Modelling Air Pollutant Concentration in Utah, Wyoming, Nevada, Colorado, and Arizona in 2019

Young Jun Jeong[1] and Hanwen Cao[2]

[1] 1006046468; youngjun.jeong@mail.utoronto.ca

[2] 1005112637; hanwen.cao@mail.utoronto.ca

Abstract: This study aims to investigate the spatial patterns and distribution of air pollution concentrations in Utah and its four contiguous states in the United States, focusing on nitrogen dioxide (NO_2), ozone (O_3), and particulate matter 2.5 ($PM_{2.5}$). By using Spatial interpolation methods such as IDW and Kriging, air pollution surfaces were produced to demonstrate their spatial patterns. In addition, spatial clustering was used to classify areas with similar air quality.The results of this study enhance our understanding of the exposure to air pollution and its potential effects. Moreover, it may help decision-making policymakers and the governments on air pollution vulnerable area monitoring and providing solutions to the issues.

Keywords: Air Pollutant; Spatial Interpolation; Spatial Clustering

1. Introduction

A healthy environment is closely related to human health, as humans interact with their surroundings all the time. Air quality is a variable that determines whether an environment is healthy or not. Inhaling relatively clean air may not have negative effects on the human body, but if inhaled polluted air for a long time, there is a high probability of causing negative effects on the respiratory systems of the human body. Direct negative effects include the onset or exacerbation of long-term respiratory diseases such as Chronic Obstructive Pulmonary Disease (COPD) and asthma, or possibly dyspnea significantly fatal to an individual's life. For example, the Great Smog of London. In 1952, due to the popularity of industrial civilization in the city of London, a large number of harmful gaseous substances had been emitted from vehicles and factories, resulting in smog. People at that time did not understand the dangers of air pollution and how to deal with or avoid it, which caused many deaths. To avoid similar phenomena, intensive monitoring and solution provision on the vulnerable regions may be necessary . Whether it is to prevent the air from being polluted or to properly deal with it when the air is polluted. Therefore, knowing how different harmful gaseous pollutants are produced enables preventing such pollutions by the drastic increment of the average concentration, knowing the acceptable upper limit of each air pollution enables managing the exceedance of the air pollutant concentration standards, also knowing the damage each air pollution enables rapid, focused, and prioritised treatment of citizens residing in the pollutant-affected at-risk areas.

This article will focus on Utah and its neighbour in 2019, with three types of air pollution: Nitrogen dioxide, ozone and $PM_{2.5}$. All three types of pollutants are well-known substances that fatally damage human respiratory systems. Nitrogen dioxide is the product of industrial synthesis of nitric acid, mainly from chemical factories. Ozone exceeding the standard is mainly due to the reaction of exhaust gas emitted by factories and vehicles under sunlight and high temperature. The excessive $PM_{2.5}$ concentration is mainly caused by wood burning either by wildfires or logging.

In the following analysis, this research will focus on the pollution indicators of each region and the pollution level of each time period, so as to analyze the risk level of each region and the trend of air pollution in the future. This may help in deciding how to prevent or deal with air pollution in different regions.

2. Methods

2.1 Study Area

The state of Utah was chosen as the study area of the research, located in the western region of the United States, characterised by the Rocky Mountains that traverse the area from north to south, as well as numerous lakes and forests, known for recreational opportunities, but also is subject to variety of air pollution sources, mostly due to the state being alpine region and its basin topography, has fatally shortened the majority of the residents' life expectancy. Beside Salt Lake City, the state capital of Utah, the rest of the state consists of mountainous regions. The economy of the state is led by Salt Lake City's service industry primarily, but also the electronics manufacturing industry in the northern region of the state is active. In terms of socioeconomic status of the residents, Utah is known to have one of the lowest unemployment rates among all the states in the U.S., low crime rate, high education levels, and multilinguality acquired by the frequent oversea missionary work experience of the religion-based community in Utah, taking one third proportion of the resident population.

The state of Arizona is located in the western region of the United States and bordered by Wyoming, Nebraska, Kansas, Utah, Oklahoma and New Mexico. It is the sunniest area in the world and has very low precipitation. The state is known for deserted landscapes and cacti. It is also famous for scenery like the Grand Canyon. The economy is driven by tourism, mining and manufacturing. Arizona has a high Native American population and a strong Hispanic heritage.

The state of Colorado is located in the western region of the United States and bordered by Wyoming, Nebraska, Kansas, Utah, Oklahoma and New Mexico. The rocky mountain penetrates through the state. It is indeed the highest mountainous region in the United States; the state capital, Denver, is called Mile High City. It has extreme weather as it snows for half a year and because of the geographical characteristic, it is home to some famous slopes and ski ranges. The state is also full of energy resources, such as foil fuel, natural gas, etc.

The state of Wyoming is located in the western region of the United States and bordered by Montana, South Dakota, Nebraska, Utah and Colorado. The state is known for the vastly opened land, famous for national parks like Yellowstone National Park. Wyoming is the least populated state in the country excluding Alaska. They rely heavily on mineral extraction especially coal as they are in charge of 40% of total coal production in the United States. Real median income is fairly amongst the other states because the housing cost is one of the lowest in the country. In addition, due to the fact the home ownership ratio is quite high.

The state of Nevada is also located in the western region of the United States and bordered by Oregon, Idaho, Utah and Arizona. Rapid increase in population in recent years, the state is known for deserted landscape, entertainment, etc. Geographically, the state is mainly high mountain areas or desert. Due to the fact, it was hard for agriculture industries to develop, instead the development was centred on livestock and mining. The largest city, Las Vegas is home of resorts, casinos, entertainments and is a major tourist destination.

Refer to Figure 1, the map of five states of the study area, divided in census tract level for better display of the whereabouts of mountainous regions and main urbanised regions of the state.

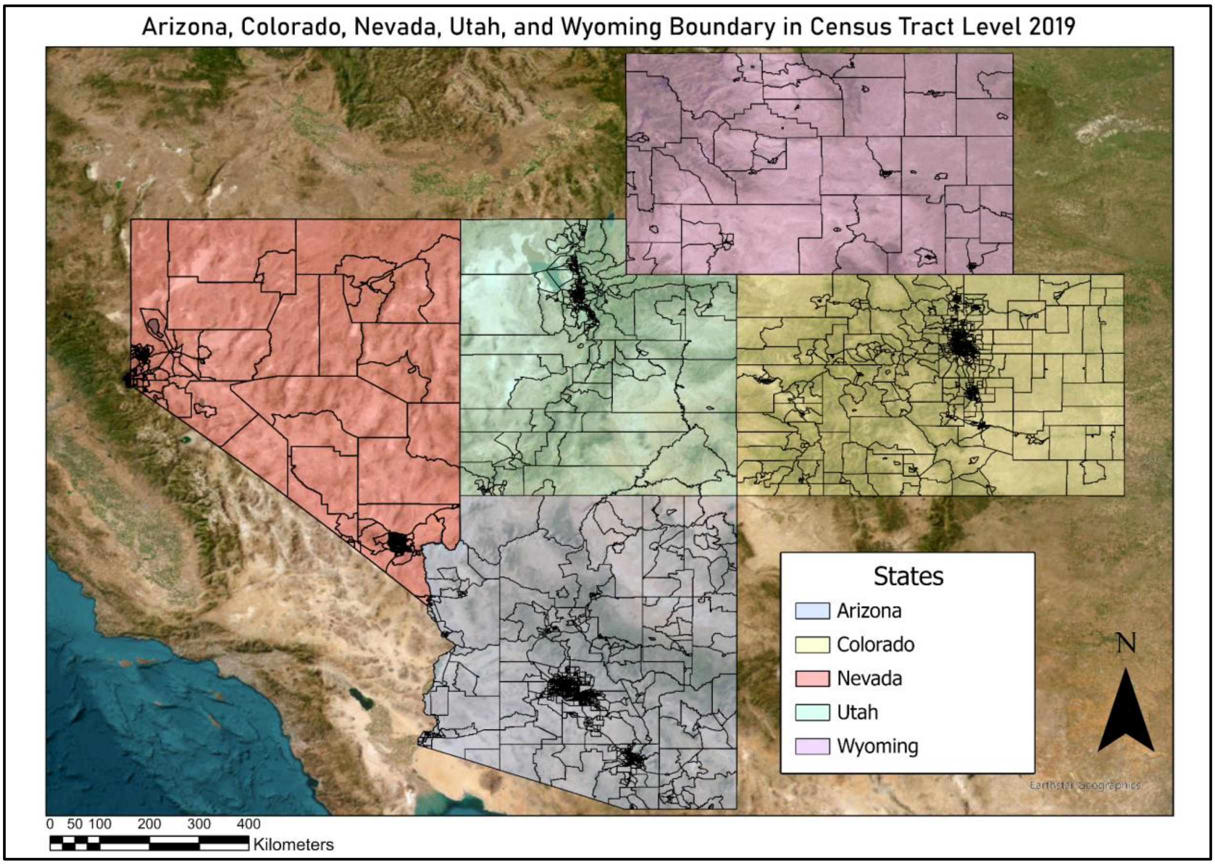

Figure 1. Arizona, Colorado, Nevada, Utah, and Wyoming Boundary in Census Tract Level (2019)

2.2 Data

In order to understand which types of air pollution are highly concentrated in specific regions, this article uses the data from United States Environmental Protection Agency (US EPA) , AirData Concentration by Monitor 2019, representing the annual average air pollutant concentrations in the U.S in 2019,. The Utah State Boundary shapefile by Utah Automated Geographic Reference Center (AGRC) ArcGIS Hub, DOR - Wyoming State Boundary shapefile by Wyoming Department of Revenue ArcGIS Hub, USGS National Boundary Dataset (NBD) in Arizona State or Territory shapefile by United States Geological Survey (USGS), Colorado State Boundary shapefile by Colorado Department of Public Health and Environment, and Nevada State Boundary shapefile by US EPA represents the combined area of research.

2.3 Statistical Analysis

2.3.1 IDW Interpolation

The Inverse Distance Weight method in R is used to estimate the unobserved hence unknown air quality index for each air pollutant of interest over the study area, in this case, the combined boundary of the state of Utah, Wyoming, Arizona, Nevada, and Colorado. The IDW method takes a model power parameter k as the exponent of decay regards to the distance of consideration. If k is set to 1, then the weight given to each point is simply the inverse of the distance between the known point and the unknown point. If k is set to a higher value, then the weights of the surrounding points are further increased, resulting in a more localised interpolation. Conversely, if k is set to a lower value, the weights of the surrounding points are decreased, resulting in a more generalised interpolation.

2.3.2 Kriging

Kriging method is also used to estimate the air pollutant concentration values at unobserved locations within the defined state boundaries of choice, however, in this method, the spatial dependence among data points are modelled in a form of variogram in which provides the visualization of the Kriging assumption of the data following stationarity, meaning the statistical properties of the data do not vary with location. The variogram enables the estimation of covariance between combinations of data points and uses this information to generate a spatially continuous estimate of unknown values as mentioned above. The general parameters that most of the variogram formulations defined are: s for Sill, the value at which the model flattens out, r for Range, the distance at which model first flattens out, and n for Nugget, the value at which the semi-variogram almost intercepts the y-axis. The semi-variance from the empirical values has to be observed to select a proper variogram model. By observing the semi-variance of the empirical data, a model of similar shape which would best represent the relationship between distance and semivariance, can be selected to fit the model to empirical data. For the optimization of parameters, the parameters that minimise the Sum of Squared Errors (SSE) are selected.

2.3.3 Clustering

As one of the low computational cost clustering methods and comparably low complexity of the variables of the air pollutant concentration data, k-means clustering method in GeoDa, an open-source software package for explanatory spatial data analysis, is chosen to identify areas that contain monitors that are most similar in air quality regards to each air pollutant of interest. In this case, the three to four variables from each data set are taken into account for the clustering: latitude, longitude, arithmetic, and optionally state code. The latitude and longitude defines the geographical location of each station, the arithmetic is the level of concentration for each pollutant, and state code is another variable that considers geopolitical partitioning, only if the clustering with the first three variables seems not performing well geographically.

3. Results

3.1 Descriptive Statistics

3.1.1 Key Results

- Nitrogen dioxide has the highest mean and maximum values
- Ozone has the lowest mean and maximum values
- Ozone has the least variability of its data distribution
- $PM_{2.5}$ concentrations are left-skewed but very close to normal distribution.
- Ozone and Nitrogen dioxide concentration are right-skewed distribution (Mean > Median)

Visualising these data distributions provides preparation for the normality assumption checking and possible decision on log-transformation of the data for Ordinary Kriging in our next steps (see Table 1 and Figure 2).

3.1.2 Figures, Tables and Schemes

Table 1. Summary Statistics of Air Pollutants of Interest Data: Ozone, Nitrogen dioxide, and $PM_{2.5}$

Pollutant	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
Ozone	0.03151	0.04806	0.05044	0.05029	0.05257	0.05750
Nitrogen dioxide	0.821	5.505	11.434	15.788	27.907	44.626
$PM_{2.5}$	0.6117	3.9879	5.3410	5.4596	6.8787	11.370

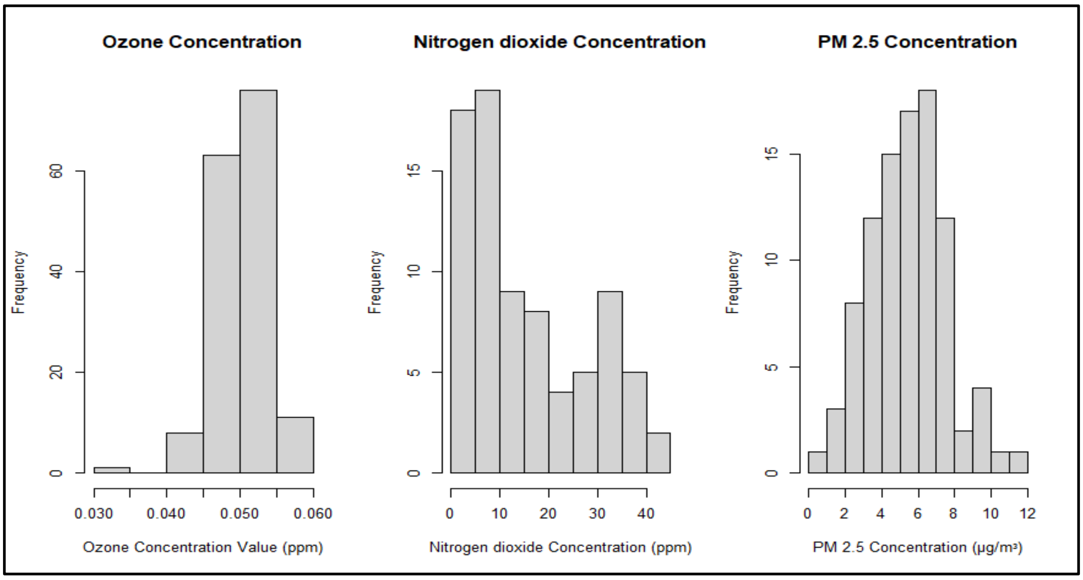

Figure 2. Histogram of Air Pollutant of Interest Values: Ozone, Nitrogen dioxide, and $PM_{2.5}$

3.2 IDW Interpolation Results

3.2.1 Key Results

- Each air pollutant data undergoes iterative LOOCV for optimal parameter selection
- *k* values referring to minimal RMSE are chosen as the final IDW model parameter
- 40,000 random sample points predicted and observed values are displayed

Figure 3 indicates the 100 sample points of prediction locations over the merge states boundary that are randomly generated with the choice of the sample size. Starting from the initial power parameter is set to $k = 2$, known as the default parameter set of the rule of thumb, a sample of locations are visualised in a way that each predicted value falls under the defined interval of the concentration values (see Figure 4).

According to Figure 5, by observing the residual visualization, the IDW interpolation with parameter $k = 2$ is having a fairly high class of residual values, especially for Nitrogen dioxide concentration and $PM_{2.5}$ concentration. In order to have less error in model accuracy, in RMSE, the power parameter k is adjusted for minimal RMSE value. From $k = 0$ to $k = 5$ of +0.1 increment interval.

The parameter is evaluated based on Leave-One Out Cross Validation (LOOCV) of n-fold in which n is the number of observations, and plotted for the result RMSE and regarding k value that refers to the lowest RMSE value for each model is selected, respectively, having a trough for each RMSE trend regards to k value increment (see Figure 6). For each trough, an exact parameter value k and respective RMSE value is recorded to use as the final IDW model for each air pollutant: $k = 1$ for Ozone, $k = 1.3$ for Nitrogen dioxide and $PM_{2.5}$, however, the optimal parameter selection has not drastically reduced RMSE for IDW model of Nitrogen dioxide and $PM_{2.5}$ (see Table 2). This result is possibly indicating that spatial interpolation using IDW method would not be optimal for the two aforementioned air pollutants, if high accuracy criteria of RMSE value less than 1 is demanded.

40 000 location points are randomly generated within the boundary of the study, visualised on the map in a way that points of the prediction covers the region of interest which enables the representation of the area's predictive air pollutant concentration with some unavoidable loss of accuracy according to the respective RMSE value even after the parameter optimization (see Figure 7).

3.2.2 Figures, Tables and Schemes

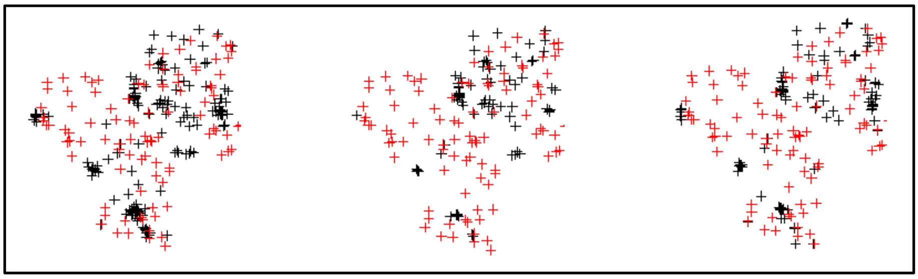

Figure 3. Random Sample Points of Prediction Locations (n = 100): Ozone, Nitrogen dioxide, and PM2.5

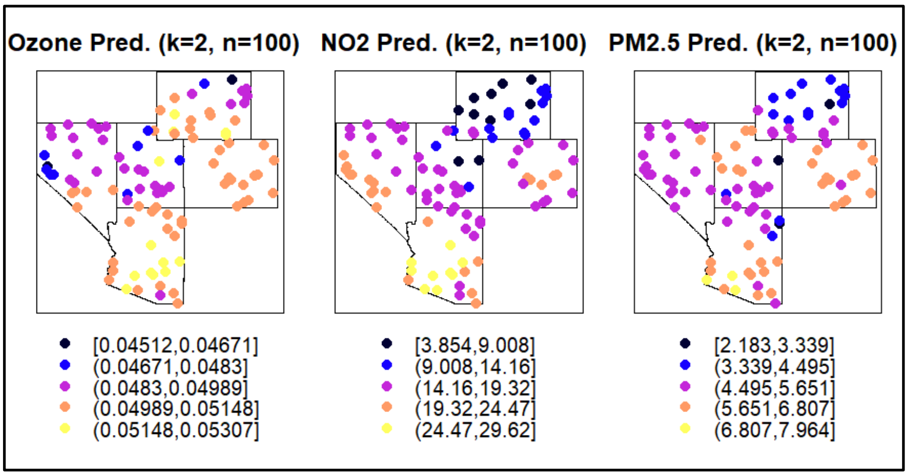

Figure 4. Inverse Distance Weighted Interpolation of Air Pollutant Concentrations (k = 2, n = 100): Ozone, Nitrogen dioxide, and PM2.5

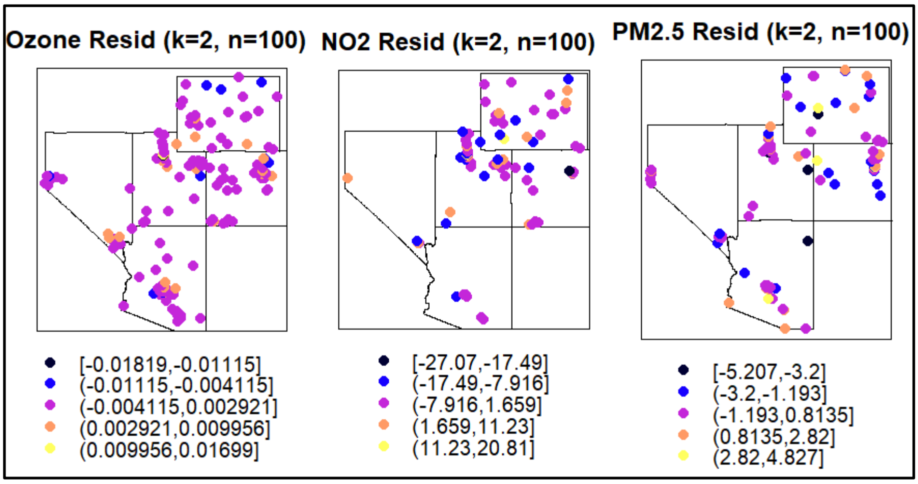

Figure 5. Residuals of Inverse Distance Weighted Interpolation of Air Pollutant Concentrations Predictions (k = 2, n = 100): Ozone, Nitrogen dioxide, and PM2.5

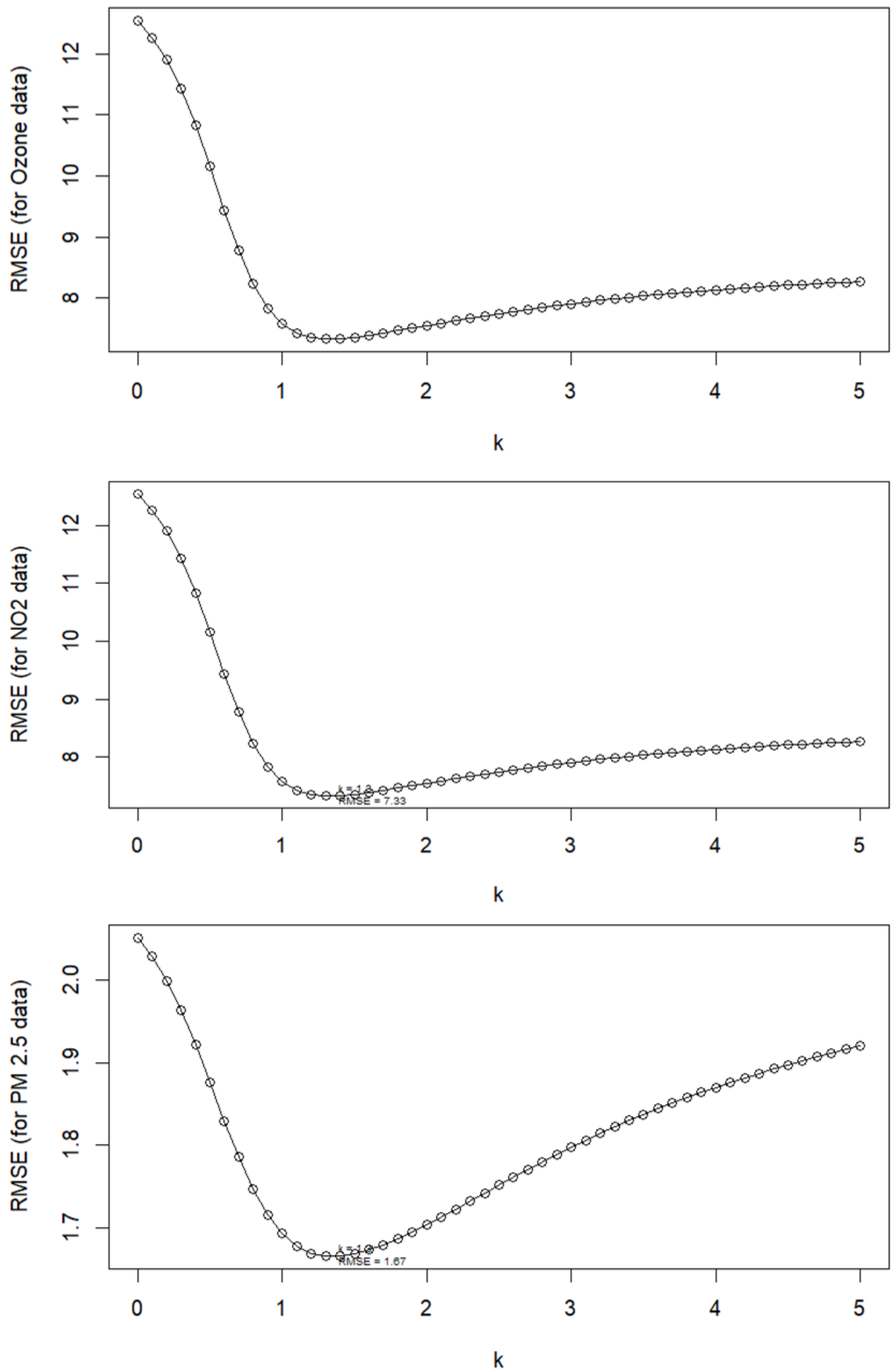

Figure 6. RMSE vs of IDW Model Power Parameter *k* from Iterative LOOCV
for Air Pollutant of Interest Concentration Estimation: Ozone, Nitrogen dioxide, and $PM_{2.5}$

Table 2. Optimal IDW Model Parameter *k* minimising RMSE of Prediction Residuals

Pollutant of Prediction	Optimal IDW Parameter (*k*)	RMSE
Ozone	1	0.003045932
Nitrogen dioxide	1.3	7.327283
$PM_{2.5}$	1.3	1.665076

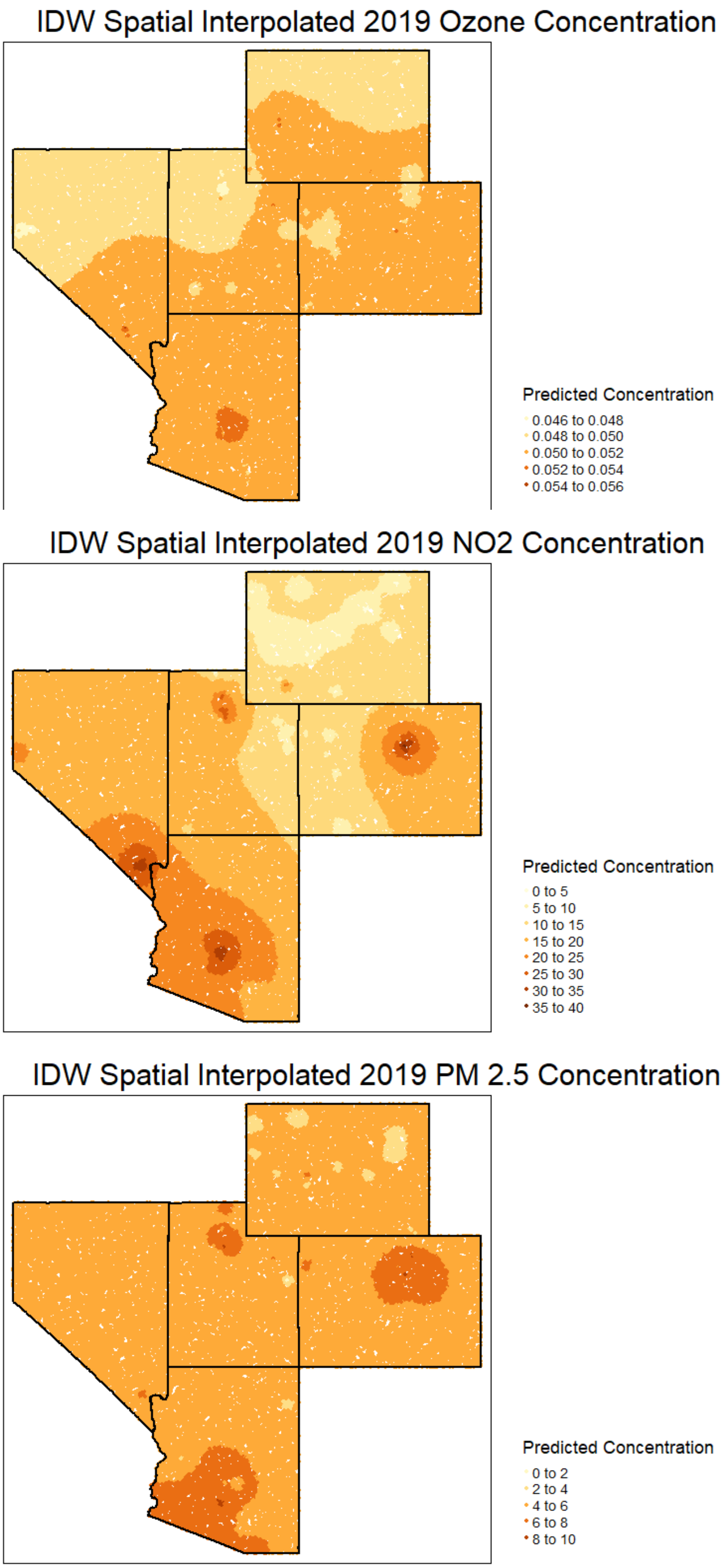

Figure 7. Inverse Distance Weighted Interpolation of Air Pollutant Concentration of Interest: Ozone, Nitrogen dioxide, and $PM_{2.5}$ (n = 40,000)

3.3 Kriging Results

3.3.1 Key Results

- Each air pollutant data had specific variogram model to be fitted
- The comparative model fitting performance was Ozone > NO2 > $PM_{2.5}$ based on the variogram model fitted
- All three Kriging interpolation models have much better predictive performance compared to IDW interpolation models in terms of RMSE values acquired from LOOCV.

As introduced on Section 3.1.2 (see Figure 2), due to skewness of the distribution of Ozone concentration data and Nitrogen dioxide concentration data, the two data are log-transformed in a way that the distribution is log-normal to pass the normality assumption of Kriging interpolation (see Figure 8). A few extreme outliers existed, however, were not deleted to avoid any bias in the data and also because the small number of outliers would not significantly control over the prediction results. By observing the pattern of empirical concentration value by each air pollutant (see Figure 10), some specific behaviours sharing in common with the model variograms of Figure 9: the "Hol" (Hole), "Gau" (Gausiaan), and "Wav" (Wave) model variogram for Ozone, Nitrogen dioxide, and $PM_{2.5}$, respectively. This can be validated by comparing with other candidate model variogram types in terms of how the model variogram line fits the empirical data points (see Figure 11) hence the model type for Kriging variogram is chosen as it was mentioned.

Table 3 presents the model parameters *pSill*, *Range*, and *Type* for model variogram, with the prediction accuracy based on RMSE value and goodness-of-fit based on SSError. In this case, the model fitting performance was the highest in Ozone variogram and the lowest in $PM_{2.5}$ variogram. To compare the prediction accuracy, all three Kriging interpolation models are found having much less RMSE than IDW interpolation models. Based on the air pollutant concentration on a two-dimensional grid of longitude and latitude with each of the locations containing a concentration value, the spatial polygon of convex hull and grid of prediction points are generated which assigns the predictive value on a grid cell the value belongs to spatially (see Figure 12).

Finally, Kriging interpolated air pollutant concentrations for each pollutant are created as a surface of cell size equal to 0.1, then displayed beneath the collection of state boundaries of the study areas for more comprehensive interpretation of high-valued areas of real-life locations. Comparing the common areas of highest concentration estimated air pollutants, it was found that South to Southwest end of Arizona: the central urban area of Arizona, East of Colorado which is a combination of mountainous region and also the central urban area of Colorado, and the upper centre of Utah: the combination of urban areas. Wyoming, on the other hand, had fairly moderate to low concentration over all three air pollutant categories. The approximate locations of the urban areas can be found by observing the regions with comparatively small sizes of census tracts clustered within the state boundary (see Figure 1).

3.3.2 Figures, Tables and Schemes

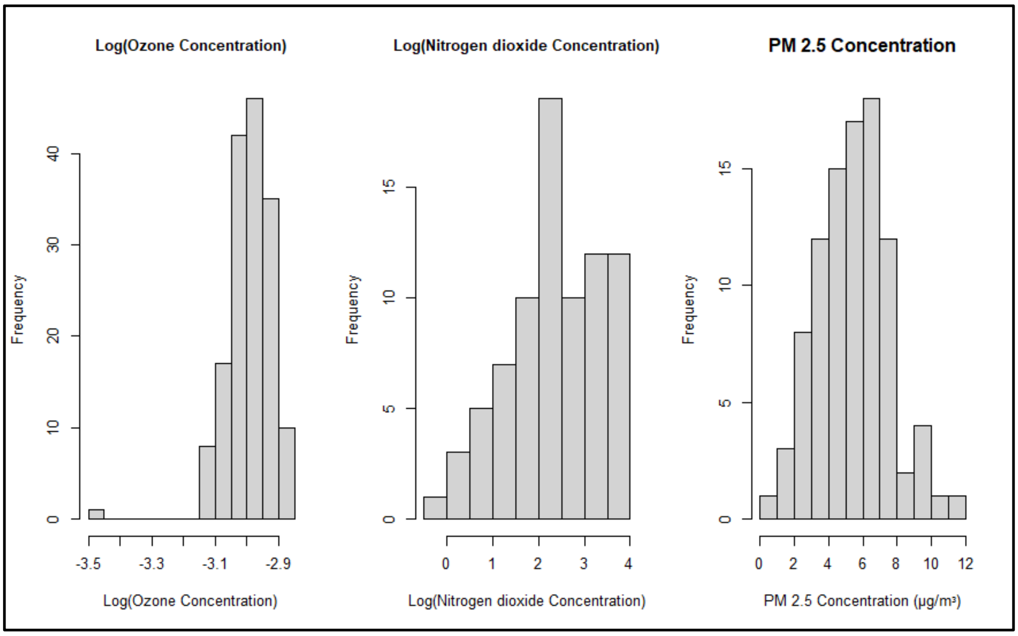

Figure 8. Histogram of Necessary Log-transformation of Air Pollutant Concentration Data: Log(Ozone), Log(Nitrogen dioxide), and Original $PM_{2.5}$

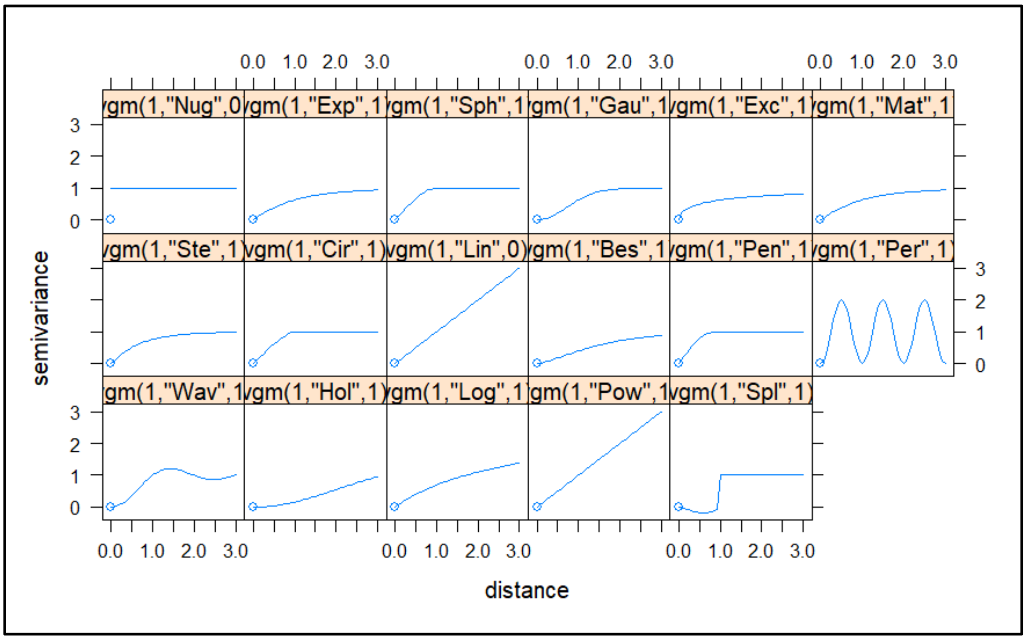

Figure 9. List of Plot Variogram Models for Model Type Decision

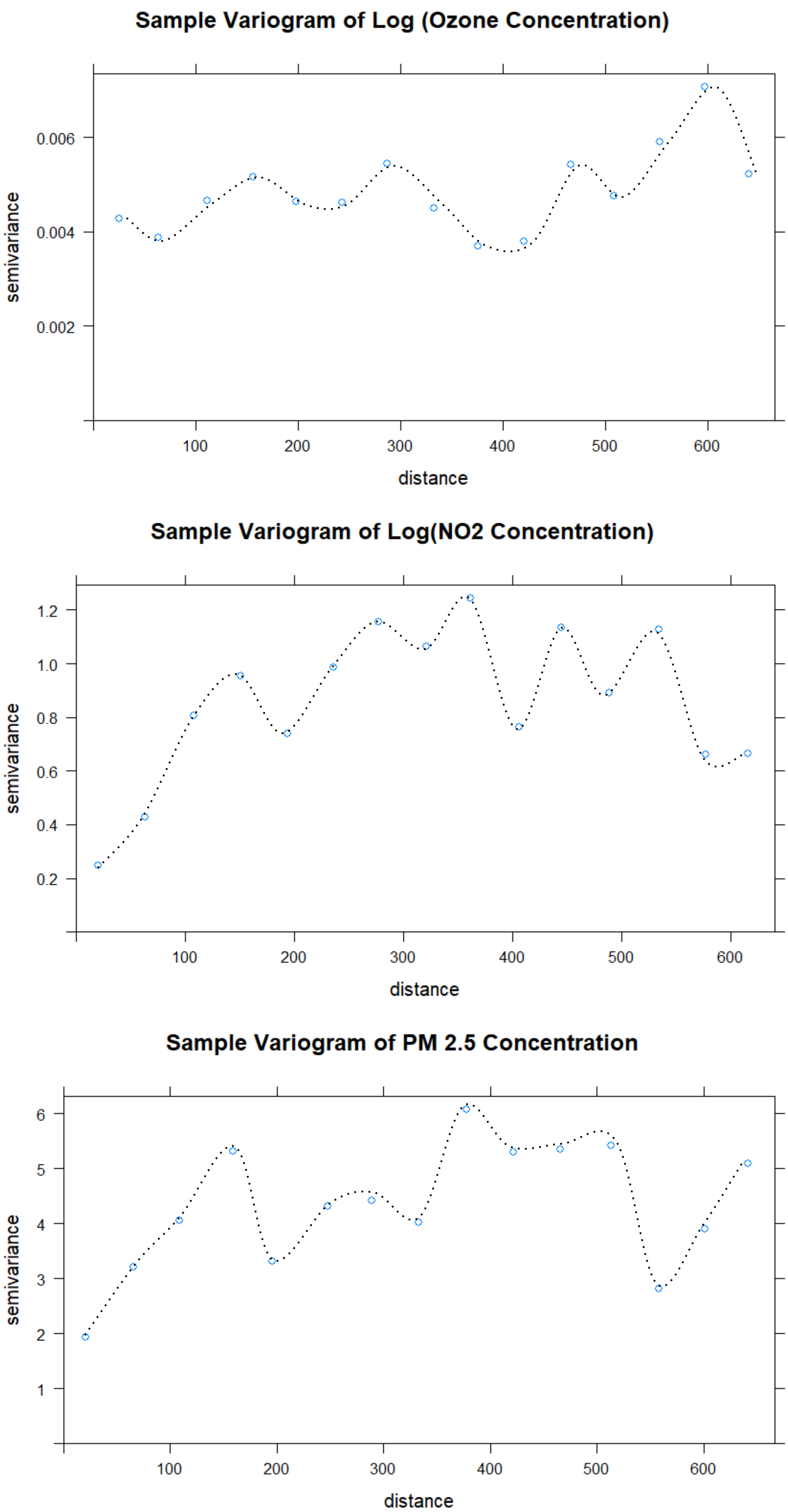

Figure 10. Sample Variogram of Air Pollutant Concentration: Log(Ozone), Log(Nitrogen dioxide), and $PM_{2.5}$

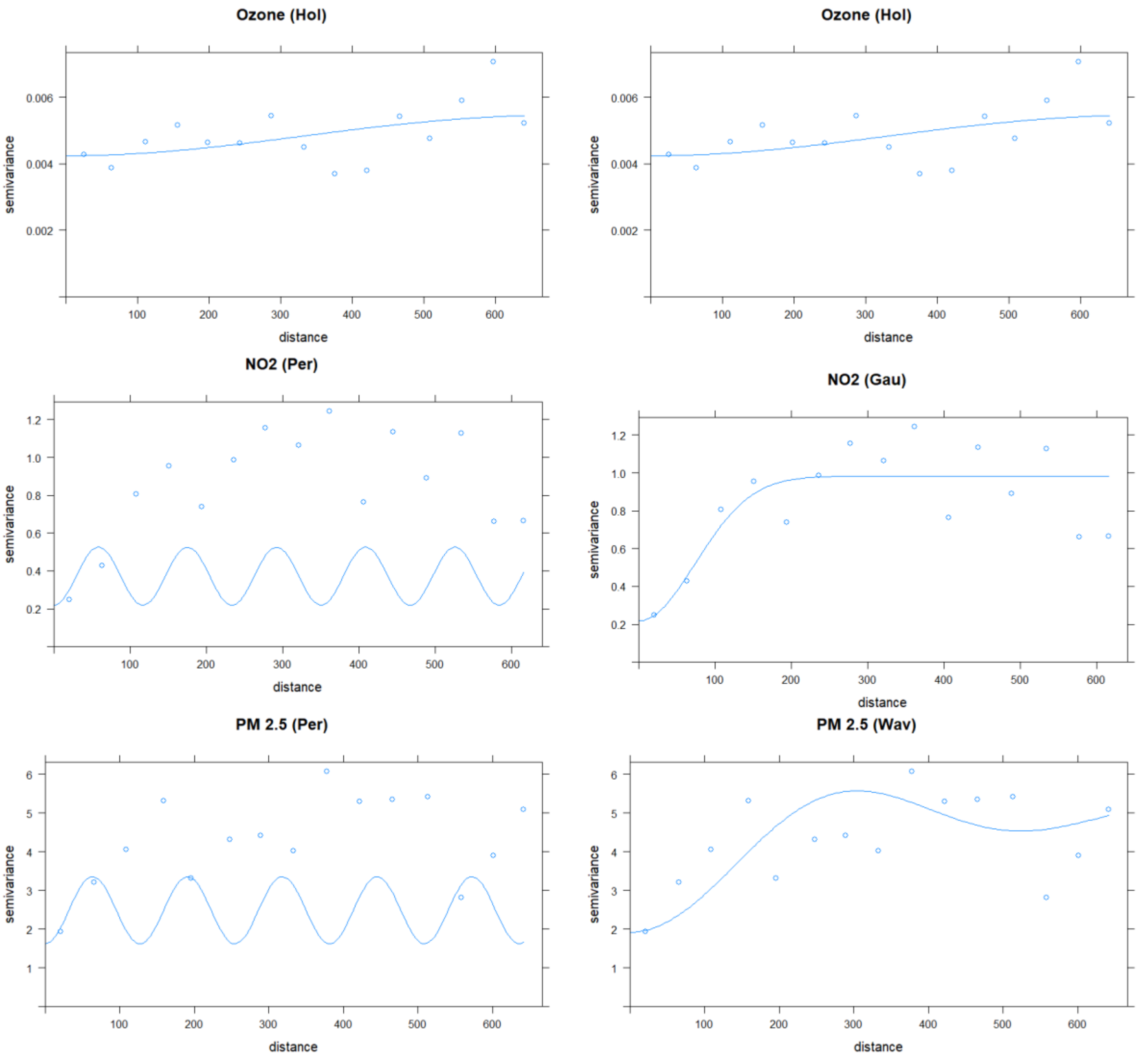

Figure 11. Candidate Variogram Model Selection by the Trend of Fitting Empirical Air Pollutant Concentration

Table 3. Selected Kriging Variogram Model with Parameters and Model Evaluation LOOCV Results

Pollutant	pSill	Range	Model	RMSE	SSError
Ozone	0.001009458	158.1515	Hole	0.06515891	0.00000004553801
Nitrogen dioxide	0.7650246	104.2188	Gaussian	0.7425717	0.0006423967
$PM_{2.5}$	3.000621	213.0957	Wave	1.789184	0.04930431

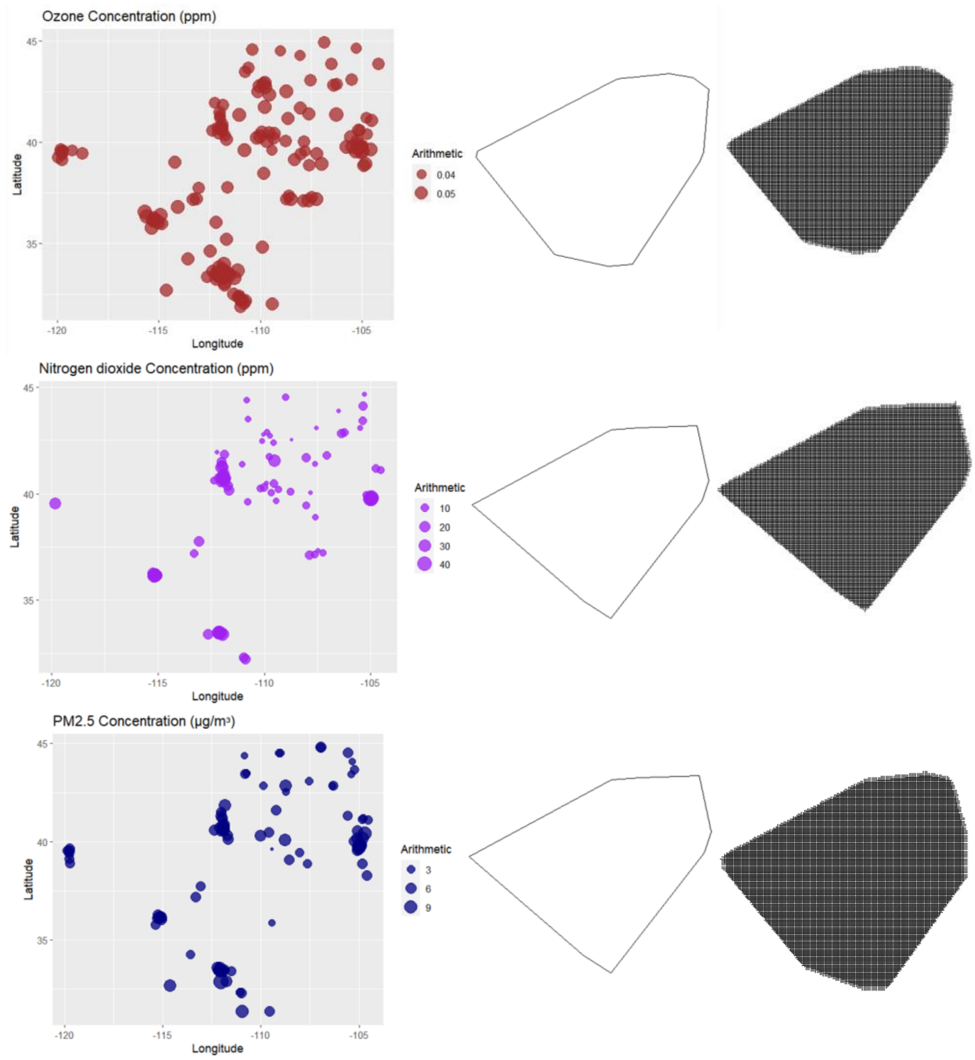

Figure 12. Convex Hull and Grid Formation of Air Pollutant Spatial Data

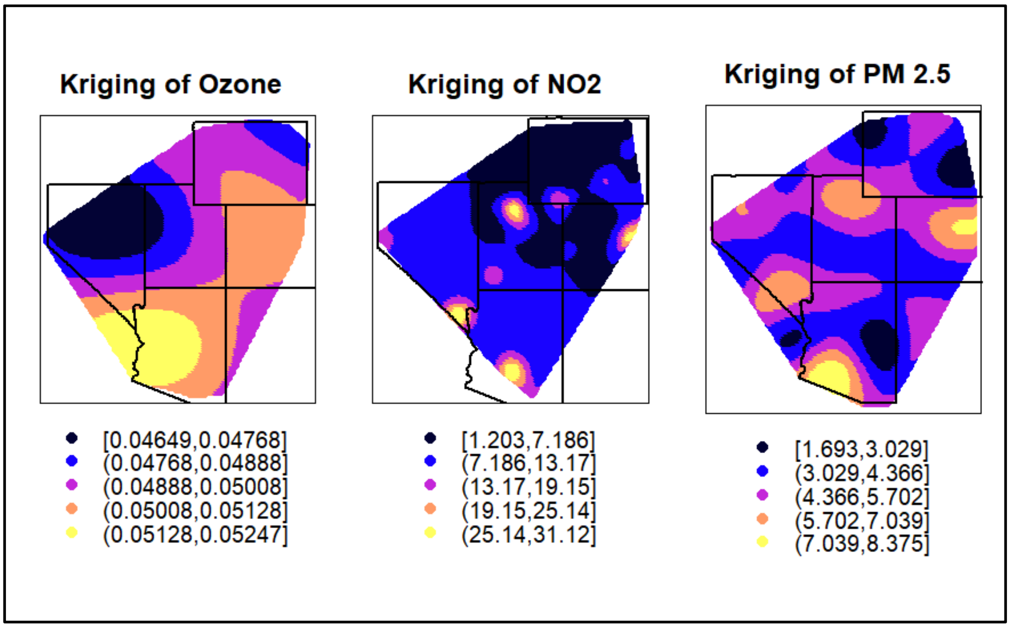

Figure 13. Kriging Interpolated (Cell Size = 0.1) over Convex Hull Grid for Air Pollutant Concentration of Interest: Ozone, Nitrogen dioxide, and $PM_{2.5}$ with Merge States Boundary overlay

3.4 Clustering Results

3.4.1 Key Results

- For the Ozone Concentration map, there are three main clusters which are clustered by Longitude and Latitude.
  Subclusters within the main clusters contain high concentration values.
- For the Nitrogen Dioxide Concentration map, seven clusters were generated.
- For $PM_{2.5}$ Concentration map, four major clusters were generated and two of them have Sub-clusters which contain high concentration values.

This article used the K-means clustering method to generate the results, however, due to the emergence of sub-clusters, the hierarchical clustering method could be used (see Figure 14-16).

3.4.2 Figures, Tables and Schemes

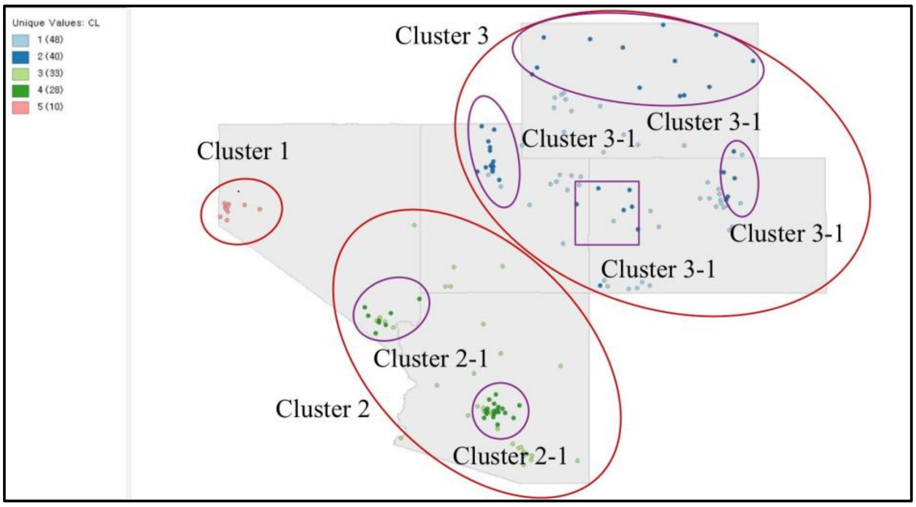

Figure 14. *k*-means Clustered (k = 5) Ozone Concentration Data
Variables of Consideration: Latitude, Longitude, Ozone Concentration

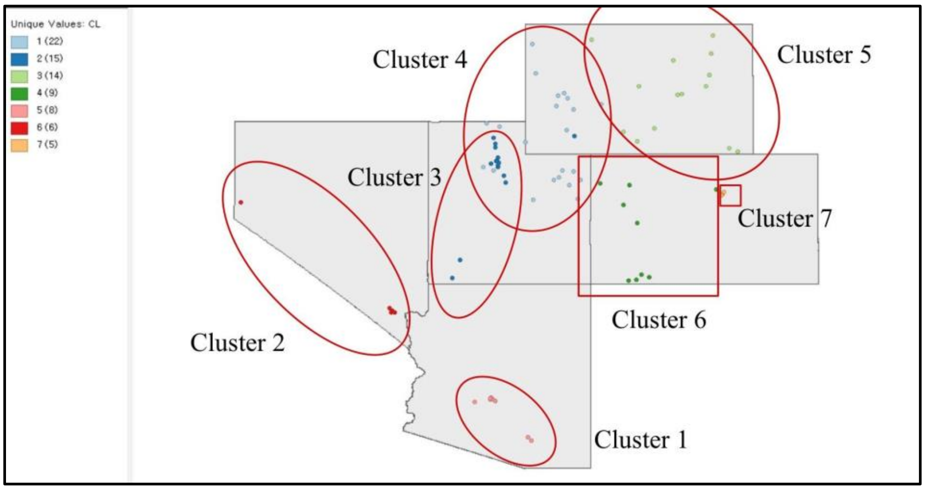

Figure 15. *k*-means Clustered (k = 7) Nitrogen dioxide Concentration Data
Variables of Consideration: Latitude, Longitude, Ozone Concentration, State Code

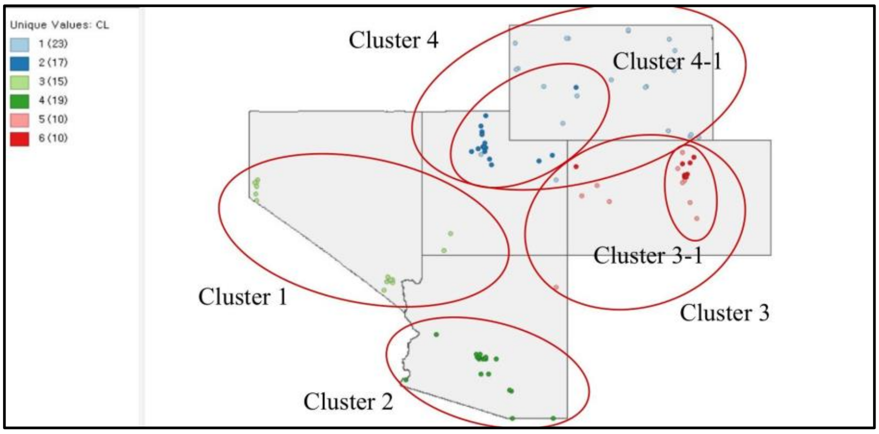

Figure 16. *k*-means Clustered (k = 6) $PM_{2.5}$ Concentration Data
Variables of Consideration: Latitude, Longitude, Ozone Concentration, State Code

4. Discussion

Pollution from ozone affects the environment the least, whereas $PM_{2.5}$ affects the most from the given results. Compared our predicted results to data collected in Ecuador , we can see that all three pollutants have significantly smaller concentrations than the Borja-Urbano research area [1]. As for the result in Mexico cities, ozone shows relatively high ozone concentration compared with Nitrogen dioxide concentration which is different from the result given in this article. Average concentrations of nitrogen dioxide and ozone are larger than the given result in this article because Mexico is a developing country and America is a developed country which means that they utilise more energy, so its pollution is high [2]. In Madrid, similarly, ozone concentration is also higher than nitrogen dioxide concentration [3]. Data collected in Mexico cities, Ecuador and in this article all represent that $PM_{2.5}$ polluted the environment the most.

Therefore, the increment of $PM_{2.5}$ should be monitored and concerned more intensively. The best way to alleviate $PM_{2.5}$ should be less emission produced by chemical plants and replacing gasoline-based cars into electric cars more. One point to notice is that data collected in Madrid and Mexico cities are long-term, both above 10 years. Other potential influencing factors are wind speed and season interchange in which could be an underlying factor of disturbance that we could not catch the hidden highly pollutant concentrated areas. Hence, improvement should be made for this analysis for future data collection. One adjustment should be the timeline, at least five years of data collecting is necessary to make the data more stable and precise.

5. Conclusions

In conclusion, this study utilised Spatial interpolation and spatial clustering methods to analyze air pollution patterns across five contiguous US states. Our findings offer valuable insights for environmental conservation and policymaking etc. It also provides a better insight of air pollutants and its spatial pattern. However, there are two potential flaws that could be improved in our project:We could've used co-kriging instead of ordinary kriging to obtain smaller errors. In addition, for spatial clustering, hierarchical clustering could have been considered as an alternative approach over k-means clustering in our case, analyzing the underlying factors of spatial process undiscovered such as socioeconomic variables of the residential areas, industrial zoning densities, loss of vegetation mass, and topography or climatology of the region considering the altitude, temperature, air pressure and precipitation levels. Also for spatial interpolation, the Co-Kriging method could be considered for further studies which takes into account multiple explanatory variables in a relationship with the variable to be predicted

References

1. Borja-Urbano, S., Rodríguez-Espinosa, F. L., Luna-Ludeña, M., & Toulkeridis, T. (2021b). Valuing the Impact of Air Pollution in Urban Residence Using Hedonic Pricing and Geospatial Analysis, Evidence From Quito, Ecuador. *Air, Soil and Water Research*, *14*, 117862212110532. https://doi.org/10.1177/11786221211053277

2. García-Burgos, J., Miquelajauregui, Y., Vega, E., Namdeo, A., Ruíz-Olivares, A., Mejía-Aranguré, J. M., Resendiz-Martinez, C. G., Hayes, L., Bramwell, L., Jaimes-Palomera, M., Entwistle, J., Núñez-Enríquez, J. C., Portas, A., & McNally, R. J. (2022). Exploring the Spatial Distribution of Air Pollution and Its Association with Socioeconomic Status Indicators in Mexico City. Sustainability, 14(22), 15320. https://doi.org/10.3390/su142215320

3. Gómez-Losada, Á., Santos, F. M., Gibert, K., & Pires, J. C. (2019). A data science approach for spatiotemporal modelling of low and resident air pollution in Madrid (Spain): Implications for epidemiological studies. Computers, Environment and Urban Systems, 75, 1–11. https://doi.org/10.1016/j.compenvurbsys.2018.12.005

Appendix A: R Code Used for the Analysis

```
# Loading Libraries
library(tidyverse)
library(lubridate)
library(tidyr)
library(sf)
library(rgdal)
library(sp)
library(gstat)
library(raster)
library(ggplot2)
library(leaflet)
library(tmap)
library(geojsonio)
library(geojsonlint)
library(lattice)
library(gridExtra)

# Reading AQI Data GeoJSON
AQI <- readOGR("Data/2019AnnualConc/2019AnnualConc.geojson")
States <- readOGR("Data/StateMerged/StateMerged.geojson")

# Filtering and Subsetting by Variable with Duplicate Removals
categories_to_keep <- c("Ozone", "Nitrogen dioxide (NO2)", "PM2.5 - Local
Conditions")
AQI_filtered <- AQI[AQI$Parameter1 %in% categories_to_keep,]

ozone_data <- AQI_filtered[AQI_filtered$Parameter1 == "Ozone",]
no2_data <- AQI_filtered[AQI_filtered$Parameter1 == "Nitrogen dioxide
(NO2)",]
pm25_data <- AQI_filtered[AQI_filtered$Parameter1 == "PM2.5 - Local
Conditions",]

ozone_data_unique <- ozone_data[!duplicated(ozone_data@data[c("Latitude",
"Longitude")]),]
no2_data_unique <- no2_data[!duplicated(no2_data@data[c("Latitude",
"Longitude")]),]
pm25_data_unique <- pm25_data[!duplicated(pm25_data@data[c("Latitude",
"Longitude")]),]

# Ozone Concentration Summary Statistics
ozone_summary <- ozone_data_unique@data$Arithmetic %>%
  summary()
# NO2 Concentration Summary Statistics
no2_summary <- no2_data_unique@data$Arithmetic %>%
  summary()
# PM 2.5 Concentration Summary Statistics
pm25_summary <- pm25_data_unique@data$Arithmetic %>%
  summary()

ozone_summary; no2_summary; pm25_summary

# Histograms of Air Pollutant Concentrations
par(mfrow=c(1,3))
hist(ozone_data_unique@data$Arithmetic, main="Ozone Concentration",
     xlab = "Ozone Concentration Value (ppm)")
hist(no2_data_unique@data$Arithmetic, main="Nitrogen dioxide Concentration",
     xlab = "Nitrogen dioxide Concentration (ppm)")
hist(pm25_data_unique@data$Arithmetic, main="PM 2.5 Concentration",
```

```
    xlab = "PM 2.5 Concentration (µg/m³)")


# Checking Projections
proj4string(ozone_data_unique)
proj4string(no2_data_unique)
proj4string(pm25_data_unique)

# Counting Measure Stations
ozone_station_count <- ozone_data_unique@data %>%
  distinct(Latitude, Longitude, .keep_all = TRUE) %>%
  group_by(Latitude, Longitude) %>%
  summarise(num_stations = n())

nrow(ozone_station_count) #159 stations

no2_station_count <- no2_data_unique@data %>%
  distinct(Latitude, Longitude, .keep_all = TRUE) %>%
  group_by(Latitude, Longitude) %>%
  summarise(num_stations = n())

nrow(no2_station_count) #79 stations

pm25_station_count <- pm25_data_unique@data %>%
  distinct(Latitude, Longitude, .keep_all = TRUE) %>%
  group_by(Latitude, Longitude) %>%
  summarise(num_stations = n())

nrow(pm25_station_count) #94 stations

# IDW Spatial Interpolation - Location Plot with Points to be Predicted
n_points <- 100
predictions <- spsample(States, n_points, "random")

par(mfrow=c(1,3))
plot(ozone_data_unique)
plot(predictions, add=TRUE, col = "RED")
plot(no2_data_unique)
plot(predictions, add=TRUE, col = "RED")
plot(pm25_data_unique)
plot(predictions, add=TRUE, col = "RED")

# IDW Spatial Interpolation - IDW Predicted Plots
Ozonepred <- idw(Arithmetic ~ 1,
            ozone_data_unique,
            predictions,
            idp = 2)

NO2pred <- idw(Arithmetic ~ 1,
            no2_data_unique,
            predictions,
            idp = 2)

PM25pred <- idw(Arithmetic ~ 1,
            pm25_data_unique,
            predictions,
            idp = 2)

par(mfrow=c(1,3))
```

```r
SamplePlot1 <- spplot(Ozonepred, "var1.pred", main = "Ozone Pred. (k=2, 
n=100)",
                      sp.layout = list(polygons(States)))
SamplePlot2 <- spplot(NO2pred, "var1.pred", main = "NO2 Pred. (k=2, n=100)",
                      sp.layout = list(polygons(States)))
SamplePlot3 <- spplot(PM25pred, "var1.pred", main = "PM2.5 Pred. (k=2, 
n=100)",
                      sp.layout = list(polygons(States)))

grid.arrange(SamplePlot1, SamplePlot2, SamplePlot3, ncol = 3)

# Demo of Sample IDW Cross-Validation
LOOCV_Ozone <- krige.cv(Arithmetic ~ 1,
                        ozone_data_unique,
                        nfold = nrow(ozone_data_unique),
                        set = list(idp = 2))

LOOCV_Ozone@data

LOOCV_NO2 <- krige.cv(Arithmetic ~ 1,
                      no2_data_unique,
                      nfold = nrow(no2_data_unique),
                      set = list(idp = 2))

LOOCV_NO2@data

LOOCV_PM25 <- krige.cv(Arithmetic ~ 1,
                       pm25_data_unique,
                       nfold = nrow(pm25_data_unique),
                       set = list(idp = 2))

LOOCV_PM25@data

# LOOCV Residual Plotting
loocvOzonePlot <- spplot(LOOCV_Ozone, "residual", main="Ozone Resid (k=2, 
n=100)",
                         sp.layout = list(polygons(States)))
loocvNO2Plot <- spplot(LOOCV_NO2, "residual", main="NO2 Resid (k=2, n=100)",
                       sp.layout = list(polygons(States)))
loocvPM25Plot <- spplot(LOOCV_PM25, "residual", main="PM2.5 Resid (k=2, 
n=100)",
                        sp.layout = list(polygons(States)))
grid.arrange(loocvOzonePlot, loocvNO2Plot, loocvPM25Plot, ncol = 3)

# Measurement of Fit
# RMSE of Air Pollutant LOOCV

RMSE_resid <- function(x){
  return(sqrt(mean(x^2)))
}

RMSE_resid(LOOCV_Ozone@data$residual) # RMSE.ozone: 0.003207004
RMSE_resid(LOOCV_NO2@data$residual) # RMSE.no2: 7.546449
RMSE_resid(LOOCV_PM25@data$residual) # RMSE.pm25: 1.703405

# Generally, RMSE of Residuals lower than 1 indicates good agreement so
# the parameter k of NO2 has to be adjusted

# Refitting with new exponent parameter k for IDW (k=3)
```

```
LOOCV_NO2_k3 <- krige.cv(Arithmetic ~ 1,
                      no2_data_unique,
                      nfold = nrow(no2_data_unique),
                      set = list(idp = 3))

LOOCV_NO2_k3@data

LOOCV_PM25_k3 <- krige.cv(Arithmetic ~ 1,
                       pm25_data_unique,
                       nfold = nrow(pm25_data_unique),
                       set = list(idp = 3))

LOOCV_PM25_k3@data

RMSE_resid(LOOCV_NO2_k3@data$residual) # RMSE.no2: 7.902865 (increased RMSE)
RMSE_resid(LOOCV_PM25_k3@data$residual) # RMSE.pm25: 1.797648 (increased
RMSE)

# Refitting with new exponent parameter k for IDW (k=1.5)
LOOCV_NO2_k15 <- krige.cv(Arithmetic ~ 1,
                      no2_data_unique,
                      nfold = nrow(no2_data_unique),
                      set = list(idp = 1.5))

LOOCV_NO2_k15@data

LOOCV_PM25_k15 <- krige.cv(Arithmetic ~ 1,
                       pm25_data_unique,
                       nfold = nrow(pm25_data_unique),
                       set = list(idp = 1.5))

LOOCV_PM25_k15@data

RMSE_resid(LOOCV_NO2_k15@data$residual) # RMSE.no2: 7.356047 (decreased RMSE)
RMSE_resid(LOOCV_PM25_k15@data$residual) # RMSE.pm25: 1.668088 (decreased
RMSE)

# Refitting with new exponent parameter k for IDW (k=0.5)
LOOCV_NO2_k.5 <- krige.cv(Arithmetic ~ 1,
                      no2_data_unique,
                      nfold = nrow(no2_data_unique),
                      set = list(idp = 0.5))

LOOCV_NO2_k15@data

LOOCV_PM25_k.5 <- krige.cv(Arithmetic ~ 1,
                       pm25_data_unique,
                       nfold = nrow(pm25_data_unique),
                       set = list(idp = 0.5))

LOOCV_PM25_k.5@data

RMSE_resid(LOOCV_NO2_k.5@data$residual) # RMSE.no2: 10.15379 (increased RMSE)
RMSE_resid(LOOCV_PM25_k.5@data$residual) # RMSE.pm25: 1.876017 (increased
RMSE)


# Refitting with new exponent parameter k for IDW (k=1)
LOOCV_NO2_k1 <- krige.cv(Arithmetic ~ 1,
                      no2_data_unique,
```

```
                      nfold = nrow(no2_data_unique),
                      set = list(idp = 1))

LOOCV_NO2_k1@data

LOOCV_PM25_k1 <- krige.cv(Arithmetic ~ 1,
                       pm25_data_unique,
                       nfold = nrow(pm25_data_unique),
                       set = list(idp = 1))

LOOCV_PM25_k1@data

RMSE_resid(LOOCV_NO2_k1@data$residual) # RMSE.no2: 7.356047 (increased RMSE)
RMSE_resid(LOOCV_PM25_k1@data$residual) # RMSE.pm25: 1.668088 (increased
RMSE)


# Loop for parameter k

# Looping parameter k LOOCV for Ozone Concentration Data IDW
RMSE_values_ozone <- c()
k_values <- seq(0, 5, by = 0.1)
for (k in k_values) {
  RMSE_values_ozone <- c(RMSE_values_ozone, RMSE_resid(krige.cv(Arithmetic ~
1,
                                                    ozone_data_unique,
                                                    nfold =
nrow(ozone_data_unique),
                                                    set = list(idp =
k))@data$residual))
}
plot(k_values, RMSE_values_no2, type = "o", xlab = "k", ylab = "RMSE (for
Ozone data)")

min_RMSE_ozone <- min(RMSE_values_ozone)
min_k_ozone <- k_values[which.min(RMSE_values_ozone)]
text(min_k_ozone, min_RMSE_ozone,
     paste0("k = ", min_k_ozone, "\nRMSE = ", round(min_RMSE_ozone, 2)), pos
= 4, cex=0.5)

# Looping parameter k LOOCV for NO2 Concentration Data IDW
RMSE_values_no2 <- c()
k_values <- seq(0, 5, by = 0.1)
for (k in k_values) {
  RMSE_values_no2 <- c(RMSE_values_no2, RMSE_resid(krige.cv(Arithmetic ~ 1,
                                                    no2_data_unique,
                                                    nfold =
nrow(no2_data_unique),
                                                    set = list(idp =
k))@data$residual))
}
plot(k_values, RMSE_values_no2, type = "o", xlab = "k", ylab = "RMSE (for NO2
data)")

min_RMSE_no2 <- min(RMSE_values_no2)
min_k_no2 <- k_values[which.min(RMSE_values_no2)]
text(min_k_no2, min_RMSE_no2,
     paste0("k = ", min_k_no2, "\nRMSE = ", round(min_RMSE_no2, 2)), pos = 4,
cex=0.5)
```

```
# Looping parameter k LOOCV for PM2.5 Concentration Data IDW
RMSE_values_pm25 <- c()
k_values <- seq(0, 5, by = 0.1)
for (k in k_values) {
  RMSE_values_pm25 <- c(RMSE_values_pm25, RMSE_resid(krige.cv(Arithmetic ~ 1,
                                                    pm25_data_unique,
                                                    nfold =
nrow(pm25_data_unique),
                                                    set = list(idp =
k))@data$residual))
}
plot(k_values, RMSE_values_pm25, type = "o", xlab = "k", ylab = "RMSE (for PM
2.5 data)")

min_RMSE_pm25 <- min(RMSE_values_pm25)
min_k_pm25 <- k_values[which.min(RMSE_values_pm25)]
text(min_k_pm25, min_RMSE_pm25,
     paste0("k = ", min_k_pm25, "\nRMSE = ", round(min_RMSE_pm25, 2)), pos =
4, cex=0.5)

# IDW Model Parameter Selection for Each Pollutant

# Ozone IDW Model Parameter and RMSE
print(min_k_ozone)  # k = 1
print(min_RMSE_ozone) # RMSE: 0.003045932

# NO2 IDW Model Parameter and RMSE
print(min_k_no2) # k = 1.3
print(min_RMSE_no2) # RMSE: 7.326283

# PM2.5 IDW Model Parameter and RMSE
print(min_k_pm25) # k = 1.3
print(min_RMSE_pm25) # RMSE: 1.665076

# IDW Spatially Interpolated Raster Map for Each Pollutant
# Ozone dataset
data(ozone_data_unique)

# NO2 dataset
data(no2_data_unique)

# PM2.5 dataset
data(pm25_data_unique)

k_ozone <- 1
k_no2 <- 1.3
k_pm25 <- 1.3

# Coordinate Retrieval
ozone_coords <- ozone_data_unique[, c("Longitude", "Latitude")]
no2_coords <- no2_data_unique[, c("Longitude", "Latitude")]
pm25_coords <- pm25_data_unique[, c("Longitude", "Latitude")]

# Check for duplicated coordinates
duplicated(ozone_coords@data)
duplicated(no2_coords@data)
duplicated(pm25_coords@data)

# Creating random sample points and create a graduated colour dot map for IDW
```

```r
SampPoints <- spsample(x = States, n = 40000, type = "random")

IDW_Ozone = gstat::idw(formula = Arithmetic ~ 1,
                       locations = ozone_data_unique,
                       newdata = SampPoints ,
                       idp = 1)

IDW_NO2 = gstat::idw(formula = Arithmetic ~ 1,
                     locations = no2_data_unique,
                     newdata = SampPoints ,
                     idp = 1.3)

IDW_PM25 = gstat::idw(formula = Arithmetic ~ 1,
                      locations = pm25_data_unique,
                      newdata = SampPoints,
                      idp = 1.3)


tmapOzone = tm_shape(IDW_Ozone) +
  tm_dots(col = "var1.pred", size = 0.05, title = "Predicted Concentration")
+
  tm_shape(States) +
  tm_borders(col = "black", lwd=2) +
  tm_layout(main.title = "IDW Spatial Interpolated 2019 Ozone Concentration",
            main.title.position = "center",
            legend.position = c("center","bottom"),
            legend.outside = TRUE)

tmapNO2 = tm_shape(IDW_NO2) +
  tm_dots(col = "var1.pred", size = 0.05, title = "Predicted Concentration")
+
  tm_shape(States) +
  tm_borders(col = "black", lwd=2) +
  tm_layout(main.title = "IDW Spatial Interpolated 2019 NO2 Concentration",
            main.title.position = "center",
            legend.position = c("center","bottom"),
            legend.outside = TRUE)

tmapPM25 = tm_shape(IDW_PM25) +
  tm_dots(col = "var1.pred", size = 0.05, title = "Predicted Concentration")
+
  tm_shape(States) +
  tm_borders(col = "black", lwd=2) +
  tm_layout(main.title = "IDW Spatial Interpolated 2019 PM 2.5
Concentration",
            main.title.position = "center",
            legend.position = c("center","bottom"),
            legend.outside = TRUE)

par(mfrow=c(1,3))
tmapOzone; tmapNO2; tmapPM25

# Exporting the Data Cleaned as GeoJSON for Clustering in GeoDa
Ozone_States <- SpatialPolygonsDataFrame(States, data =
data.frame(ozone_data_unique@data, row.names = row.names(ozone_data_unique)))
NO2_States <- SpatialPolygonsDataFrame(States, data =
data.frame(no2_data_unique@data, row.names = row.names(no2_data_unique)))
PM25_States <- SpatialPolygonsDataFrame(States, data =
data.frame(pm25_data_unique@data, row.names = row.names(pm25_data_unique)))
```

```r
OzoneGJ <- geojson_json(Ozone_States)
NO2GJ <- geojson_json(NO2_States)
PM25GJ <- geojson_json(PM25_States)

write(OzoneGJ, file = "OzoneGJ.geojson")
write(NO2GJ, file = "NO2GJ.geojson")
write(PM25GJ, file = "PM25GJ.geojson")

OzoneIDWGJ <- geojson_json(IDW_Ozone)
write(OzoneIDWGJ, file = "OzoneIDWGJ.geojson")

# Air Pollutant Concentration over Longitude and Latitude
ggplot(data = ozone_data_unique@data)+
  geom_point(mapping = aes(x = Longitude, y = Latitude, size = Arithmetic),
             color = "brown", alpha = 3/4) +
  ggtitle("Ozone Concentration (ppm)") +
  coord_fixed()

ggplot(data = no2_data_unique@data)+
  geom_point(mapping = aes(x = Longitude, y = Latitude, size = Arithmetic),
             color = "purple", alpha = 3/4) +
  ggtitle("Nitrogen dioxide Concentration (ppm)") +
  coord_fixed()

ggplot(data = pm25_data_unique@data)+
  geom_point(mapping = aes(x = Longitude, y = Latitude, size = Arithmetic),
             color = "navy", alpha = 3/4) +
  ggtitle("PM2.5 Concentration (µg/m³)") +
  coord_fixed()

# Creating Bounding Box
bbox(ozone_data_unique)
bbox(no2_data_unique)
bbox(pm25_data_unique)

# Normality Assumption Checking
par(mfrow=c(1,3))
hist(ozone_data_unique@data$Arithmetic, main="Ozone Concentration",
     xlab = "Ozone Concentration Value (ppm)")
hist(no2_data_unique@data$Arithmetic, main="Nitrogen dioxide Concentration",
     xlab = "Nitrogen dioxide Concentration (ppm)")
hist(pm25_data_unique@data$Arithmetic, main="PM 2.5 Concentration",
     xlab = "PM 2.5 Concentration (µg/m³)")

# Log-transformation of Air Pollutant Concentration Data
# Log-transformation of Ozone Concentration Data
ozone_data_unique@data$logArith <- log(ozone_data_unique@data$Arithmetic)

# Log-transformation of Nitrogen dioxide Concentration Data
no2_data_unique@data$logArith <- log(no2_data_unique@data$Arithmetic)

# Histogram of Air Pollutant Concentration after necessary Log-transformation
par(mfrow=c(1,3))
hist(ozone_data_unique@data$logArith, main="Log(Ozone Concentration)",
     xlab = "Log(Ozone Concentration)", cex.main = 1)
hist(no2_data_unique@data$logArith, main="Log(Nitrogen dioxide
Concentration)",
     xlab = "Log(Nitrogen dioxide Concentration)", cex.main = 1)
hist(pm25_data_unique@data$Arithmetic, main="PM 2.5 Concentration",
```

```
     xlab = "PM 2.5 Concentration (µg/m³)")

# All passes Normality Assumption

# Calculating Sample (Empirical) Variogram
# Ozone Concentration Sample Variogram
ozone.vgm <- variogram(logArith ~ 1, ozone_data_unique)

# NO2 Concentration Sample Variogram
no2.vgm <- variogram(logArith ~ 1, no2_data_unique)

# PM 2.5 Concentration Sample Variogram
pm25.vgm <- variogram(Arithmetic ~ 1, pm25_data_unique)

plot(ozone.vgm, main = "Sample Variogram of Log (Ozone Concentration)")
plot(no2.vgm, main = "Sample Variogram of Log(NO2 Concentration)")
plot(pm25.vgm, main = "Sample Variogram of PM 2.5 Concentration")

# List of Possible Model Variograms
show.vgms()

# List of semivariance value
SampleOzoneSemiVar <- ozone.vgm$gamma
SampleNO2SemiVar <- no2.vgm$gamma
SamplePM25SemiVar <- pm25.vgm$gamma

print("-----------Ozone Semivariance----------")
print(SampleOzoneSemiVar)
print("------------NO2 Semivaraince-----------")
print(SampleNO2SemiVar)
print("----------PM 2.5 Semivariance----------")
print(SamplePM25SemiVar)

# Sample Ozone Variogram would fit for "Nug" or "Per" or "Hol"
# Sample NO2 Variogram would fit for "Gau"
# Sample PM 2.5 Variogram would fit for "Wav" or "Per"

# Fitting possible model variograms
# Ozone Variogram of "Per", "Hol"
ozonePer.fit <- fit.variogram(ozone.vgm, model = vgm("Per"))
ozoneHol.fit <- fit.variogram(ozone.vgm, model = vgm("Hol"))

# NO2 Variogram of "Per", "Gau"
no2Per.fit <- fit.variogram(no2.vgm, model = vgm("Per"))
no2Gau.fit <- fit.variogram(no2.vgm, model = vgm("Gau"))

# PM 2.5 Variogram of "Per", "Wav"
pm25Per.fit <- fit.variogram(pm25.vgm, model = vgm("Per"))
pm25Wav.fit <- fit.variogram(pm25.vgm, model = vgm("Wav"))

# Plotting each possible fit
# Ozone Possible Variogram Fit Plot
plot(ozone.vgm, ozonePer.fit, main = "Ozone (Per)")
plot(ozone.vgm, ozoneHol.fit, main = "Ozone (Hol)") # Best represents

# NO2 Possible Variogram Fit Plot
plot(no2.vgm, no2Per.fit, main = "NO2 (Per)")
plot(no2.vgm, no2Gau.fit, main = "NO2 (Gau)") # Best represents

# PM 2.5 Possible Variogram Fit Plot
```

```r
plot(pm25.vgm, pm25Per.fit, main = "PM 2.5 (Per)")
plot(pm25.vgm, pm25Wav.fit, main = "PM 2.5 (Wav)") # Best represents

# Model Parameter Specification and Evaluations
# vgm(0.001009458, "Hol", 158.1515)
ozoneHol.fit # psill: 0.001009458    , range: 158.1515

# vgm(0.7650246, "Gau",104.2188)
no2Gau.fit # psill: 0.7650246  , range: 104.2188

# vgm(3.000621, "Wav", 213.0957)
pm25Wav.fit # psill: 3.000621 , range: 213.0957

LOOCV_krige_ozone <- krige.cv(logArith ~ 1,
                        ozone_data_unique,
                        model = ozoneHol.fit)

LOOCV_krige_no2 <- krige.cv(logArith ~ 1,
                        no2_data_unique,
                        model = no2Gau.fit)

LOOCV_krige_pm25 <- krige.cv(Arithmetic ~ 1,
                        pm25_data_unique,
                        model = pm25Wav.fit)

LOOCVKrigOzoneResid <- LOOCV_krige_ozone@data$residual
LOOCVKrigNO2Resid <- LOOCV_krige_no2@data$residual
LOOCVKrigPM25Resid <-LOOCV_krige_pm25@data$residual

RMSE_resid(LOOCVKrigOzoneResid)
RMSE_resid(LOOCVKrigNO2Resid)
RMSE_resid(LOOCVKrigPM25Resid)

options(scipen=999)
attr(ozoneHol.fit, "SSErr")
attr(no2Gau.fit, "SSErr")
attr(pm25Wav.fit, "SSErr")

# Creating a Convex Hull
convex_hull_point_ids_ozone <- ozone_data_unique@coords %>%
  chull()

convex_hull_point_ids_no2 <- no2_data_unique@coords %>%
  chull()

convex_hull_point_ids_pm25 <- pm25_data_unique@coords %>%
  chull()

# Sub-setting the points for convex hull and convert to data frame
convex_hull_points_ozone <-
ozone_data_unique@coords[convex_hull_point_ids_ozone,] %>%
  as.data.frame()

convex_hull_points_no2 <-
no2_data_unique@coords[convex_hull_point_ids_no2,] %>%
  as.data.frame()

convex_hull_points_pm25 <-
pm25_data_unique@coords[convex_hull_point_ids_pm25,] %>%
  as.data.frame()
```

```
# Plotting convex hull points for each air pollutant
plot(convex_hull_points_ozone)
plot(convex_hull_point_ids_no2)
plot(convex_hull_point_ids_pm25)

# Creating a Convex Hull from Polygon from the Convex Hull Points
# Creating a matrix of the convex hull point coordinates
coordsOzone <- coordinates(convex_hull_points_ozone)
coordsNO2 <- coordinates(convex_hull_points_no2)
coordsPM25 <- coordinates(convex_hull_points_pm25)

# Creating a Polygon
polyOzone <- sp::Polygon(coordsOzone)
polyNO2 <- sp::Polygon(coordsNO2)
polyPM25 <- sp:: Polygon(coordsPM25)

# Preparing data for a "sp" object
ID <- "Minimum Boundary"
PlsOzone <- Polygons(list(polyOzone), ID = ID)
PlsNO2 <- Polygons(list(polyNO2), ID = ID)
PlsPM25 <- Polygons(list(polyPM25), ID = ID)

SPlsOzone <- SpatialPolygons(list(PlsOzone))
SPlsNO2 <-SpatialPolygons(list(PlsNO2))
SPlsPM25 <- SpatialPolygons(list(PlsPM25))

plot(SPlsOzone)
plot(SPlsNO2)
plot(SPlsPM25)

# Creating evenly-spaced grid
gridOzone <- makegrid(SPlsOzone, cellsize = 0.1)
gridNO2 <- makegrid(SPlsNO2, cellsize = 0.1)
gridPM25 <- makegrid(SPlsPM25, cellsize = 0.1)

coordinates(gridOzone) <- ~ x1+x2
coordinates(gridNO2) <- ~ x1+x2
coordinates(gridPM25) <- ~ x1+x2

plot(gridOzone)
plot(gridNO2)
plot(gridPM25)

# Selecting only grid points within convex hull
# Checking which grid cells are in the polygon
grid_in_poly_ozone <- sp::over(gridOzone, SPlsOzone)
grid_in_poly_no2 <- sp::over(gridNO2, SPlsNO2)
grid_in_poly_pm25 <- sp::over(gridPM25, SPlsPM25)

# List of NAs and valid grid cells
grid_in_poly_ozone
grid_in_poly_no2
grid_in_poly_pm25

# Selecting only points that are not NA
grid_predict_ozone <- gridOzone[!is.na(grid_in_poly_ozone),]
grid_predict_no2 <- gridNO2[!is.na(grid_in_poly_no2),]
grid_predict_pm25 <- gridPM25[!is.na(grid_in_poly_pm25),]
```

```
# Plotting
plot(grid_predict_ozone)
plot(grid_predict_no2)
plot(grid_predict_pm25)

# Predicting Air Pollutant Concentration Values at New Points
# Setting CRS of grid_predict's of every air pollutants to the spatial data
set
crs(grid_predict_ozone) <- CRS("+proj=longlat +datum=WGS84 +no_defs")
crs(grid_predict_no2) <- CRS("+proj=longlat +datum=WGS84 +no_defs")
crs(grid_predict_pm25) <- CRS("+proj=longlat +datum=WGS84 +no_defs")

# Kriging of Ozone Concentration over Grid Cells
ozone_krige <- krige(log(Arithmetic) ~ 1,
                        ozone_data_unique,
                        grid_predict_ozone,
                        model = ozoneHol.fit)

# Kriging of NO2 Concentration over Grid Cells
no2_krige <- krige(log(Arithmetic) ~ 1,
                        no2_data_unique,
                        grid_predict_no2,
                        model = no2Gau.fit)

# Kriging of PM 2.5 Concentration over Grid Cells
pm25_krige <- krige(Arithmetic ~ 1,
                        pm25_data_unique,
                        grid_predict_pm25,
                        model = pm25Wav.fit)

# Plotting Air Pollutant Concentration Predictions
# Exponential-transformation of Log(Ozone) back to original scale
ozone_krige_copy <- ozone_krige
ozone_krige_copy@data$Pred <- exp(ozone_krige@data$var1.pred)

# Exponential-transformation of Log(NO2) back to original scale
no2_krige_copy <- no2_krige
no2_krige_copy@data$Pred <- exp(no2_krige@data$var1.pred)

# Plotting the original scale of Kriging Prediction Air Pollutant
Concentrations
KrigOzonePlot <- spplot(ozone_krige_copy, "Pred", main="Kriging of Ozone",
                         sp.layout = list(list(States, fill=alpha("white",
0.01), lwd=2, first=FALSE)))

KrigNO2Plot <- spplot(no2_krige_copy, "Pred", main = "Kriging of NO2",
                      sp.layout = list(list(States, fill=alpha("white",
0.01), lwd=2, first=FALSE)))

KrigPM25Plot <- spplot(pm25_krige, "var1.pred", main = "Kriging of PM 2.5",
                       sp.layout = list(list(States, fill=alpha("white",
0.01), lwd=2, first=FALSE)))
grid.arrange(KrigOzonePlot, KrigNO2Plot, KrigPM25Plot, ncol = 3)
```