

Spatial study of particulate matter distribution, based on climatic indicators during major dust storms in the State of Arizona

Amin MOHEBBI (✉)¹, Fan YU¹, Shiqing CAI¹, Simin AKBARIYEH², Edward J. SMAGLIK¹

¹ Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA

² Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo, CA 93407, USA

© Higher Education Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract Arizona residents have been dealing with the suspended particulate matter caused health issues for a long time due to Arizona's arid climate. The state of Arizona is vulnerable to dust storms, especially in the monsoon season because of the anomalies in wind direction and magnitude. In this study, a high-resolution Weather Research and Forecasting (WRF) model coupled with a chemistry module (WRF-Chem) was simulated to compute the particulate matter spatiotemporal distribution as well as the climatic parameters for the state of Arizona. Subsequently, Ordinary Least Square (OLS), spatial lag, spatial error, and Geographically Weighted Regression (GWR) techniques were utilized to develop predictive models based on the climatic indicators that impacted the formation and dispersion of the particulate matter during dust storms. Census tracts were adopted to create local spatial averages for the chosen variables. Terrain height, temperature, wind speed, and vegetation fraction were designated as the most significant variables, whereas base state and perturbation pressures, planetary boundary layer height and soil moisture were adopted as supplementary variables. The determination coefficient for OLS, spatial lag, spatial error, and GWR models peaked at 0.92, 0.93, 0.96, and 0.97, respectively. These models provide a better understanding of the current distribution of the particulate matter and can be used to forecast future trends.

Keywords particulate matter, dust storm, Weather Research and Forecasting, census tracts, Ordinary Least Square, Geographically Weighted Regression

1 Introduction

Suspended particles in the air, which are called particulate matter, have been one of the hot topics for air quality studies in recent years (Fuzzi et al., 2015). Particulate matter may be directly emitted (primary sources) or formed because of the chemical reactions in the atmosphere (secondary sources). The primary source of particulate matter is either anthropogenic (human-made) or natural. Examples of anthropogenic activities are agricultural burning, farming and deforestation, industrial processes, wood and fossil fuel combustion, and entrainment of road dust into the air as a result of construction and demolition activities. Examples of natural activities are dust emission from drylands, sea salt emission from marine, and biomass burning from wildfires. Generally, anthropogenic activities generate finer particles with less than 2.5 μm in diameter ($\text{PM}_{2.5}$), whereas natural activities generate coarser particles with less than 10 μm in diameter (PM_{10}) (Karambelas, 2013).

Dust storms, as one of the major contributors to airborne soil dust particulate matter, are prevalent in regions with little to no soil moisture that are swept by high winds. The state of Arizona, with more than seven million residents, is at the risk of particulate matter due to its arid to semi-arid climate. This problem is exacerbated in the monsoon season, where the wind magnitude and direction exhibit rapid fluctuations. The volume of particulate matter generated as a result of these dust storms is approximately ten orders of magnitude larger than usual particulate matter from anthropogenic sources.

The study of particulate matter generated from dust episodes comes with high levels of uncertainty. The reason for this can be the difficulty of conducting large-scale experimental studies for natural events, limited

computational resources for numerical analysis, and lack of high-resolution observational data for theoretical developments. Another reason is that the short duration of dust episodes causing large spikes in particulate matter sensors may be treated as outliers in statistical analysis.

The purpose of this study is to develop a predictive model for PM_{10} in the state of Arizona based on climatic indicators. Considering all the shortcomings listed above, it is believed that the best way to tackle this problem is to utilize a hybrid method of a numerical simulation coupled with statistical techniques. The statistical analysis has the benefit of resulting in simplified predictive models that can be utilized by industry, academia, and state agencies to forecast, plan and alleviate the particulate matter caused problems. In this study, the statistical model is calibrated for the state of Arizona but can be extended to any region in the world, given the widespread availability of the climatic indicators.

The remainder of the paper is organized as follows: first, the previous work done in the area of the particulate matter concerning climatic indicators is summarized. Next, the Methodology section contains a brief overview of the Weather Research and Forecasting (WRF) model (the numerical solver) as well as a condensed mathematical theory of the statistical methods being used in this research. Ordinary Least Square (OLS), spatial lag, spatial error, and Geographically Weighted Regression (GWR) methods along with quality assessment techniques such as determination coefficient (r^2), Akaike Information Criterion (AIC), log-likelihood, local Moran's I , and Variance Inflation Factor (VIF) are reviewed, as well. The climatic variables, their definitions, and description are discussed next. The major contribution of this work, which is a regression equation for PM_{10} based on climatic indicators is calibrated and assessed in the Results and Discussion Section. The work is finalized with recommended practical applications, conclusions, and an in-depth limitation and future work discussion.

2 Literature review

A comprehensive literature review reveals studies that target world megacities with a severe air pollution problem in the form of particulate matter and other airborne pollutants such as nitrogen dioxide, sulfur dioxide, and ozone. The climatic parameters used in these studies were air temperature, air relative humidity, atmospheric pressure, solar radiation, mixing layer height, rainfall/precipitation, wind speed, and wind direction. The temporal scale of data collection campaigns in each study varies from one to 18 years. The nature of particulate matter is mainly anthropogenic from the volume of traffic or the use of fossil fuels contrary to what is being investigated in this manuscript. Nevertheless, the ideas, methods, and applications of these works form the basis of the current study.

- Albuquerque, NM; El Paso, TX; Las Vegas, NV; Phoenix, AZ; and Tucson, AZ.: Main meteorological factors influencing particulate matter and Ozone were investigated in five major metropolitan areas located at the South-western United States (Wise and Comrie, 2005). Results indicated that air moisture levels (relative humidity) are highly correlated with the particulate matter concentration accounting for 20%–50% of particulate matter variability. The article showed that the meteorological variability typically accounts for 20%–50% of PM variability.

- Athens, Greece: Statistical analysis on particulate matter was performed to better understand the air pollution in the Greater Athens Area (GAA) of Greece (Sfetsos and Vlachogiannis, 2010). This study proposed using a series of algorithms such as dimension reduction and Positive Matrix Factorization (PMF) coupled with the k -means clustering. The identified clusters of data were used as input in the Granger Causality method aided by the Pearson correlation. The research resulted in causal relationships between PM_{10} and meteorological patterns.

- Freiburg, Karlsruhe, Mannheim, and Stuttgart Germany: The influence of atmospheric exchange conditions represented by solar radiation, air temperature, wind speed, mixing-layer height, precipitation, and backward-trajectories on particulate matter was investigated (Rost et al., 2009). Statistical analysis revealed a negative correlation of mixing-layer height that is high particulate matter was correlated with low mixing-layer height. In the absence of precipitation, high particulate matter values from the roadside stations were reported. The determination coefficient used to assess the quality of regression was different for each case study ranging between 0.77 and 0.98. Surprisingly, no clear correlation was reported between wind speed and particulate matter concentration. It was hypothesized that the presence of the surrounding building was the reason for this irregularity.

- Seoul, South Korea: The origin and nature of particulate matter and meteorological conditions were investigated for the city of Seoul, South Korea using back trajectory analysis (Lee et al., 2011). From 2001 to 2008, the number of the days PM_{10} exceeded $100 \mu\text{g}/\text{m}^3$ were identified as 254 days. This study distinguished between biogenic and anthropogenic sources. The biogenic source, in this case, was the dust from the Gobi Desert. Through back trajectory analysis, the PM_{10} sources were identified as internal (Seoul) and external (the industrial areas in inland China and the Gobi Desert). It was shown that the high pressure over the study area and the low pressure near the emission sources were positively correlated with the particulate matter concentration.

- Switzerland: The influence of meteorology on PM_{10} trends and variability in Switzerland from 1991 to 2008 was investigated using measurements and statistical analysis (Barnpadimos et al., 2011). It was hypothesized that the relationship between PM_{10} and meteorological

variables is nonlinear. Using the Generalized Additive Models method, regression models were generated for each station and season separately. Boundary layer height, precipitation, and wind speed variability and their impact on PM_{10} were investigated, revealing a different seasonal behavior.

- Xinjiang and Beijing, China: Particulate matter, sulfur dioxide, and nitrogen dioxide levels in Urumqi, the capital of Xinjiang in north-west China, were analyzed (Mamtimin and Meixner, 2011). The focus of this study was air pollution, and temperature inversion simply called the inversion. Inversion events were identified in more than 237 days of each year, where the inversion layer height was below 850 hPa. The seasonal changes in the particulate matter were similar to other pollutants except for the winter season, which was attributed to the increase in consumption rates of fossil fuels for domestic heating. The trend of Urumqi's air pollution indicated an increase in mean annual concentrations during the study time frame. The Beijing metropolitan region PM_{10} and meteorological parameters were investigated using Wavelet and Gray analysis (Tian et al., 2014). The most significant meteorological variables were designated as relative humidity, wind speed, and atmospheric pressure. The study reported an improvement in air quality in the Beijing metropolitan region over the last decade. Through analysis of PM_{10} and meteorological factors, the study proposed measures to prevent and reduce atmospheric pollution.

The study conducted here only targets the particulate matter generated from the dust storms. Moreover, rather than using statistical methods built based on measured data, it introduces a hybrid approach of numerical/statistical. Contrary to the studies prevalent in literature, the study here is only focused on the spatial distribution of the dust storm rather than temporal evolution. Therefore, only spatial specific models are adopted.

3 Methodology

The study was conducted in two phases with phase I modeling and validation of the necessary variables using the Weather Research and Forecasting (WRF) open-source software with Chemistry Module (WRF-Chem) and phase II developing predictive models based on regression analysis. To validate the particulate matter, a point by point comparison was performed utilizing the eight measurement stations located at the Pinal County, AZ (PCAQCD 2016). Figure 1 shows a logical block diagram of the steps discussed above with phase I and phase II of the project in different colors. In the next sections, the WRF, regression models, and quality diagnostics used for assessment are discussed in more detail.

3.1 Weather Research and Forecasting with Chemistry (WRF-Chem) Module

Weather Research and Forecasting (WRF) model is a numerical solver to simulate atmospheric (synoptic and mesoscale climatic), surface, and a few subsurface hydrological variables (Skamarock, 2008; Wang et al., 2009). The WRF dynamic solver core is capable of solving equations of the conservation laws in differential form coupled with physical parameterization (Michalakes et al., 2001). To successfully conduct a WRF, first, land data and climate data are interpolated to the study area grids using the four-point bi-linear interpolation (WRF Preprocessing System). Then, WRF numerically solves the predefined equations for each node using the third-order Runge-Kutta method. WRF adopts a terrain-following coordinate system also called eta coordinate system with eta equal to one assigned to the earth surface and eta equal to zero assigned to the top of the atmosphere. A pressure of 5000 Pa is typically chosen for the top of the atmosphere.

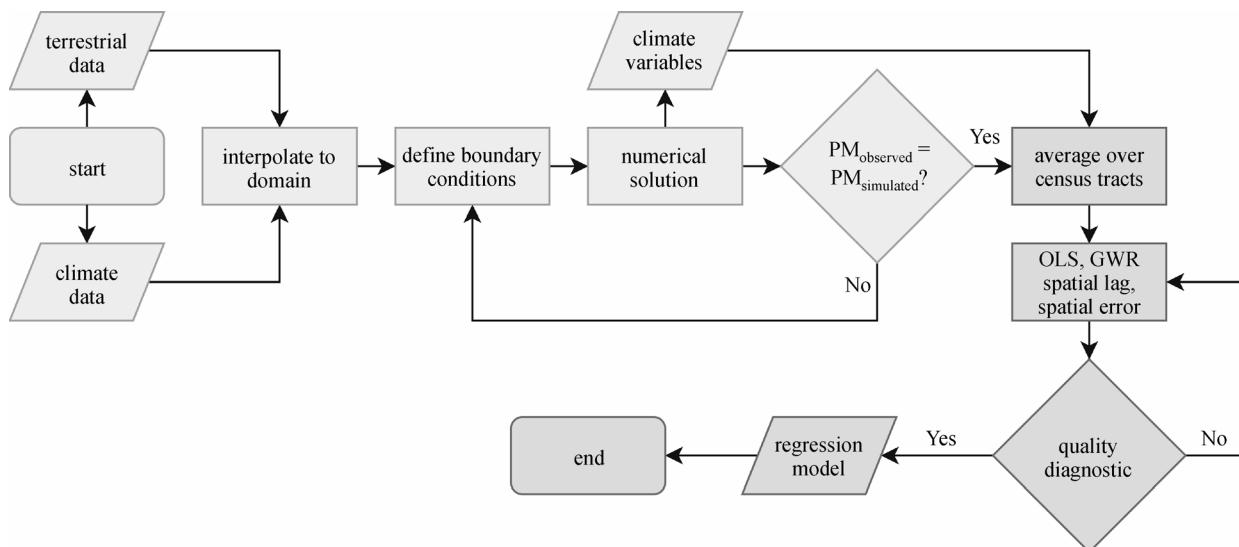


Fig. 1 Project process.

WRF users are granted the right to study and customize the source code for their projects. One of the modules that has been successfully developed and coupled with WRF is called the chemistry module (WRF-Chem) (Grell et al., 2005). The coupled weather prediction/dispersion model is used to simulate the release and transport of constituents in the air. The online nature of the WRF-Chem enables it to access anthropogenic and biogenic emission data from the Model of Emissions of Gases and Aerosols from Nature (MEGAN), Environmental Protection Agency Biogenic Emissions Inventory System (BEIS), Emission Database for Global Atmospheric Research (EDGAR), and RE analysis of the TROpospheric (RETRO) to improve the quality of the forecast.

One of the constituents that has received significant interest among others is the dust aerosol transport. There are three working dust emission schemes in WRF-Chem which are Goddard Chemistry Aerosol Radiation and Transport (GOCART) (Ginoux et al., 2001), Air Force Weather Agency (AFWA) (LeGrand et al. 2019) and University of Cologne (UoC) (Shao et al., 2011). The AFWA model has been successfully validated for the past nine dust episodes in the State of Arizona in the pilot study (Mohebbi et al., 2017 and 2019). The model-specific information such as version, set up, resolution, and microphysics scheme is thoroughly discussed in the pilot study (Mohebbi et al., 2017 and 2019) and also summarized in the Results and Discussion Section.

3.2 Linear regression analysis

Linear regression analysis is an attempt to establish a relationship between a dependent (y) and a set of independent/explanatory variables (x) by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots, \quad (1)$$

where β is an optimal vector of parameters that minimizes the error between the dependent variable (\mathbf{y}) and the hypothesized explanatory variable vector (\mathbf{x}) (De Smith et al., 2007). The existence of multiple explanatory variables and the linear nature of β classify Eq. (1) as a multiple linear regression method. The regression coefficients (β) that generates the least residuals (ϵ_i) is calculated by minimizing the difference between the fitted model and the observed values as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \epsilon_i, \quad (2)$$

or in a vector format

$$\mathbf{y} = \mathbf{x}\beta + \boldsymbol{\epsilon}, \quad (3)$$

where i represents a different instance of the observations or measurements. The null hypothesis in multiple linear regression is that the residuals are normally distributed. Equation (3) is an over-determined system solved by minimizing the sum of squared residuals as

$$\hat{\boldsymbol{\beta}} = (\mathbf{xx}^T)^{-1} \mathbf{x}^T \mathbf{y}, \quad (4)$$

where the hat symbol denotes the sample nature of the residuals, and superscript T denotes the transpose of a vector. This method is called the Ordinary Least Square (OLS) regression. OLS is a global regression model with no spatial associations between the observations. That is why, in spatial analysis, OLS is used as an exploratory method to identify the essential explanatory variables.

OLS is modified to incorporate the local spatial characteristics by

$$\mathbf{y} = \mathbf{x}\beta + \lambda \mathbf{W}\boldsymbol{\epsilon} + \boldsymbol{\epsilon}, \quad (5)$$

or

$$\mathbf{y} = \mathbf{x}\beta + \rho \mathbf{W}\mathbf{y} + \boldsymbol{\epsilon}, \quad (6)$$

where \mathbf{W} is the weight matrix typically built based on a queen or rook contiguity, and λ and ρ are error and lag coefficients, accordingly (Anselin, 2002 and 2013). Equations (5) and (6) are called the spatial error model and the spatial lag model, respectively. Zero λ and ρ reverts the model back to the classical OLS.

3.3 Geographically Weighted Regression

Geographically Weighted Regression (GWR) belongs to the family of local spatial regression models that allow parameters in β to vary spatially. The difference between GWR and OLS model is that the former constructs a unique equation for every feature in the spatial domain, allowing regression coefficients to vary in space. Depending on the type of feature (point or polygon), GWR utilizes the coordinates of each sample point or zone centroid (t_i), modifying Eq. (3) to

$$\mathbf{y} = \mathbf{x}\beta(\mathbf{t}) + \boldsymbol{\epsilon}, \quad (7)$$

where $\beta(\mathbf{t})$ is calculated by defining neighborhood features of each sample. This neighborhood criterion, in its simplest form, is a circle with a known radius (R) around the feature. More sophisticated functions have been developed for this purpose based on distance decay function, $f(d)$, accompanied by a kernel density estimation as

$$f(d) = \left[1 - \left(\frac{d}{h} \right)^2 \right]^2 \text{ for } d < R \text{ and } f(d) = 0, \text{ otherwise} \quad (8)$$

where d is the distance from the feature centroid, and h is the bandwidth parameter best calculated by minimizing the Akaike Information Criterion (AIC) explained later in the Quality Diagnostics Section. The kernel function and bandwidth are calculated for each feature (t) to create a weight matrix, $\mathbf{W}(\mathbf{t})$. The final answer to GWR is a modified version of Eq. (4) as

$$\hat{\boldsymbol{\beta}}(\mathbf{t}) = (\mathbf{x}^T \mathbf{W}(\mathbf{t}) \mathbf{x})^{-1} \mathbf{x}^T \mathbf{W}(\mathbf{t}) \mathbf{y}. \quad (9)$$

3.4 Quality diagnostics

Quality diagnostics are required to ensure the quality of the input data and the output regression model. These measures are reviewed for this work.

3.4.1 Pearson correlation coefficient

Correlation coefficient is the measure of similarity between two or more data sets and is defined as the ratio of covariance to the product of the standard deviations. Correlation Coefficient is a standardized form of covariance which falls in the range $[-1 \ 1]$ with negative one being the perfect negative correlation, positive one being a perfect positive correlation and zero as an indication of no correlation. The sample correlation coefficient is defined as

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (10)$$

with

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}, \quad (11)$$

and

$$\bar{y} = \sum_{i=1}^n \frac{y_i}{n}, \quad (12)$$

where n is the number of the samples, \bar{y} and \bar{x} are mean values of the dependent and explanatory variables, respectively. To avoid negative values in the correlation coefficient, it is common to use the Determination Coefficient (r^2) defined as the square of the correlation coefficient.

3.4.2 Akaike Information Criterion (AIC)

Akaike Information Criterion is a relative measure of the quality of a model (Akaike, 1998). *AIC* is calculated by

$$AIC = -2\ln L + 2k, \quad (13)$$

or in the corrected form as

$$AICc = -2\ln L + 2k \left(\frac{n}{n-k-1} \right), \quad (14)$$

where L is the maximum value of the likelihood function (Davidson et al., 2004) for the model, and k is the number of the parameters used in the model. It is recommended to use Eq. (14) unless $n/k > 40$ (De Smith et al., 2007). The relative nature of *AIC* only makes it useful in comparing a model to other models, and it does not provide any information about the absolute quality of a model. Given a set of candidate models, the one with the lowest *AIC* is the preferred one. The logarithmic form of the likelihood

function (L) can also be used to assess the quality of the model where the highest $\log(L)$ is relatively a more desired model.

3.4.3 Local Spatial Autocorrelation (Moran's I)

Moran's I is one of the spatial autocorrelation methods, which reveals the clustering of data with high or low values (Moran, 1950). Local Moran's I is calculated by

$$I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1, j \neq i}^n W_{ij} (x_j - \bar{x}), \quad (15)$$

with

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n W_{ij} (x_j - \bar{x})^2}{n-1}, \quad (16)$$

where i and j are different features. This method compares the data to its neighboring value based on a rook, queen, and/or a user-defined contiguity-based spatial weight. A positive I value would indicate a classification of significant locations as high-high (a high data point surrounded by high data points) and low-low (a low data point surrounded by low data points) spatial clusters. On the other hand, a negative I value is an indication of dissimilarity (high-low or low-high) in the spatial distribution of the data, which is interpreted as an outlier. Similar to other data quality assessment methods, p -values (probability of a data point being similar to its surrounding) and z -scores (deviation of a data point from its surroundings) can be calculated for I (Natrella, 2013).

3.4.4 Variance Inflation Factor (VIF)

Variables chosen for regression analysis must be independent of each other. This assumption is violated when it comes to a physical phenomenon where every variable is correlated to some extent to the other variables. The correlation between variables is called multicollinearity which is quantified by calculating Variance Inflation Factor as

$$VIF = \frac{1}{1-r^2}, \quad (17)$$

where r is the Pearson correlation coefficient for each variable (Allison, 1999). The VIF is the degree to which a regression coefficient is inflated due to multicollinearity in a model. In simple terms, the independent variables used in a regression model should not tell a similar story. VIF controls the degree of multicollinearity in regression models with VIFs larger than ten flagged as a severe model problem (Hair, 2006). The model built based on the variables with large VIFs may result in imprecise partial regression coefficients.

4 Study area

The study area for this project was the State of Arizona due to the consistent occurrence of dust storms in the vicinity of Phoenix (capital) and Tucson (major city). The State of Arizona is located in the south-western United States (latitude: 34.0489°N, Longitude: 111.0937°W) bounded on the north by Utah, on the east by New Mexico, on the west by California and Nevada, and on the south by Mexico's State of Sonora (Fig. 2). Arizona is home to four major deserts, the Great Basin, Chihuahuan, Mojave, and the Sonoran. The average statewide annual rainfall is as low as ~250 mm, making it susceptible to particulate matter issues. Most of Central and Southern Arizona has an arid climate, whereas northern Arizona has an arid to semi-arid climate. Arizona's average annual high temperature is ~30°C in the central and southern regions, and the annual low temperature is ~17 °C in the northern region (US climate data, 2019).

The WRF-Chem model was centered at latitude 34.000° N and longitude 111.684° W with a spatial resolution of 100 nodes in south-north and 100 nodes in the

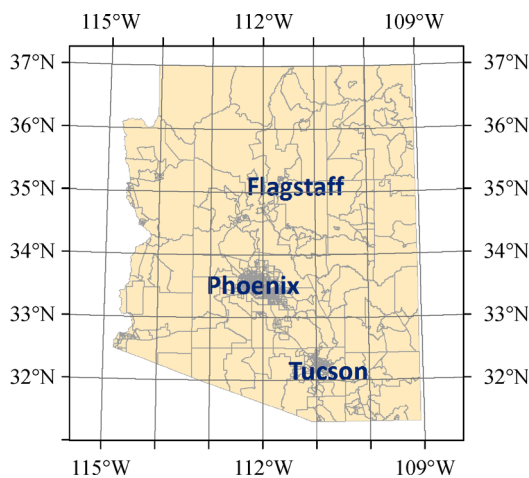


Fig. 2 Study area.

west-east with a distance of 7.5 km between each node in each direction. The model was 10% to 20% larger than the state of Arizona to generate lateral boundary conditions called relaxation zones. Map projection was Mercator with a standing longitude of 108.000°W. Land use, and topography data were from the United States Geological Survey (USGS) with a spatial resolution of 10 m and 30 arc seconds, respectively. The temporal resolution of the model was extremely high to capture dust episodes (45 s). The simulation time in each case study was five days with two days before the nominal dust episode to allow enough model spin-up time and two days after to ensure full dust storm coverage. There were 31 vertical levels in the model with the top pressure set as 5000 Pa. The model was created based on the eta coordinate system, which is a specific form of the terrain-following coordinate system used in atmospheric science.

5 Climatic indicators influencing the particulate matter

In meteorology, Planetary Boundary Layer (PBL) is the closest layer to the earth's surface characterized by strong temperature, wind speed, pressure, and moisture gradients (Arya, 1999). Coarser particulate matter transport occurs in this layer because of the limitation of threshold air shear velocity and particulate matter density. Taking into consideration the physics of the particulate matter transport, the explanatory variables can be classified as land base variables including terrain height (HGT), soil moisture (SMOIS) and vegetation fraction (VEGFRA) and atmospheric variables including temperature (T2), wind speed (U10 and V10), perturbation and base state pressures (P and PB), and Planetary Boundary Layer Height (PBLH). The land-based variables facilitate the formation, and the atmospheric variables assist in the transport and dispersion of the particulate matter.

Table 1 lists the dependent and independent variables used in this study with their official WRF-Chem name,

Table 1 Hydrological and climatic variables influencing particulate matter formation and transport

Official WRF name	Description	Dimension	Unit
P10	particulate matter with a diameter less than 10 μm	4D	$\mu\text{g} \cdot \text{m}^{-3}$
P2_5_DRY	particulate matter with a diameter less than 2.5 μm	4D	$\mu\text{g} \cdot \text{m}^{-3}$
HGT	terrain height from mean sea level	3D	m
T2	temperature at 2 m above ground	3D	K
U10	wind speed at 10 m above ground in west-east direction	3D	$\text{m} \cdot \text{s}^{-1}$
V10	wind speed at 10 m above ground in north-south direction	3D	$\text{m} \cdot \text{s}^{-1}$
VEGFRA	vegetation fraction	3D	
P	perturbation pressure	4D	Pa
PB	base state pressure	4D	Pa
PBLH	planetary boundary layer height	3D	m
SMOIS	soil moisture at four layers	4D	

description, dimension, and units. Name and unit columns are official names and units used in the WRF-Chem. These variables are either Three-Dimensional (3D) with dimensions of time, latitude, and longitude or Four-Dimensional (4D) with dimensions of time, latitude, longitude, and elevation. 3D variables were used directly in the regression analysis, but 4D variables were post-processed to remove the elevation direction by choosing the closest data layer to the earth's surface.

In this study, there were 11 variables (Table 1) and 9 case studies (Table 2) totaling 99 different spatial distributions. Figures 3 and 4 depict the spatial distribution of the

variables during the July 4th, 2014 dust storm listed in Table 2 as Case F (PM_{2.5} was omitted because of the similarity to PM₁₀). Case F was randomly selected for illustration and did not serve any particular purpose. The data are temporally averaged for the duration of the simulation and spatially for each census tract (explained later). The inspection of Figs. 3 and 4 shows that all variables are clustered in a north-east direction except for the wind speed (U10). Height (HGT) varies from zero near the Gulf of California to its highest ~2500 m near the north; Air temperature varies from ~297 K (24 °C) in the north to ~307 K (34 °C) in the south; Wind speed shows a random

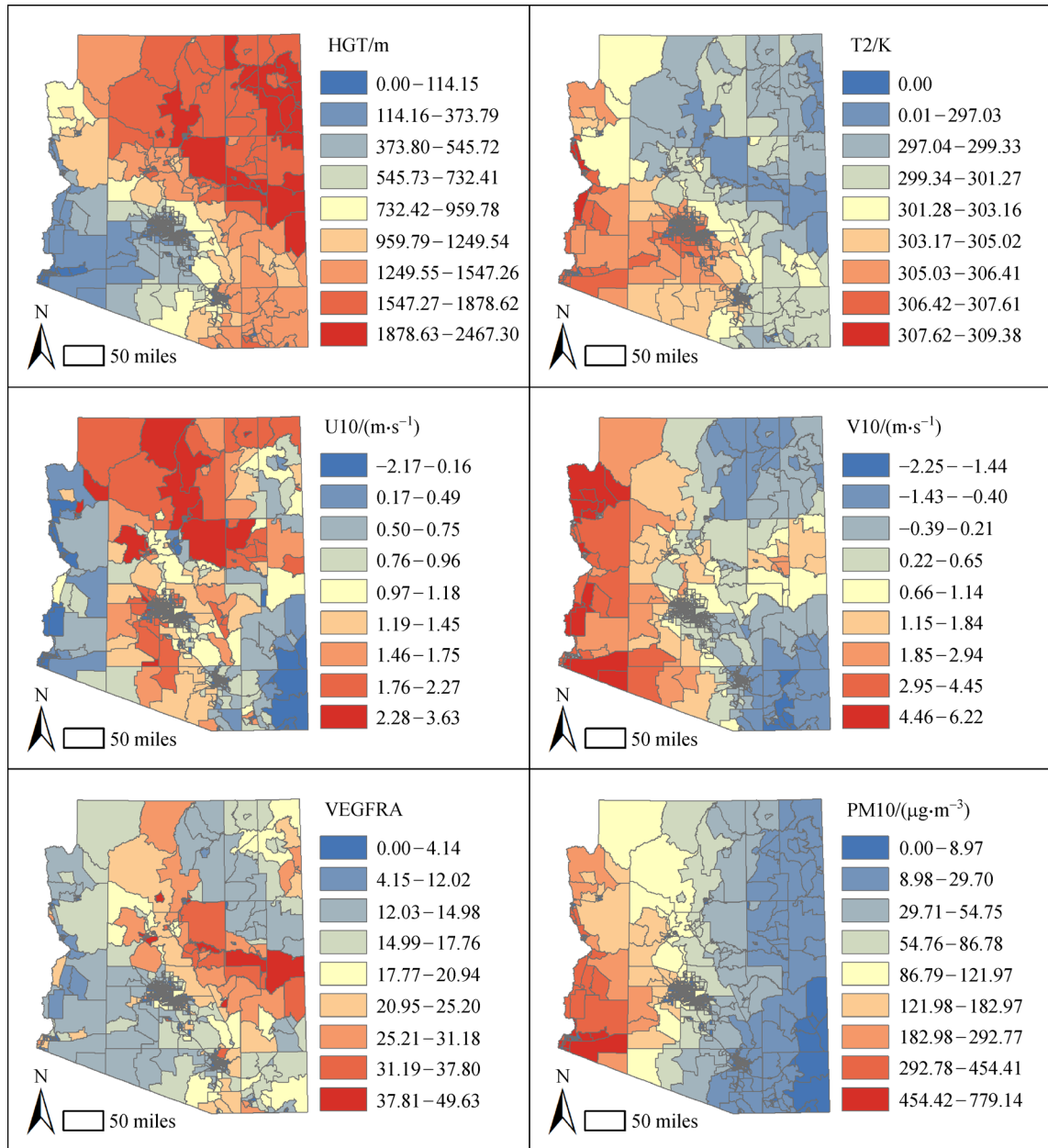


Fig. 3 Significant explanatory variables during July 4th, 2014 dust storm. The variables are averaged temporally for the duration of the simulation and spatially for each census tract.

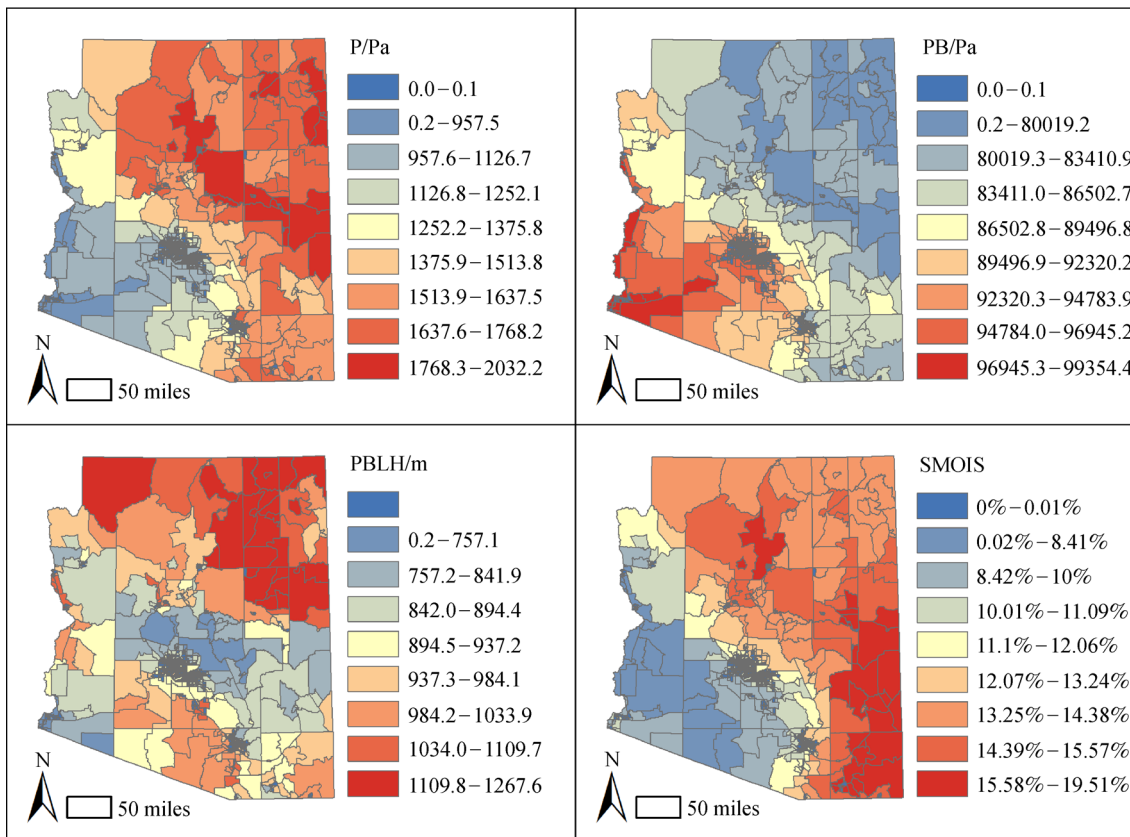


Fig. 4 Supplementary explanatory variable during July 4th, 2014 dust storm. The variables are averaged temporally for the duration of the simulation and spatially for each census tract.

Table 2 Maximum PM_{10} and $PM_{2.5}$ for the duration of simulation averaged over census tracts

Case	Model start date	Model end date	PM_{10} ($\mu\text{g}\cdot\text{m}^{-3}$)		$PM_{2.5}$ ($\mu\text{g}\cdot\text{m}^{-3}$)	
			Simulated	Observed	Simulated	Observed
A	Jul 10, 2009	Jul 15, 2009	2344.5	2133.5	395.9	360.3
B	Jan 19, 2010	Jan 24, 2010	17506.3	17156.2	2912.3	2825.0
C	Jul 3, 2011	Jul 8, 2011	4032.4	3992.0	682.1	668.5
D	Jul 19, 2012	Jul 24, 2012	687.4	680.5	115.7	120.3
E	Jun 28, 2013	Jul 3, 2013	6265.7	5451.2	1057.0	972.4
F	Jul 1, 2014	Jul 6, 2014	5992.6	6172.3	1021.0	1051.7
G	Jul 6, 2014	Jul 11, 2014	1927.4	1946.6	328.2	334.7
H	Jun 26, 2015	Jul 1, 2015	1987.1	2006.9	334.4	294.3
I	Aug 13, 2016	Aug 18, 2016	498.3	N/A	83.5	N/A

pattern, but generally, the highest is observed at the north and western part of the state with up to ~ 7 m/s; Vegetation fraction is a static variable (for a duration of simulation) with its highest located in the national forests near the city of Prescott; particulate matter concentration has its highest near the south-eastern deserts; Perturbation pressure ranges from zero near the Gulf of California to ~ 2000 Pa near the

city of Flagstaff which is positively correlated with the height; Base state pressure ranges from ~ 80 kPa to ~ 100 kPa which is negatively correlated with the height; Planetary Boundary Layer is highly dynamic with its highest near the mountainous region; Soil moisture is minimum ($\sim 4\%$) near the highest particulate matter concentration.

6 Results and discussion

During severe dust episodes, dust particles are the major contributor to the particulate matter concentration overshadowing the contribution of the background particulate matter. Validation of the particulate matter during dust storms because of the low temporal resolution of the sensors is problematic. Another issue is the large sudden pulses of dust concentrations, which typically are missed by the traditional sensors. With all these obstacles, a fair agreement between the model and observations were achieved (Fig. 5). The abscissa of Fig. 5 is the hourly average of the particulate matter for eight particulate matter

sensors, whereas the ordinate is the hourly average particulate matter calculated from 45 s simulations. The closer the distribution of the data points to the dashed redline (1 to 1 match), the higher is the precision of the model. Nevertheless, visualization can be misleading when it comes to correlation analysis. That is why the quality diagnostic of the Pearson correlation coefficient introduced in the Methodology section was used for assessment. The best correlation was achieved for Case A with a correlation coefficient of 0.87, and the worst correlation was achieved for Case D with a correlation coefficient of 0.69 (Fig. 5). Also, it looks like the model overestimates the particulate matter concentration in most occasions based on the

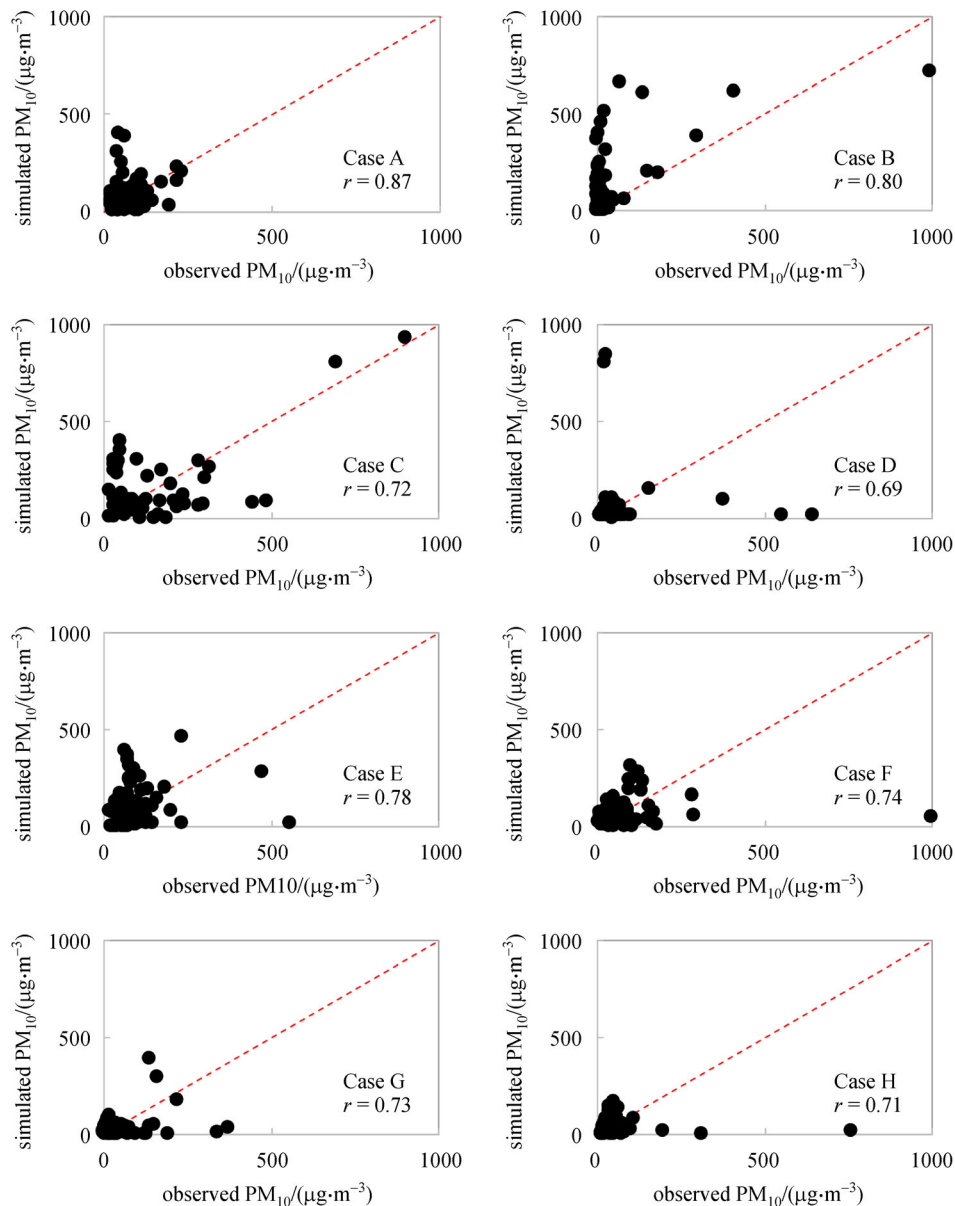


Fig. 5 Comparison of PM_{10} simulated by WRF-Chem and PM_{10} observed in Pinal County, AZ air quality stations.

distribution of the majority of data points above the dashed redline (Mohebbi et al., 2019). It is also worth to mention that to create a uniform and consistent plot, only 0 to 1000 $\mu\text{g}\cdot\text{m}^{-3}$ of PM_{10} was visualized, which contained up to 96% of the data points.

The validated particulate matter data at the time of the dust storm was used to develop OLS, spatial lag, spatial error, and GWR models. Data was averaged in the time direction to remove the time dependency as the goal of the study was to study the dust statistics, spatially. Based on the theoretical background of the dust governing equations, dust emission sources are highly impacted by their surrounding topography (Ginoux et al., 2001). That is why data was aggregated into census tracts (Census, 2010) using spatial mean resulting in local averages. Each census tract covers up to 8000 people, described by the United States Bureau of Census. The choice of census tracts was made because of their high density near dust hot spots (better capturing data variations) as well as their easy accessibility.

6.1 Spatial distribution of PM_{10} in census tracts

The novelty of this work was using the census tracts to establish a spatial regression model. That is why the data was converted from point to surface. This process was different for the simulation and the observation data because of their different spatial and temporal resolutions. The high-resolution simulation data was averaged over the census tracts, while the low-resolution observation data was first interpolated/extrapolated over the state of Arizona and then downscaled over census tracts. The observation values were interpolated/extrapolated from pointwise sensor readings using three methods of the nearest neighbor, natural neighbor, and linear. The best results were achieved using the natural neighbor method (reported values in Table 2). Although Table 2 values suggest a fair agreement, the interpolated/extrapolated observation values should be used with caution due to the uncertainty of the interpolation/extrapolation methods as well as the limited number of available sensors.

Figure 6 shows the normalized mean distribution of PM_{10} for nine dust episodes listed in Table 2. $\text{PM}_{2.5}$ results were very similar and hence omitted here. Case studies showed similarities in terms of spatial distribution. Maximum dust concentration occurred in the south-western part of the state near La Paz, Yuma and Pima counties. According to Table 2, PM_{10} and $\text{PM}_{2.5}$ reached 17506.32 $\mu\cdot\text{gm}^{-3}$ and 2912.33 $\mu\cdot\text{gm}^{-3}$, respectively, which were considerably high. The high values are documented (Hyde et al., 2018) and expected due to the Sonoran Desert dust composition and low moisture as well as high and fluctuating wind patterns because of the North American Monsoon (Mohebbi et al., 2019).

6.2 Global regression

6.2.1 Case by case exploratory study

The primary reason to use regression analysis is to model a phenomenon to better understand or predict it. Classic OLS regression was performed on mean PM_{10} aggregated for each census tract. For the sake of uniformity in visualization, OLS results were normalized by dividing the original PM_{10} to its respective maximum in time and space. The outcome after normalization is depicted in Fig. 7 with the dashed redline as a perfect one to one match between the regression and the WRF-Chem output and the solid blue line as the best linear fit for the data. The determination coefficient range was 0.75 to 0.92 (Table 3), which shows a promising agreement between the regression model and WRF-Chem simulation. Other OLS models such as spatial lag and spatial error were also tested utilizing a weight matrix built based on first-order queen contiguity. The spatial lag model improved the determination coefficient by 0 to 9%, whereas the spatial error improved the determination coefficient by 4% to 16% (Table 3). In both cases, the more significant improvement was case B with the lowest determination coefficient ($r^2 = 0.76$), and the lowest improvement was case H with the highest determination coefficient ($r^2 = 0.91$). This improvement is due to the logarithmic nature of the spatial lag and spatial error models, which lowers the variable ranges as well as the utilization of the weight matrix, which creates a kernel that prioritizes the nearest data. The average spatial lag ρ was calculated as 0.19 with a standard deviation of 0.10, and the average spatial error λ was calculated as 0.72, with a standard deviation of 0.11. *AIC* and log-likelihood calculated for the OLS model designated case I as the best fit (2734, and -1357, respectively) whereas the determination coefficient designated case A and case C as the best fit ($r^2 = 0.92$) (Table 3). Although the spatial lag and spatial error models provided a better fit, classic OLS was chosen due to the simplicity and fewer parameters needed in model calibration.

The climatic variables used in this study are listed from the most significant to the least significant in Table 4. The degree of significance was decided based on variables' *p*-values, *z*-scores, and VIFs. Terrain height, temperature, wind speed at west-east and north-south directions and vegetation fraction were the most significant variables whereas perturbation pressure, base state pressure, planetary boundary layer height, and soil moisture either were not physically significant in particulate matter transport or were collinear with other variables (supplementary variables). In particular, perturbation and base state pressures were collinear with temperature, vegetation fraction was collinear with soil moisture, and planetary boundary layer height was collinear with pressure and temperature.

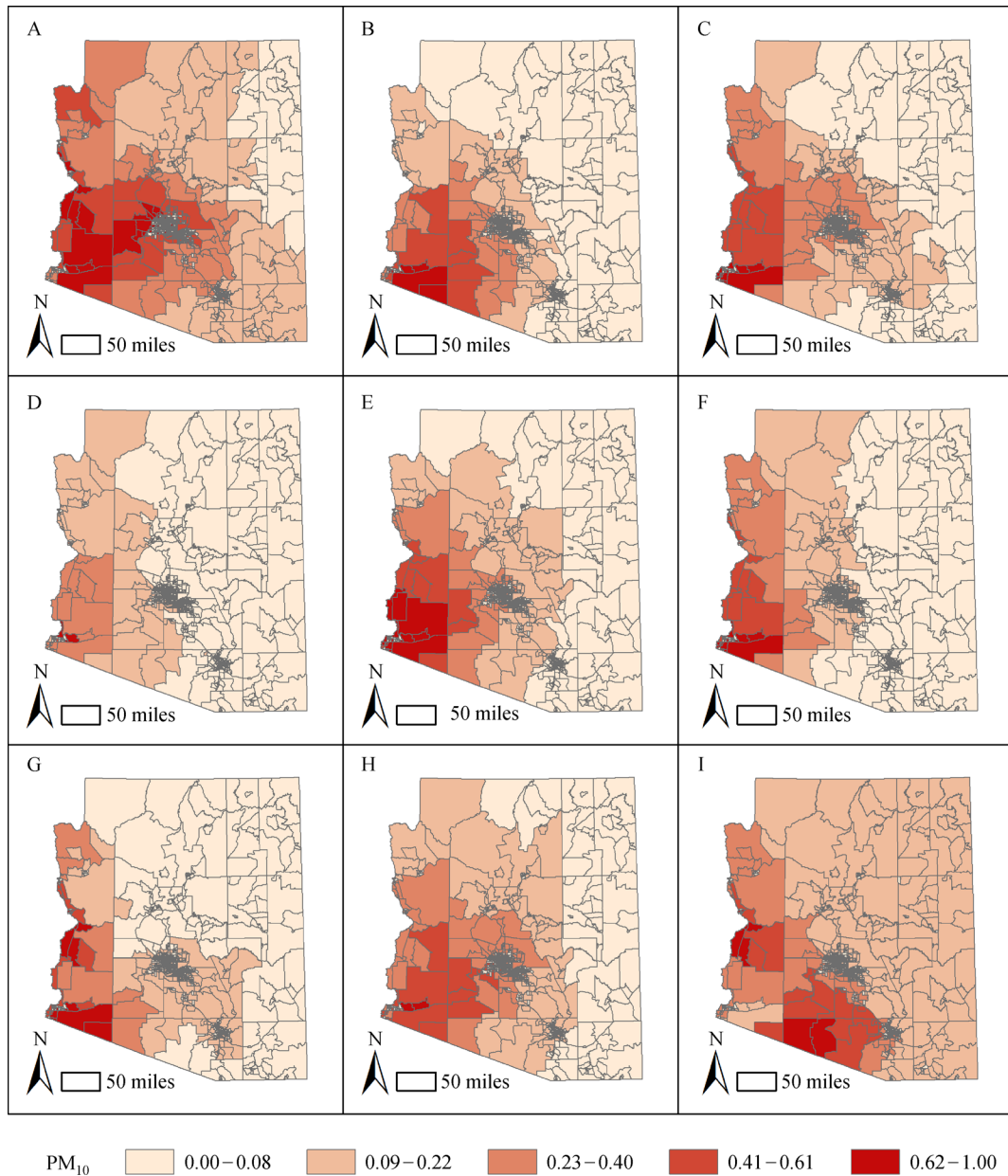


Fig. 6 Normalized mean PM_{10} modeled by WRF-Chem averaged temporally for the duration of the simulation and spatially for each census tract.

Relative humidity was also selected in the preliminary study consistent with the past literature (Wise and Comrie, 2005), but eventually, it was dropped because of the high collinearity with temperature. The collinearity of the variables was verified by the VIF diagnostic, where all the variables mentioned above had positive VIFs less than ten. However, their collinearity was not high enough to entirely forgo the variable from the analysis except for the relative humidity.

Regression coefficients for each case study are listed in Table 4. The intercepts were all equal to zero up to four decimal places, which could also be confirmed from Fig. 7.

Generally speaking, very small coefficients are an indication of the respective variable not being significant. It was noted that the soil moisture coefficients were more substantial than other variables, but this was mainly because of the low soil moisture in the State of Arizona. Another interesting result is that the particulate matter was negatively correlated with perturbation pressure, whereas it was positively correlated with base state pressure. This was noted in the earlier pilot study when both pressures were combined to reduce the degree of freedom in the optimization process, and the result was not satisfactory. This conclusion is not based on the pressure coefficients in

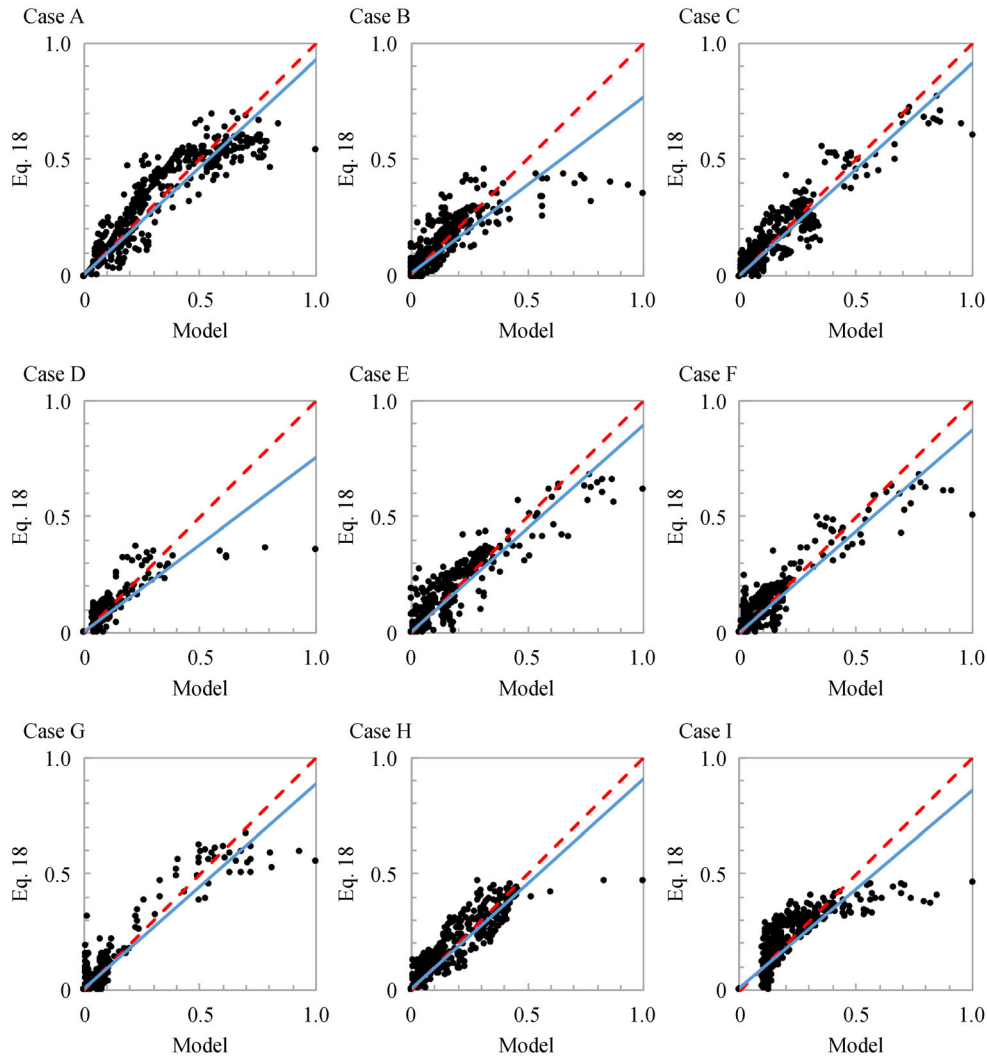


Fig. 7 Normalized PM_{10} calculated by classical OLS versus modeled PM_{10} , the dashed red line represents a one to one match, and the solid blue line represents the best linear fit.

Table 3 Classic, spatial lag, and spatial error performance comparison

Case	r^2			AIC			Log (L)		
	Classic	Spatial lag	Spatial error	Classic	Spatial lag	Spatial error	Classic	Spatial lag	Spatial error
A	0.92	0.93	0.96	8408	7680	8408	-4194	-4144	-3830
B	0.76	0.83	0.89	15971	15475	15061	-7976	-7727	-7521
C	0.92	0.92	0.95	12772	12689	12023	-6376	-6333	-6002
D	0.75	0.76	0.79	3689	3659	3527	-1835	-1818	-1754
E	0.89	0.91	0.95	14250	13943	13286	-7115	-6960	-6633
F	0.87	0.88	0.92	14469	14312	13824	-7224	-7145	-6902
G	0.89	0.90	0.93	9091	8989	8583	-4536	-4483	-4281
H	0.91	0.91	0.94	9701	9663	9125	-4841	-4820	-4552
I	0.85	0.90	0.93	2734	2234	1823	-1357	-1106	-901

Table 4 Coefficients used in the classic OLS regression for PM₁₀

Case	C_HGT	C_T2	C_U10	C_V10	C_VEGFRA	C_P	C_PB	C_PBLH	C_SMOIS
A	0.02	2.60	0.61	-0.49	0.16	-0.13	-0.01	-0.03	-136.18
B	0.52	-28.69	35.92	0.83	-0.24	-0.14	0.09	-0.14	-343.65
C	0.43	-19.75	-16.77	18.18	-0.25	0.10	0.06	-0.05	-973.55
D	0.03	-1.23	0.05	1.07	-0.04	0.00	0.00	0.01	-22.45
E	0.75	-15.55	29.01	9.58	-0.80	-0.50	0.06	-0.38	0.55
F	1.49	-40.72	-11.31	30.19	-1.95	-0.78	0.14	-0.15	533.50
G	0.12	-1.22	7.78	2.58	0.20	-0.19	0.01	-0.01	-171.22
H	0.10	-2.69	7.88	1.64	0.30	-0.08	0.01	-0.01	-1.86
I	0.00	-0.13	0.00	0.28	0.01	0.00	0.00	0.01	-10.60

Table 4 but is based on the plot of particulate matter versus both pressures, which is not shown here because of the limited space. The wind speed coefficients fluctuating between negative and positive was because of the wind pattern anomaly in the monsoon season. Lastly, vegetation fraction near particulate matter hot spots was identified as shrubs based on the USGS land-use model used in the WRF-Chem simulation.

6.2.2 Global regression predictive model

The goal of the previous section was to show the capability and robustness of OLS in particulate matter transport. Now the goal is to develop an OLS model that can be used for every dust episode calculation and future forecast. To this end, all case studies were merged to create one data set, which resulted in Eq. (18) as

$$\begin{aligned}
 PM_{10} = & 0.51HGT - 15.81T2 + 11.44U10 + 10.81V10 \\
 & - 0.95VEGFRA - 0.20P + 0.05PB \\
 & - 0.20PBLH - 415.94SMOIS. \quad (18)
 \end{aligned}$$

Equation (18) will result in $r^2 = 0.88$ and AIC of 11745, which shows a good agreement with the model data. The supplementary variables may be dropped to have a simplified and more practical model as

$$\begin{aligned}
 PM_{10} = & -0.06HGT + 0.19T2 + 13.32U10 + 25.69V10 \\
 & - 1.05VEGFRA, \quad (19)
 \end{aligned}$$

with $r^2 = 0.79$ and $AIC = 12425$. This simplified model is the major contribution of this work and can be used for any case study as a predictive tool in the State of Arizona. Figure 8 illustrates the comparison between PM₁₀ calculated by Eq. (18) and Eq. (19), for all the case studies. The difference of determination coefficients between Eq. (18) and Eq. (19) is seen by comparing the slope of the solid blue line.

6.3 Local regression

It is a common practice to use OLS exploratory findings in GWR models. GWR was performed on mean PM₁₀ aggregated for each census tract. Contrary to OLS, single case studies were not analyzed because it was not feasible to summarize spatially variable coefficients to propose a universal model. Also, based on the OLS exploratory study, only five significant variables were considered in the GWR study. A fixed Gaussian kernel was used to generate geographic weighting in the model. The extent of the kernel (bandwidth) was determined using the $AICc$. The bandwidth was 110.56 km calculated based on the $AICc$ of 10276.93.

GWR is a local regression model, so it was expected to perform superior to its counterpart OLS. In this study, the GWR maximum determination coefficient increased to 0.97, which was 12.94% higher than the OLS maximum determination coefficient (Fig. 9). This high determination coefficient was observed near central Arizona, where census tracts were denser and smaller in the area. There was a downside to this increase, and that was the decrease of the determination coefficient near Arizona boundaries. Generally speaking, local spatial regression models do not perform well near boundaries due to a drop in the number of neighboring features.

Residual in regression models is treated as a test for the null hypothesis. This concept is further discussed in the next section via local spatial autocorrelation (Moran's I), but a qualitative study of residuals has also proved to be beneficial. The null hypothesis in this study was complete randomness or no clustering of residual distribution. This means if the residual distribution was not random, there was a chance that the independent variables used in the original model were not enough or perhaps inappropriate. In the case of this study, residuals close to zero were not significant, and negative/positive residuals were dispersed out, so no apparent clustering was observed (Fig. 9).

Similar to OLS, GWR passed through the origin resulting in intercepts of zero (Fig. 10). The OLS

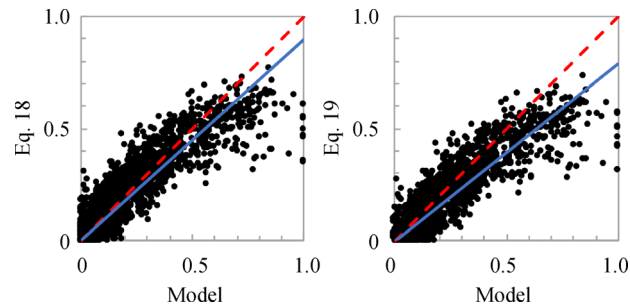


Fig. 8 Comparison of PM_{10} calculated by Eq. 18 and Eq. 19, for all the case studies.

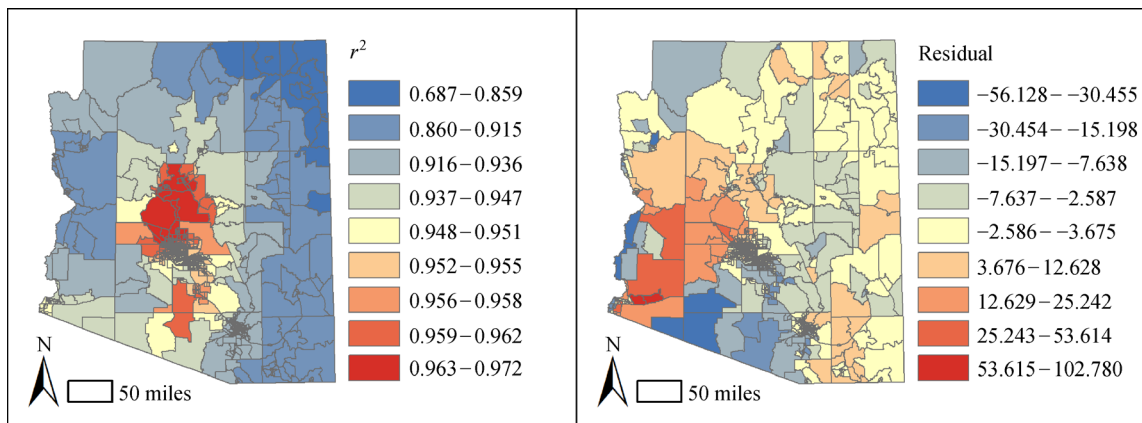


Fig. 9 GWR determination coefficient and residuals.

coefficients in Eq. (19) fell within the range of GWR coefficients except for the vegetation fraction. Taking a step back and studying OLS vegetation fraction coefficients for each case study, a fluctuation is observed, which could be due to the land-use model resolution. Nevertheless, GWR exhibits a more uniform distribution for the regression coefficients, which comes from the spatial nature of this method.

6.4 Local spatial autocorrelation (Moran's I) of residuals

Local Moran's I was computed using the residuals as an input to decide whether the independent variables used to build the regression model were appropriate. To this end, z -scores and p -values for the Moran's I statistics were calculated, resulting in the cluster map for residuals (Fig. 11). The cluster map was classified as H-H, L-L, L-H, and H-L described in the Methodology section. The weighting matrix used was constructed based on the first order queen contiguity. Ideally, complete randomness was expected for the local spatial autocorrelation output. However, some degree of clustering is always expected merely due to randomness even if the regression model performs flawlessly. In the case of this study, OLS exhibited a significant L-L and H-H clustering all over the study area, whereas GWR only exhibited clustering in

the south-western region of Arizona. Revisiting Fig. 6 showed high particulate matter concentration and frequency of occurrence in that region. This means the clustering was because of the inability of the regression model to adapt itself with sudden spikes of the particulate matter. L-H and H-L clusters were statistical outliers for a 95% confidence level.

The graphical representation of residual clusters (Fig. 11) can be well misleading because of the discrepancies in census tract areas. To have a quantitative analysis, the affected census tracts count and area were calculated (Table 5). OLS covered 47% of the significant census tracts, whereas GWR only covered 16% of the significant census tracts. Also, 51% of the residuals resulted from OLS were not significant compared to 81% of the residuals resulted from GWR. The promising performance of GWR weights the decision toward local regression models and leaves out the global regression models as excellent initial exploratory methods.

7 Practical applications

The global and local regression models developed in this study can be used to estimate particulate matter concentration during dust storms in the State of Arizona. Both

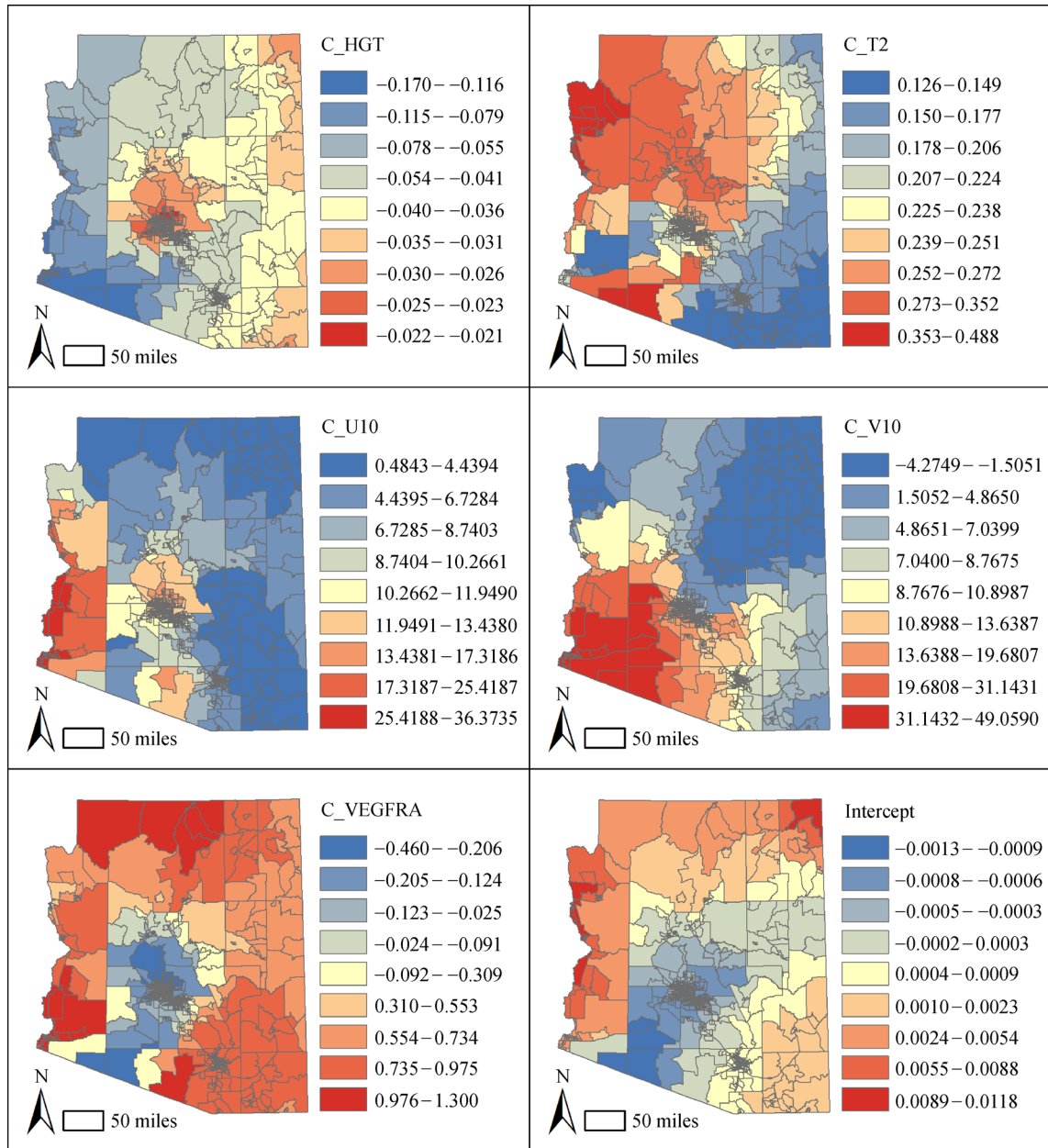


Fig. 10 GWR coefficients for PM₁₀.

models comprised of five independent variables of which terrain height and vegetation fraction are readily available. The remaining temperature and wind speed variables are available from standard *in situ* stations, radar or satellite products. Another practical aspect of the regression models is the ability to forecast future trends. Climate models such as Representative Concentration Pathways (RCP) have forecast values for climate indicators based on three greenhouse gas emission scenarios up to the year 2100 (Bruyere et al., 2015). Coupling this climate model and the regression models developed here could lead to particulate matter forecasts allowing scientists and policymakers to develop best management practices to control and reduce

the particulate matter production and dispersion rate. Further, with the particulate matter forecast in hand, it is possible to study the impact of the climate change on particulate matter formation and transport, or the impact of particulate matter on climate change, which is currently a controversial research topic.

8 Conclusions

WRF-Chem simulation results were used to develop new global and local regression models for the particulate matter formation and transport at the time of dust storms in

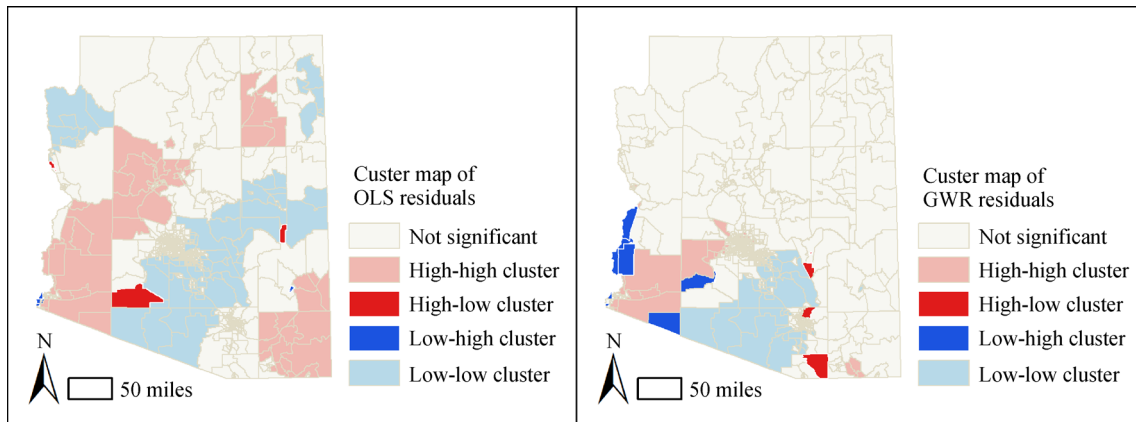


Fig. 11 Cluster map of OLS and GER residuals.

Table 5 OLS and GWR local spatial autocorrelation

Cluster	Count		Areal %	
	OLS	GWR	OLS	GWR
H-H	69	40	24%	6%
L-L	108	55	23%	10%
H-L	5	4	1%	1%
L-H	5	7	0%	2%
Not significant	1339	1420	51%	81%
Sum	1526	1526	100%	100%

the State of Arizona. The following can be concluded based on this study:

- The work benefits from the use of census tracts to aggregate the dependent and independent variables. In the process of regression, independent variables were classified into significant and supplementary based on their p -values, z -scores, and VIFs. The major contribution of the work was the regression model built for the soil dust particulate matter based on the significant variables of terrain height, temperature, wind speed, vegetation fraction, and supplementary variables of base state and perturbation pressures, planetary boundary layer height, and soil moisture.

- Particulate matter transport correlation with climatic indicators has been widely investigated for anthropogenic sources leaving a gap in the literature for soil dust particulate matter caused by dust storms. The goal of this study was to develop a regression model for soil dust particulate matter transport. It was hypothesized that the spatial nature of the soil dust particulate matter would best be handled by a spatial regression model such as GWR. To better identify the independent variables, an exploratory OLS model was developed. The determination coefficient for both methods peaked at 0.92 and 0.97, revealing the quality of the regression model developed.

- Spatial lag and spatial error models were tested, which

improved the model determination coefficient by 9% and 16%, respectively. The fitting parameters of ρ and λ were calculated as 0.19 and 0.72, accordingly. The weighting matrix was built based on the first order queen contiguity. Both models performed superior to the classic OLS, but due to the simplicity and practical nature of the study, the OLS method was recommended.

9 Limitations and future work

The analysis performed in this study has some inherent limitations, listed below:

- There are 1526 census tracts in the State of Arizona with a higher density near the cities of Phoenix and Tucson. The spatial manipulation to downscale the climate simulation to census tracts results in 46 empty polygon features. Although these features were removed from the regression studies, they introduce errors to the analysis. The extent of this error is not known at this point, but it is dependent on the resolution of the climate simulation. Increasing the climate simulation resolution will lower this error but at the cost of increased run time, processor, memory, and storage. Moreover, WRF-Chem model temporal (45 s) and spatial resolution (7.5 km by 7.5 km), map projection (Mercator), and land-use data

resolution (10 m) in the study were chosen based on the validation results from the pilot study. A comprehensive sensitivity analysis is required to quantify the impact of resolution in the final regression model.

- Land use plays an essential role in dust emission source characterization. However, the use of categorical data, such as land-use, is not permitted in regression models. Use of local statistics such as the aggregation of data into census tracts and utilization of the soil moisture and vegetation fraction variables were attempted to alleviate this problem but a new erodibility function incorporating topography, land-use, soil moisture and vegetation fraction will lower the degree of freedom in the regression model and improve the quality of the fit.

- Validation of the simulation against an observational data is mandatory not only for WRF-Chem but also for any numerically solved equations. This study has been made possible with one dependent and nine independent variables through nine case studies. Validation of each one of these variables is a separate project which is out of the scope of this study. Nevertheless, the study was not built upon unreliable data because the dependent variable was extensively validated in the pilot study (Mohebbi et al., 2019), loosely giving the independent variables some credibility. Although a fair agreement was reached in the previous study, WRF-Chem occasionally fails to retrieve the particulate matter peaks (Hyde et al., 2018). The uncertainty that comes with this should be investigated to constrain the statistical model.

- Climatic phenomena are stochastic events as opposed to random events. The null hypothesis in regression analysis is the complete randomness of samples chosen from the population. The nine case studies were selected based on the extreme weather conditions covered in media, particulate matter measurement sensors, and past published literature. These samples may or may not be a random representation of the particulate matter transport phenomenon. Given the fact that running continuous numerical models is not feasible because of the limitations in resources, other statistical methods such as Getis-Ord G_i^* (Getis and Ord, 1992) should be used to ensure the independence of the case studies.

Acknowledgements The grant for this study was awarded by the University Transportation Centers of the US Department of Transportation to Northern Arizona University as a sub awardee. The author would like to thank Dr. Fernando Sánchez-Trigueros of the University of Arizona for his constructive feedback that significantly contributed to improving the final version of the manuscript. Also, the work of Mr. Gabriel Green, the first author's former graduate student in developing the initial pilot study model is greatly appreciated.

References

Akaike H (1998). Information theory and an extension of the maximum likelihood principle. In: Selected papers of Hirotugu Akaike. New

- York: Springer: 199–213
- Allison P D (1999). Multiple Regression: A Primer. New York: Pine Forge Press
- Anselin L (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3): 247–267
- Anselin L (2013). *Spatial Econometrics: Methods and Models*. New York: Springer Science & Business Media
- Arya S P (1999). *Air Pollution Meteorology and Dispersion*. New York: Oxford University Press
- Barmpadimos I, Hueglin C, Keller J, Henne S, Prévôt A S H (2011). Influence of meteorology on PM₁₀ trends and variability in Switzerland from 1991 to 2008. *Atmospheric Chemistry and Physics*, 11(4): 1813–1835
- Bruyere L, Monaghan J, Steinhoff F, Yates D (2015). Bias-corrected CMIP5 CESM data in WRF/MPAS intermediate file format. *Climate Dynamics*, 2014, 43(7–8): 1847–1856
- Census (2010). *Geography Program*. United States Census Bureau
- Davidson R, MacKinnon J G (2004). *Econometric Theory and Methods*. New York: Oxford University Press
- Fuzzi S, Baltensperger U, Carslaw K, Decesari S, Denier Van Der Gon H, Facchini M C, Fowler D, Koren I, Langford B, Lohmann U, Nemitz E, Pandis S, Riipinen I, Rudich Y, Schaap M, Slowik J, Spracklen D V, Vignati E, Wild M, Williams M, Gilardoni S (2015). Particulate matter, air quality and climate: lessons learned and future needs. *Atmos Chem Phys*, 15(14): 8217–8299
- Getis A, Ord J K (1992). The analysis of spatial association by use of distance statistics. *Geogr Anal*, 24(3): 189–206
- Ginoux P, Chin M, Tegen I, Prospero J M, Holben B, Dubovik O, Lin S J (2001). Sources and distributions of dust aerosols simulated with the GOCART model. *J Geoph Res: Atmos*, 106(D17): 20255–20273
- Grell G A, Peckham S E, Schmitz R, McKeen S A, Frost G, Skamarock W C, Eder B (2005). Fully coupled 'online' chemistry within the WRF model. *Atmos Environ*, 39(37): 6957–6975
- Hair J F (2006). *Multivariate Data Analysis*. India: Pearson Education
- Hyde P, Mahalov A, Li J (2018). Simulating the meteorology and PM₁₀ concentrations in Arizona dust storms using the Weather Research and Forecasting model with Chemistry (Wrf-Chem). *J Air Waste Manag Assoc*, 68(3): 177–195
- Karambelas A (2013). The interactions of biogenic and anthropogenic gaseous emissions with respect to aerosol formation in the United States. Dissertation for Doctoral Degree. Madison: University of Wisconsin
- Lee S, Ho C H, Choi Y S (2011). High-PM₁₀ concentration episodes in Seoul, Korea: background sources and related meteorological conditions. *Atmos Environ*, 45(39): 7240–7247
- LeGrand S L, Polashenski C, Letcher T W, Creighton G A, Peckham S E, Cetola J D (2019). The AFWA dust emission scheme for the GOCART aerosol model in WRF-Chem v3. 8.1. *Geosci Model Dev*, 12(1): 131–166
- Mamtin B, Meixner F X (2011). Air pollution and meteorological processes in the growing dryland city of Urumqi (Xinjiang, China). *Sci Total Environ*, 409(7): 1277–1290
- Michalakes J, Chen S, Dudhia J, Hart L, Klemp J, Middlecoff J, Skamarock W (2001). Development of a next-generation regional weather research and forecast model. *Developments in Teracomput-*

- ing, 269–276
- Mohebbi A, Chang H I, Hondula D (2017). WRF-Chem model simulations of Arizona Dust Storms. AGU Fall Meeting Abstracts
- Mohebbi A, Green G T, Akbariyeh S, Yu F, Russo B J, Smaglik E J (2019). Development of dust storm modeling for use in freeway safety and operations management: an Arizona case study. *Transportation Research Record*, 2673(5): 175–187
- Moran P A P (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1–2): 17–23
- Natrella M (2013). NIST/SEMATECH E-handbook of Statistical Methods. National Institute of Standards and Technology
- PCAQCD (2016). Pinal County Air Quality Control District. Pinal County, Florence, AZ
- Rost J, Holst T, Sähn E, Klingner M, Anke K, Ahrens D, Mayer H (2009). Variability of PM₁₀ concentrations dependent on meteorological conditions. *Int J Environ Pollut.*, 36(1–3): 3–18
- Sfetsos A, Vlachogiannis D (2010). A new approach to discovering the causal relationship between meteorological patterns and PM₁₀ exceedances. *Atmos Res*, 98(2–4): 500–511
- Shao Y, Ishizuka M, Mikami M, Leys J F (2011). Parameterization of size-resolved dust emission and validation with measurements. *Int J Environ Pollut*, 116(D8): D08203
- Skamarock W C, Klemp J B, Dudhia J, Gill D O, Barker D M, Wang W, Powers J G (2008). A Description of the Advanced Research WRF version 3. Tech Note, 1–96
- De Smith M J, Goodchild M F, Longley P (2007). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Leicester: Troubador Publishing Ltd
- Tian G, Qiao Z, Xu X (2014). Characteristics of particulate matter (PM₁₀) and its relationship with meteorological factors during 2001–2012 in Beijing. *Environ Pollut*, 192: 266–274
- US climate data (2019). *Climate Arizona-Phoenix*. U.S. Climate Data
- Wang W, Bruyere C, Duda M, Dudhia J, Gill D, Lin H C, Michaelakes J, Rizvi S, Zhang X (2009). *WRF-ARW Version 3 Modeling System User's Guide*. Mesoscale & Microscale Meteorology Division. Boulder: National Center for Atmospheric Research
- Wise E K, Comrie A C (2005). Meteorologically adjusted urban air quality trends in the southwestern United States. *Atmos Environ*, 39 (16): 2969–2980