasoline-Powered Cars & Drastically Reduce Demand for Fossil Fuel in California's Fight Against Climate Change Available online: <https://www.gov.ca.gov/> (accessed on Sep 28, 2020).