

Fine Particulate Matter Concentrations in the Dallas-Ft. Worth Area

Adam Pacsi

Graduate Student

Department of Chemical Engineering

University of Texas at Austin

GISWR Fall 2011

Class Project

Table of Contents

Introduction 3

Methods 4

Results 12

Conclusions 14

Appendix 1: Sample Code 16

References 18

Introduction

Project Problem

The National Ambient Air Quality Standards (NAAQS) require that fine particulate matter (PM_{2.5}) concentrations remain below 15 µg/m³ on an annual arithmetic average basis.¹ The NAAQS for fine particulate matter are set at this level in order to minimize health effects such as asthma and increased overall morbidity that have been linked to elevated PM_{2.5} concentrations.² In addition, PM_{2.5} particles can reflect sunlight in the visible spectrum of light which can lead to a hazy or foggy environments.³ In order to ensure compliance with all NAAQS, the Texas Commission on Environmental Quality (TCEQ) is charged with maintaining a series of site monitors to measure target pollutant levels throughout Texas. Due to the relatively high cost of maintaining PM_{2.5} monitors compared to other parameters such as ozone, only a small subset of TCEQ air quality monitors throughout the state gather fine particulate matter data. The purpose of this project is to determine which additional TCEQ air quality monitoring locations should be considered for new PM_{2.5} monitors based on local population density, distance from a current PM_{2.5} monitor, and predicted relative PM_{2.5} concentration. The primary assumption for this analysis is that PM_{2.5} concentrations will follow the same trend as ozone concentrations, which is an applicable assumption for many urban areas throughout the world.⁴ The Dallas-Ft. Worth (DFW) metropolitan area will be used as a case study for this project, but the techniques that were developed would be applicable to any urban area that has sufficient monitoring locations for ozone and PM_{2.5} and where fine particulate matter concentrations are found to trend with ozone concentrations. The Dallas-Ft. Worth area has 20 ozone monitors and 8 ozone monitors currently.⁵ Placing new fine particulate matter monitors in the same location as a currently-operating ozone monitor would decrease capitol costs for the TCEQ in that the organization would not need to acquire additional land or site materials (such as fencing and concrete pads).

Sources of Data

Two types of data (census and air quality) were used in the completion of this project. The location of current PM_{2.5} and ozone monitoring stations in the Dallas-Fort Worth region were obtained from the TCEQ's website for the Site Implementation Plan (SIP) for Ozone Attainment in the Dallas Fort Worth Region.⁵ This data included the latitude and longitude for each monitoring location and the types of air quality parameters that were measured there. In addition, the TCEQ website had hourly ozone concentrations at each of the current monitors that were gathered for the past three years (2008-2010). Year 2000 Census Data was found in a free, ArcGIS-ready format from ESRI under the Tiger/Line Database.⁶ This data included shapefiles for each census tract in the counties of interest as well as a database file containing relevant population information for each census tract.

Relevant GIS Techniques

The methods of implementation for the project will be described in more detail; however, the project involved several broad techniques for use in Arc-GIS. These included changing the projection for different types of data, calculating distance between points, and raster interpolation between points. In addition, a categorization code (Appendix 1) was written in Visual Basic Script for use in the Field Calculator interface within the program. While these techniques were developed within an air quality context, there is significant applicability in other fields, including hydrology.

Methods

Geographic Extent

The primary projection was set as the Albers Projection scheme used in Exercise 1 for the course.⁷ The shapefiles for the state of Texas and for the counties of Texas were added to ArcGIS in order to set the Albers projection from those datasets as the reference projection. All other datasets used in the analysis were eventually converted to this projection.

The location of each of the monitors in the Dallas-Fort Worth Area were obtained from the TCEQ⁵ and compiled in an Excel spreadsheet. Each monitoring site was also given an attribute based on the parameters that were measured there (Site_Type). The codes were as follows: (1) ozone only, (2) ozone and PM, (3) ozone and PM compliance, (4) meteorology only, and (5) PM compliance only. PM compliance meant that a site reported whether PM levels had exceeded the NAAQS of 15 $\mu\text{g}/\text{m}^3$ during each hour that was reported. The spreadsheet was exported to ARC-GIS and the points were displayed using the Display XY Data option using a NAD83 projection. The thirteen counties (as shown in Figure 1) in the DFW region that contained monitoring sites were grouped as an additional feature class and were given a new attribute (NAMES) that included the name of the county. These counties were the only ones used for further analysis.

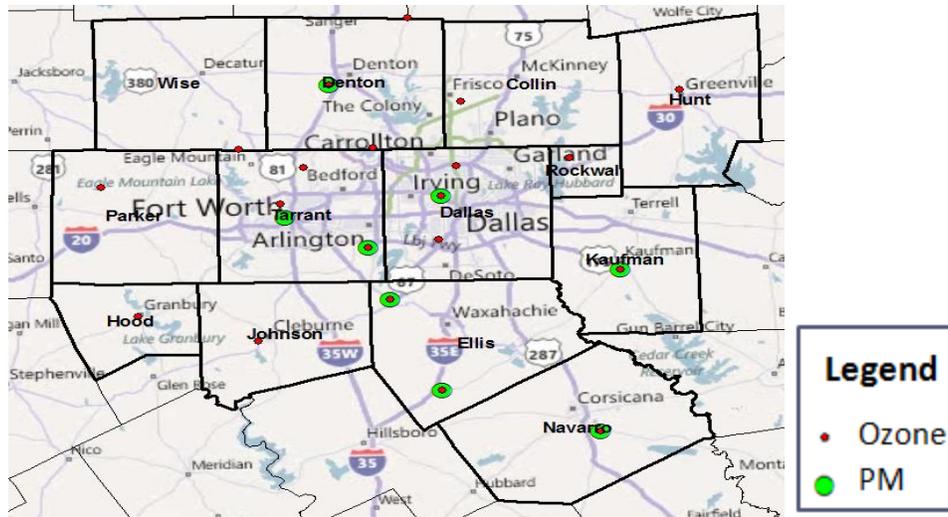


Figure 1: Thirteen Counties in DFW With Current Monitoring Sites

Population Density

The ESRI website for the TIGER/Line Data Set⁶ contained ArcGIS-compatible shapefiles for each census tract in a NAD83 geometric datum. Using the drop down menus on the website, only the shapefiles of census tracts (Census Tracts 2000) from the thirteen DFW counties were selected and downloaded from the ftp site. The census tract files were compressed by county after downloading. All folders were decompressed and added to the ArcGIS basemap using the Add Data function.

The census tract shapefiles had to be converted to the Albers Projection that was used as the basemap. In addition, most of the internal functions within ArcGIS require that the data be projected in terms of length (meters, feet, etc) rather than as decimal degrees in order to operate. While several methods were attempted to facilitate this change in spatial references, the only successful methods involved using features of ArcCatalog 10. The added shapefiles for census tracts in each county were converted each to an individual feature class. This was accomplished by exporting the data to a Geodatabase that had NAD83 as the reference datum. Then, ArcMap was closed and ArcCatalog was opened. The tool Feature Class Conversion was used to transform the shapefiles from the NAD83 to Albers Projection by using the Texas counties shapefile (Albers) as the output coordinate system. It is important to note that this function did not work in the ArcMap environment. The thirteen feature classes containing the shapefiles for the census tracts of each DFW county were combined into a single feature class (Projected_Tracts) using the Merge Tool in a single operation.

While the feature class contained the desired shapefiles, there was no population information. Using the ESRI website,⁶ the population (as well as racial make-up) data were

downloaded for all census tracts in the state of Texas as a single database (.dbf) file. There were not options to limit this request by the counties of interest. The downloaded database file was unzipped and added to ArcMap using the Add Data button. A table join was created between the Projected_Tracts feature class and the database file based on the unique identifier for each census tract (STFID). A new attribute for population (Pop) was added to Projected_Tracts and was equated to the corresponding value for population from the joined database file using the Field Calculator. The table join was then removed.

A new attribute column for population density (Pop_Density) was added to the census tracts feature class. The Field Calculator was used to fill the data in this column by dividing the population of the tract by its area. The statistics function was used to identify the largest value for population density in the study region. A new column for normalized population density (NORM_Pop_Density) was created and filled with the Field Calculator using the formula of Population Density divided by maximum population density in the DFW region in order to ensure all values were between 0 and 1. The resulting normalized relative population density map for the DFW region is shown in Figure 2 with the green to yellow to red coloration showing increasing population density:

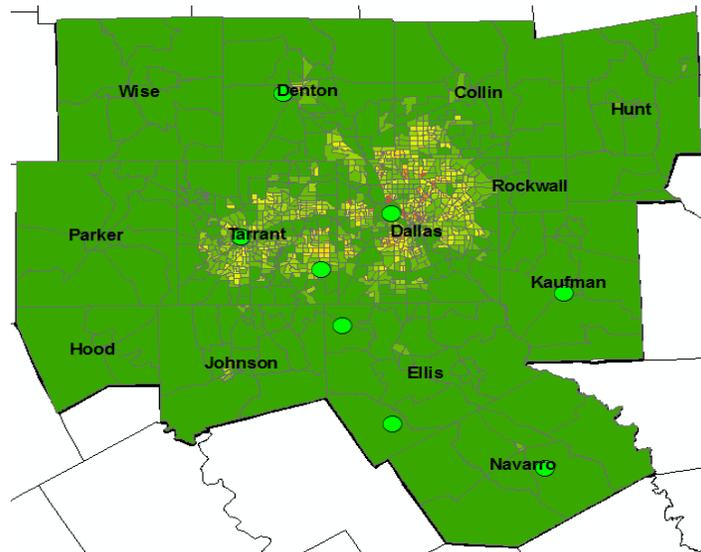


Figure 2: Population Density in the DFW Region

The distribution of the values made sense with the higher population density values being in downtown Dallas and along major highways that lead to Dallas and Ft. Worth. The surrounding counties showed very low population density.

Distance to Current PM Monitoring Site

Using Excel, the data from TCEQ that contained the location of monitoring sites was filtered to include only the sites that currently monitor fine particulate matter. This was completed by selecting based on the SITE_TYPE attribute with 2, 3, or 5 as the value. The spreadsheet data was then added to the basemap using the Add Data button. The projection scheme was changed from NAD83 (decimal degrees) to the Albers Projection using the Feature Class Conversion Tool using the same process as with the census shapefiles.

In order to calculate a distance from each census tract to the closest current PM_{2.5} monitor, the census tracts had to be represented as a point. This was accomplished through the Mean Center function with the Case Field selected as STFID. The connection to the STFID was important for later joins to the other tract data. The distance to the closest current PM_{2.5} site for each census tract center was calculated with the Analysis tool called Near. This function added the closest PM monitoring station's Object_ID and a distance value to the Attribute Table of the tract center feature class. A join was made between the tract center feature class and the tract feature class, and the Field Calculator was used to obtain a permanent distance value in the tract feature class that would remain after the table join was removed. A new normalized distance attribute (Norm_Dist) was added to the table and was calculated using the same procedure that was used for population density. It is important to note that the Near function will over-write any previous result if the same output feature class is selected. Therefore, proper table joining is essential when using the near function multiple times during a project. The normalized distance to a current PM site for the DFW region is shown in Figure 3 with the green to yellow to red conversion being related to increasing distance from the green circles (current PM monitors).

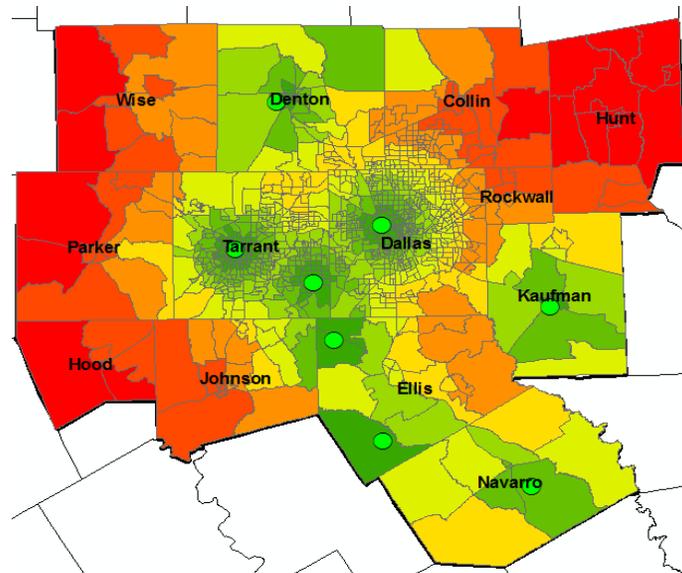


Figure 3: Normalized Distance Data to Current PM Sites

The output values seemed to be reasonable given that the sites near a PM monitor appear green while those that were further away appeared to be red. It is important to note that most the downtowns of Dallas and Ft. Worth are primarily green, which indicates that the current PM monitors are relatively well-dispersed in these areas.

Concentration Data

The analysis for this project assumes that PM_{2.5} levels trend like ozone levels in the DFW area. Since there are more than two times as many ozone monitoring locations (20) as PM_{2.5} monitoring locations (8), the interpolated ozone values will be used as a relative measure of PM_{2.5} concentration (i.e. high ozone areas are also high PM_{2.5} areas). The original spreadsheet with TCEQ monitoring sites was again filtered to contain only monitoring locations that measured ozone (Codes 1, 2, and 3). The yearly average ozone concentration at each selected monitoring site was obtained for 2008 through 2010, and a three-year average value was calculated in Excel. The new spreadsheet was export to ArcMap and added to the Albers projection in the same method as the current PM_{2.5} sites were added during the previous step.

Three interpolation methods (Natural Neighbor, Spherical Kriging, and IDW) were used to create concentration rasters. The Spherical Kriging and IDW were limited to using monitoring location values within 50 km in order to preserve the location variability of concentration values in the region. Since natural neighbor method only selected the nearest monitoring location, this restriction was not needed. The resulting concentration rasters for the three methods are shown in Figure 4. The color scheme for these maps is that the concentration decreases in the following order: white, red, yellow, and green.

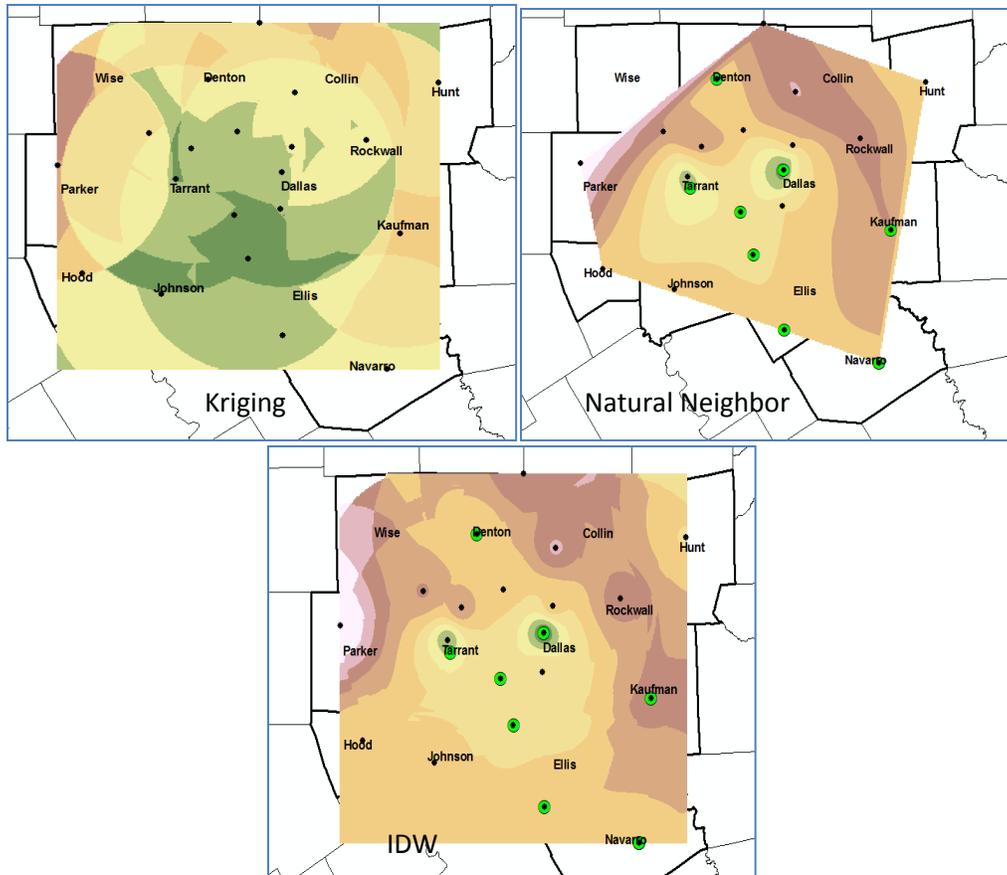


Figure 4: Ozone Interpolation Rasters from Three Methods

Since some of the census tracts (the areas that are not covered by the raster values in Figure 4) were outside of the interpolated area of the concentration rasters, an extrapolation method was developed in order to obtain a concentration in each census tract. Fortunately, the method developed for this project in order to obtain a concentration for each census tract functioned as both an interpolation and extrapolation method as will be described below. Before this method was undertaken, the float values for the raster were converted to integer values using the int function.

While several methods to convert the raster to a feature class based on the census tract were attempted, the method that worked best was a conversion of the raster cells (squares) to a single point at the center of the cell. This is a common simplification made in air quality modeling programs with raster data. The Near function was then used as described in the Distance Calculations to find the nearest raster center value to each census tract center. The values for each interpolation method were then added to the tracts feature class using the series of table joins and Field Calculator operations that were described in the Distance Calculations section of the paper. In addition, the normalization procedure that had been previously discussed was used for the data from each of the three interpolation methods. The

results were normalized concentration values by site for each of the three interpolation methods. The results are shown in Figure 5.

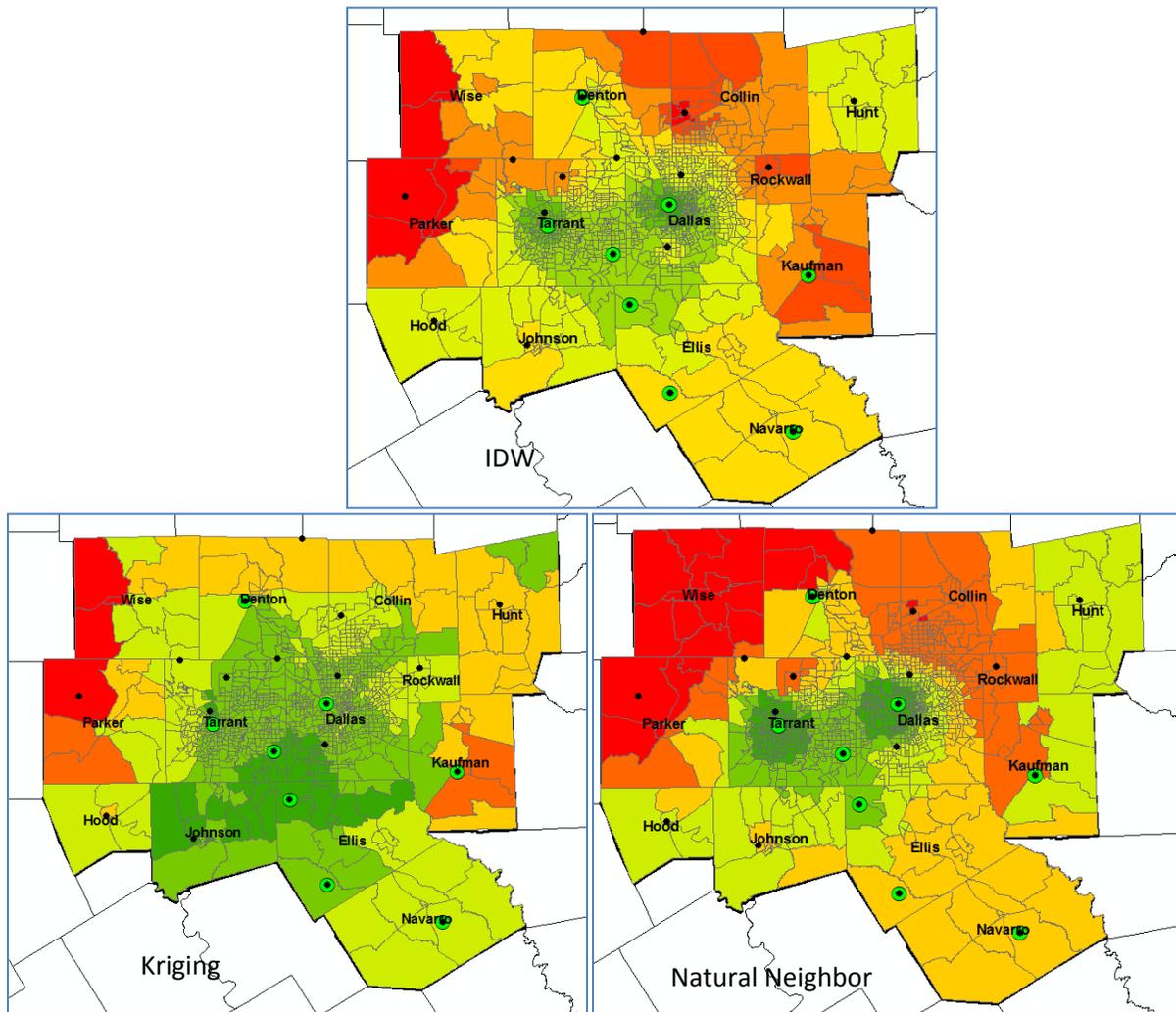


Figure 5: Normalized Concentration Values by Census Tract for Three Interpolation Methods

The results are consistent with observed ozone data from the DFW region with the worst air quality being northwest of the city, which is downwind of the urban plume of the Dallas and Fort Worth downtown areas.

Data Categorization

The first attempt at combining the three criteria for determining which ozone monitor to add to the PM_{2.5} monitoring involved simply adding the normalized values from the distance and population density calculations to the concentration value obtained from each of the three interpolation techniques that were used. The problem was that these values were dominated by the concentration value (which typically ranged from 0.8 to 1) versus the other two criteria

(which typically had values between 0 and 0.2). Therefore, a new classification method was developed based on the Field Calculator program that is shown in Appendix 1. The program assumes that the input data from an existing column (population density, distance, and each concentration) is normally distributed. The Visual Basic Script populates a new category column for the input data with an integer 1 through 5 based on the distance from the mean. The classification system is as follows: (5) if the value is greater than two standard deviations above the mean, (4) if the value is between one and two standard deviations above the mean, (3) if the value is between one standard deviation above and one standard deviation below the mean, (2) if the value is between one and two standard deviations below the mean, and (1) if the value is less than two standard deviations below the mean. The program only needs a mean, standard deviation, and input column name to run. The mean and standard deviation for the original column can be calculated using the column statistic functions in ArcGIS.

Five new columns were created to be populated with category numbers (distance, population density, and concentration from the three interpolation methods). The script was uploaded to each of column field calculators using the load button. The program was run with the correct standard deviation, mean, and original data column. Three new columns were then created for final analysis values (on a scale of 3-15) based on the addition of the category value for population density, distance, and each concentration method. The results for the three different scenarios that were examined are shown in Figure 6.

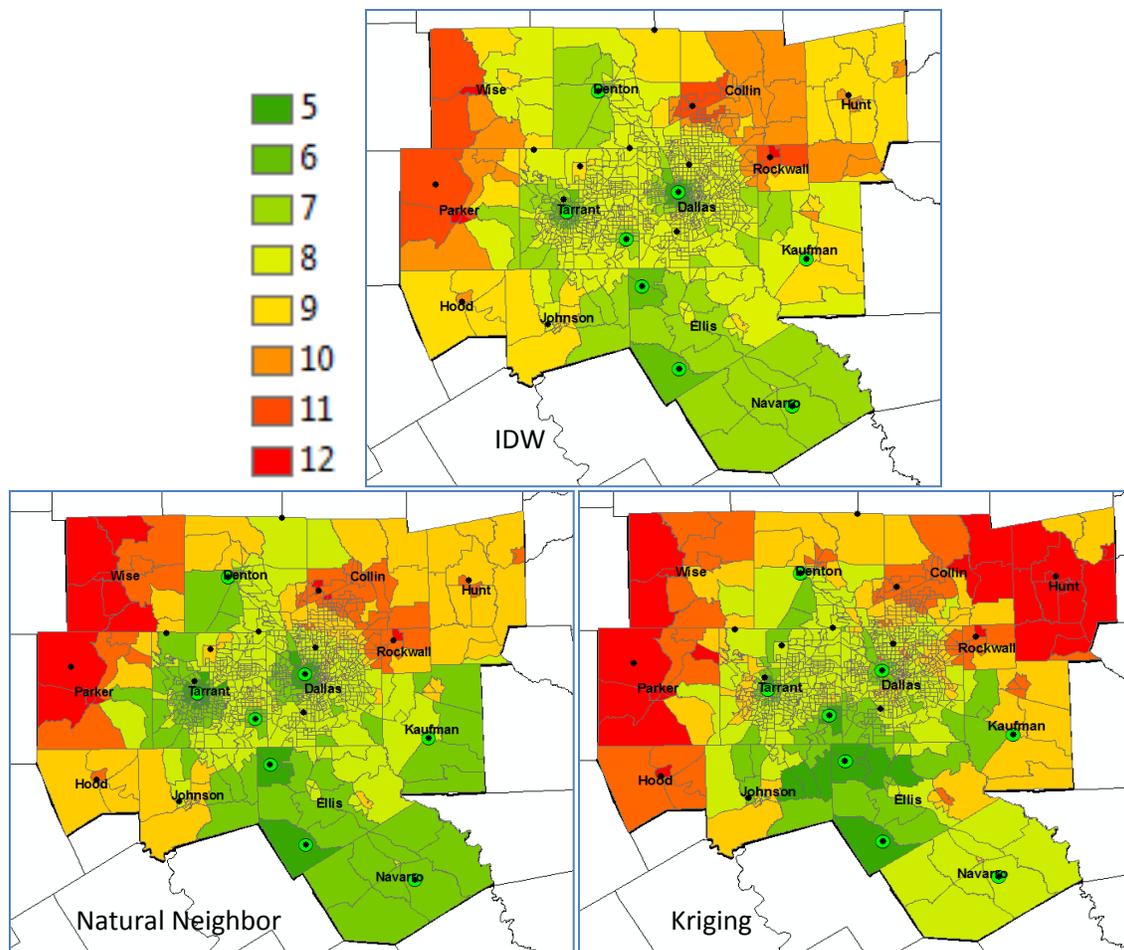


Figure 6: Final Ranking for Each Scenario by Census Tract

Results

The analysis was focused on the red areas for each scenario that were displayed in Figure 6. These values had a combined score (distance + population density + concentration) of greater than 11 and contained a current ozone monitoring location. The value of 11 was selected since it provided the top three or four monitoring locations for each scenario. The results are displayed in Figure 7:

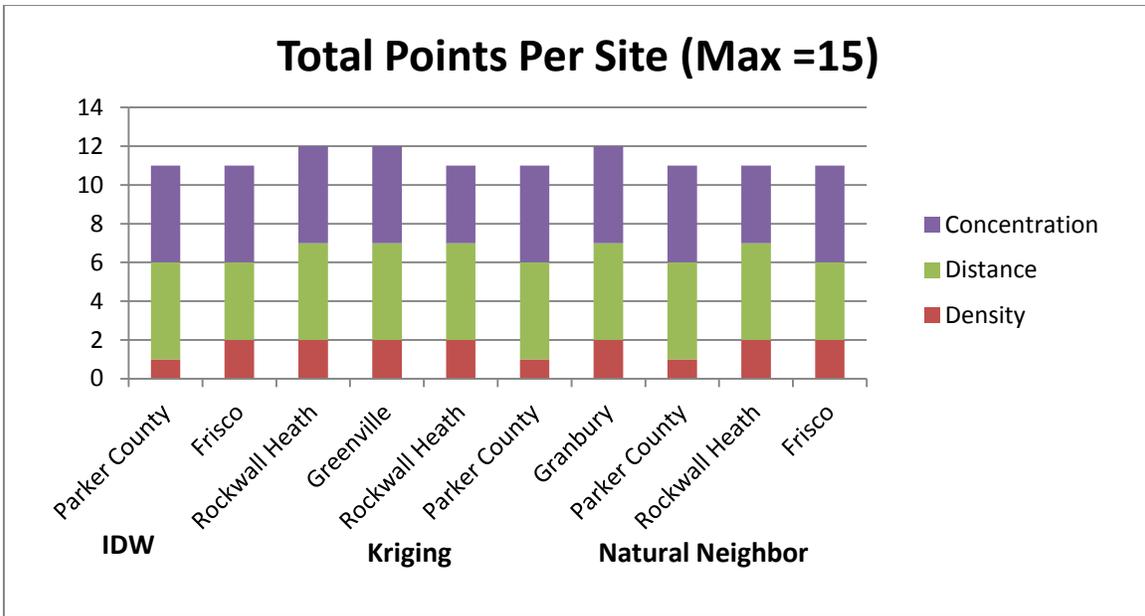


Figure 7: High Ranking Monitoring Locations for by Scenario

The final composite score for each of the high ranking sites (greater than 11) was predominately composed of Distance and Concentration scores. The population density contributed no more than 2 (out of 5) possible points to highest scoring sites in all three scenarios. This seems to indicate that the coverage in the densely populated areas in the downtowns of Dallas and Ft. Worth are already well-covered by current PM monitors. Thus, the primary factors that influence the need for an additional PM monitor were distances from a current PM site and concentration values.

The sites that appeared in multiple scenarios were selected from Figure 7 for further analysis since these values were considered to be independent of the interpolation method that was used. Rockwall Heath and Parker County (purple circles) appeared in all three scenarios, and Frisco (red circle) appeared in both the IDW and Natural Neighbor Scenarios. The location of these monitors is shown in Figure 8.

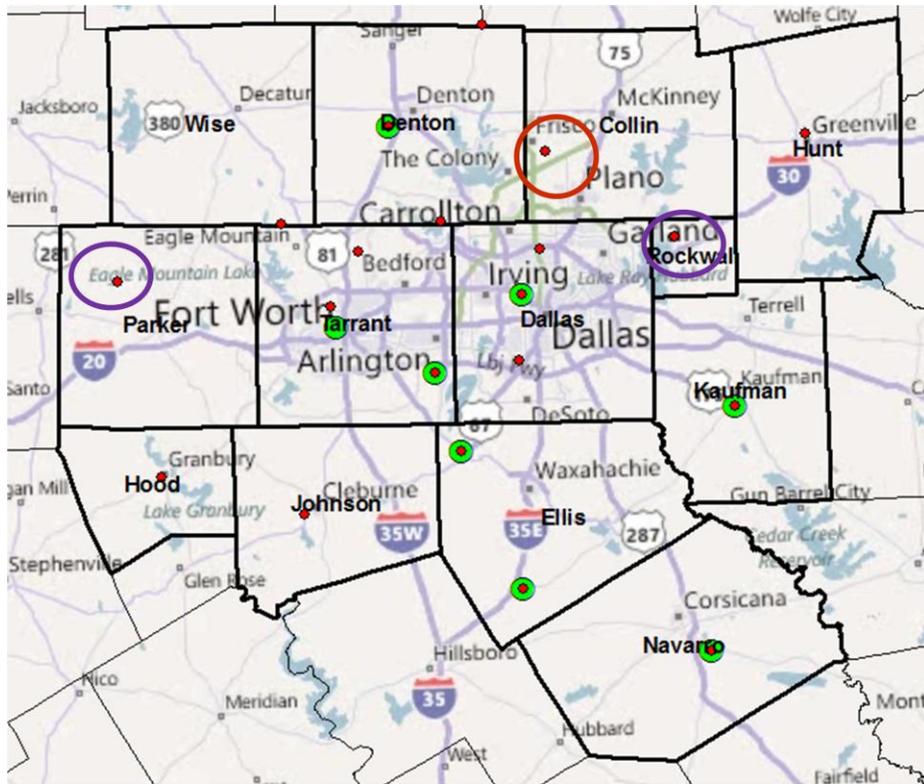


Figure 8: Location of Target Monitoring Sites for Additional PM Sites

The locations of the recommended monitoring locations were all north of the downtown areas of Dallas. This made intuitive sense since the prevailing winds from downtown Dallas (a major population center) move the pollutants towards the north and northwest on average throughout the year. Furthermore, these areas tend to be less well covered with current PM monitoring locations than the downtowns of Dallas and Ft. Worth.

Conclusions

Assuming that $PM_{2.5}$ tracks like ozone in the Dallas-Ft. Worth area, then based on distance, population density, and predicted PM concentrations, the Parker County, Rockwall Heath, and Frisco ozone monitoring stations should be considered as locations for additional $PM_{2.5}$ sites by the TCEQ. The sites are all north of the city which is the downwind direction from the major population centers in the region. Thus, concentrated plumes from these urban areas likely are making their way to this region. This transportation effect may be even more important with fine particulate matter than with ozone since the atmospheric lifetime of these pollutants is generally larger than that of ozone.³ However, this methodology is intended to be a first step in the process of citing a new PM monitor in any region of interest. Factors such as the localized wind direction at the monitoring location and localized high $PM_{2.5}$ sources should also be considered before installing a new PM monitor at the recommended locations. For

example, locating a monitor directly downwind of a coal fired powerplant may cause readings that are higher than the regional average since the higher sulfur dioxide levels can lead to high levels of particulate sulfate formation in highly localized settings. These monitors are intended to capture the average PM concentration over a wider area rather than to be used in source apportionment studies.

` In conclusion, the methods developed for this project can be applied to any urban area with sufficient existing ozone and particulate matter monitors in a metropolitan area where PM_{2.5} trends like ozone. In addition, the methods outlined in the report such as interpolation between monitoring locations and categorization of normalized data have applications in other fields even though they were developed in an air quality framework.

Appendix 1: Code

Sample Code in Visual Basic

```
mean = 0.884
sigma = 0.022
value = [Kriging_Ozone_Norm]
If value < mean-2*sigma Then
cat =1
ElseIf value > mean+2*sigma Then
cat =5
Else
cat = 0
End If
```

```
If value <= mean+2*sigma Then
If value > mean+sigma Then
cat = 4
End If
End If
```

```
If value <= mean+sigma Then
If value > mean-sigma Then
cat = 3
End If
End If
```

```
If value <= mean-sigma Then
If value > mean-2*sigma Then
cat = 2
End If
End If
```

Usage

The mean and standard deviation for the column containing the original data should be calculated using the statistical features in ArcGIS and entered as numerical values on the first two lines of code. The third line should contain the bracketed name for the original data column. The entire code can be cut and paste into the VB code block in the field calculator or saved as an external text file and uploaded to the code block.

References

- ¹Environmental Protection Agency. *National Ambient Air Quality Standards*.
<http://www.epa.gov/air/criteria.html>
- ²Pope CA, DV Bates, MA Raizenne. *Health Effects of Particulate Air Pollution: Time for Reassessment?*. Environmental Health Perspectives. Volume 103, Number 5, May 1995.
- ³Seinfeld John and Spyros Pandis. *Atmospheric Chemistry and Physics: Air Pollution to Climate Change*. Second Edition. Wiley-Interscience. 2006.
- ⁴Gramsch E, F Cereceda-Balic, P Oyola, D von Baer. *Examination of pollution trends in Santiago de Chile with cluster analysis of PM₁₀ and Ozone data*. Atmospheric Environment 40 (2006) 5464-5474
- ⁵Texas Commission on Environmental Quality. *Dallas-Ft. Worth Eight-Hour Ozone SIP Modeling (2006 Episode): Air Quality Monitoring Sites*. 2010.
http://www.tceq.texas.gov/airquality/airmod/data/dfw8h2/dfw8h2_site.html
- ⁶US Bureau of the Census. *Census 2000 TIGER/Line Data*. Via the ESRI website.
<http://www.esri.com/data/download/census2000-tigerline/index.html>
- ⁷Maidment, David. *Introduction to ArcGIS Desktop*. September 2011.
<http://www.ce.utexas.edu/prof/maidment/giswr2011/Ex1/Ex12011.htm>