U.S. Hot Spots for Ground-Level Ozone

By:
Jamie Encarnacion-Peña
Christopher M. Johnson

Mentor:
Dr. Brian J. Smith PhD.

Department of Biostatistics
University of Iowa
2009

Outline

- Background
  - What is ozone?
  - Health effects
  - Where is the data obtained?
- Objective
- Analysis
  - Multiple Regression
  - Geostatistical Analysis
- Results
- Conclusion
- Future Work
- Acknowledgements

Background

- What is ozone?
  - Ozone is a triatomic molecule, which consists of three oxygen atoms; this gas exists in different quantities throughout the Earth's atmosphere.
  - The upper atmosphere is called ozone layer, which protects the Earth from destructive ultraviolet rays.
  - At ground-level, ozone is a dangerous air pollutant that affects the respiratory system of human beings.

Background

- Health Effects
  - The minor health effects are chest pain, coughing, irritation, and congestion.
  - Worst health effects would be bronchitis, emphysema, asthma, reducing lung function, and inflammation of the linings of the lung.
  - Repeated inhalation of ozone can cause permanent scarring of lung tissue.

Background

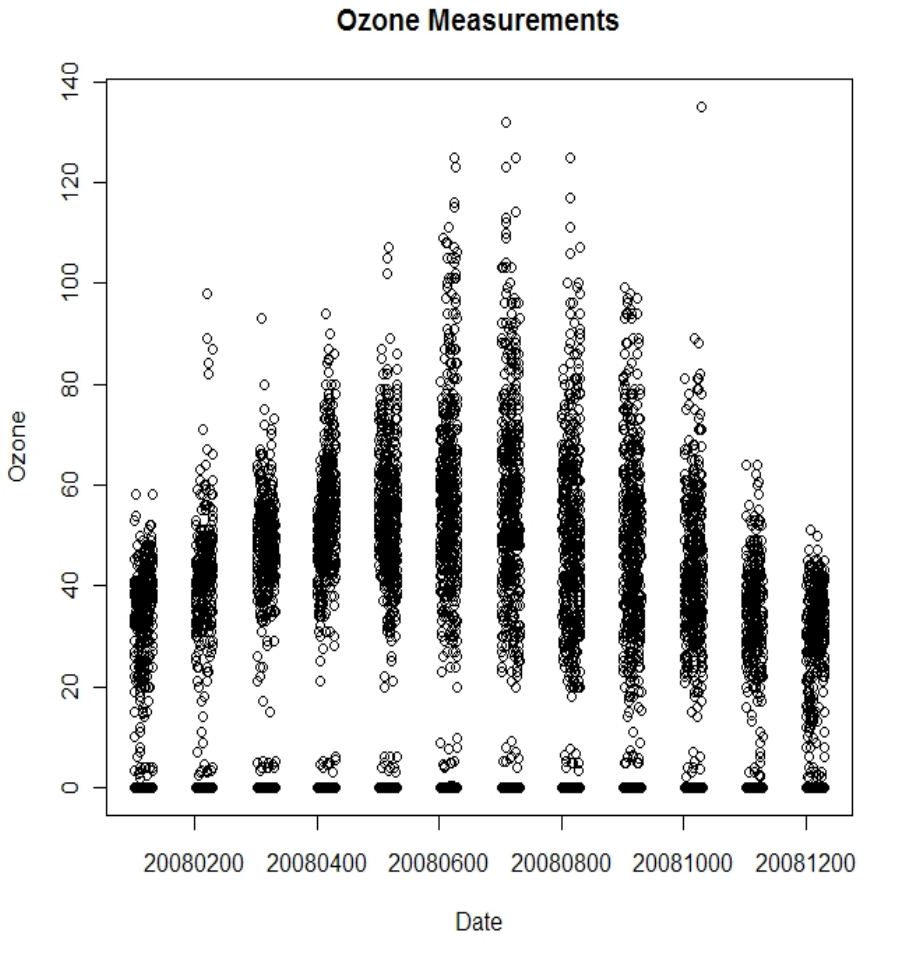
- Where is the data obtained?
  - All of the data analyzed has been collected and maintained by the Air Quality System(AQS) database.
  - The AQS contains ambient air pollution data collected by EPA, state, local, and tribal air pollution control agencies from thousands of monitoring stations.
  - AQS also contains meteorological data and descriptive information about each monitoring station (including its geographic location and its operator).

Objective

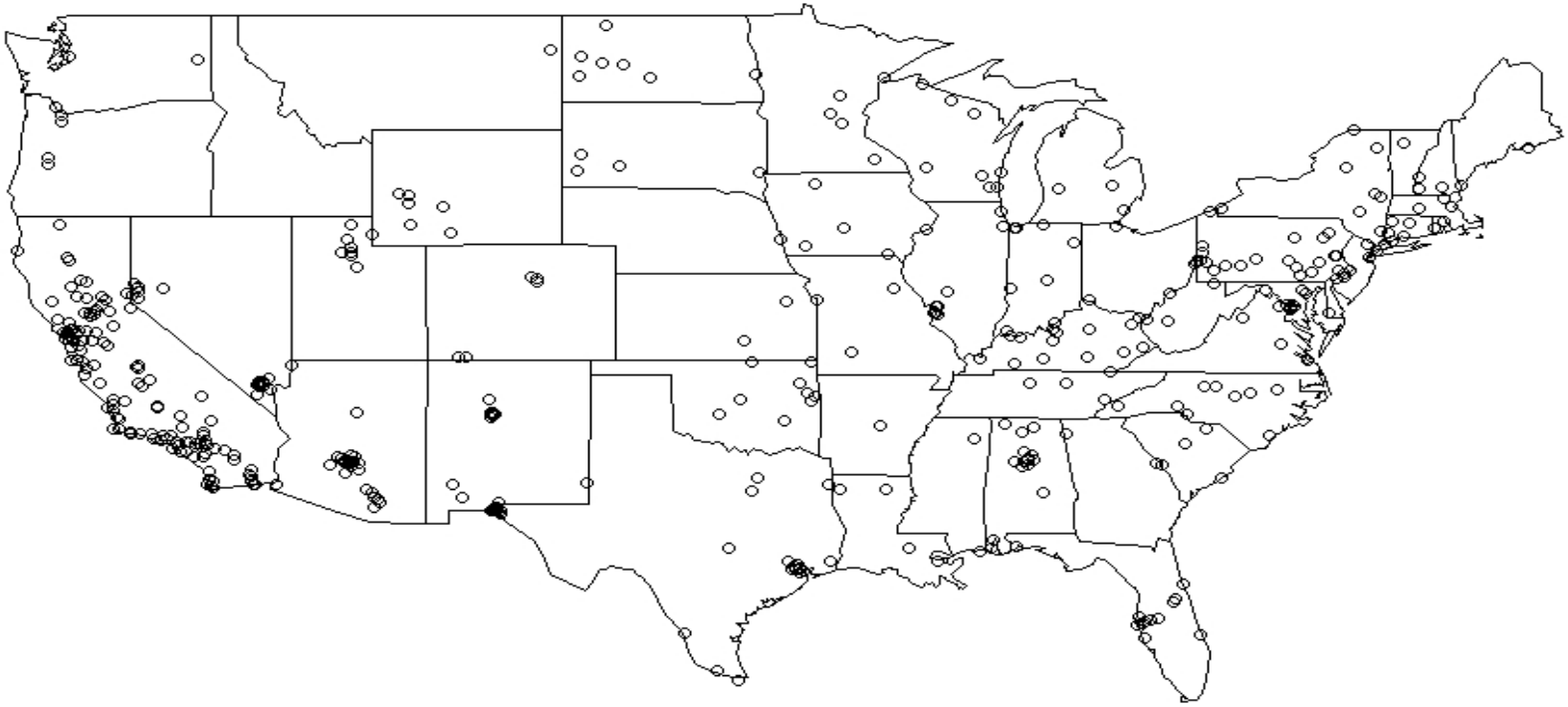
- To identify sites with high ozone concentration throughout the country, while accounting for seasonal and daily variations in order to account for other variables such as:
  - Vehicle emissions
  - Industrial sites

AQS Data Analysis

- Hourly measurements were taken from 140 cities; therefore, only 12,000 Air Quality System ozone measurements were used.
- The data was collected in 2008.
- Summary Statistics of Ozone
  - Min = .001 ppm
  - Max = 135.0 ppm
  - Mean = 20.0 ppm
  - Median = .067 ppm
  - Stnd Dev. = 25.79 ppm



Ozone Measurements Sites



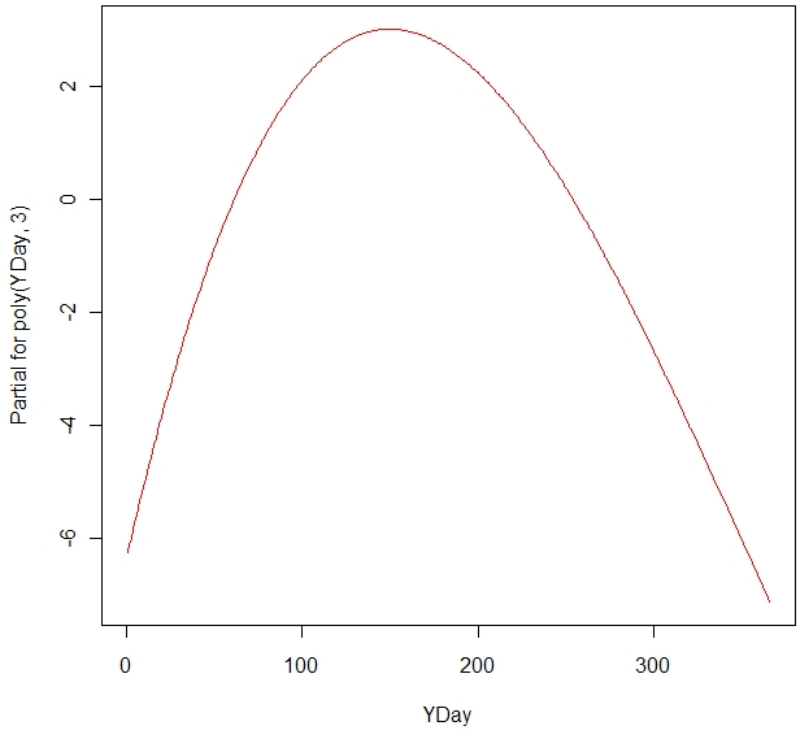
Analysis/Regression Model

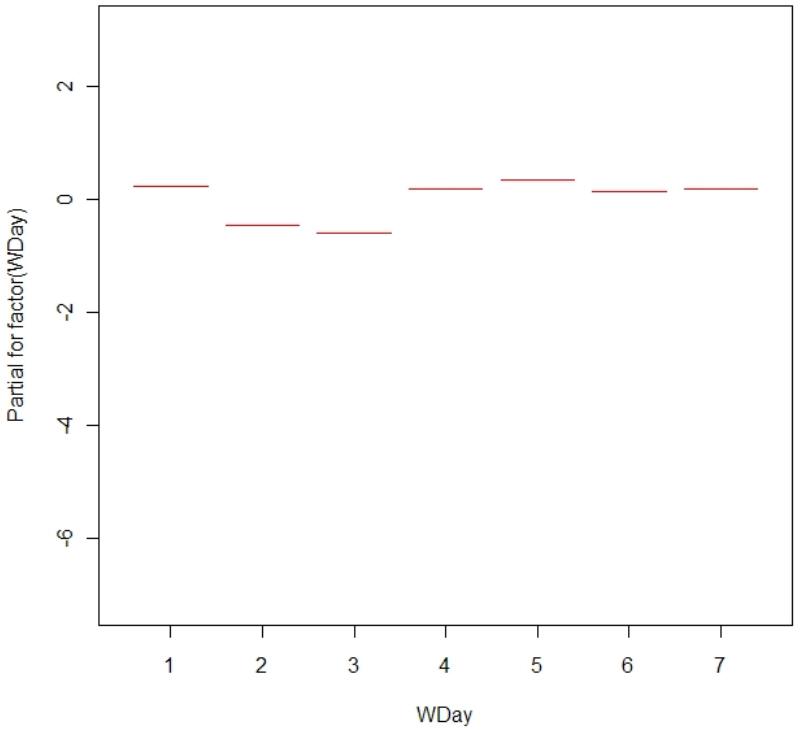
- Regression is used to estimate the relationship between ozone and the independent variables such as days of the week, month, and year.
- Multiple Regression Model
  - Let y be the actual ozone measurements
  - $x_1, x_2, \ldots, x_k$ are independent variables and possibly non-linear effects
  - $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$
  - $\alpha, \beta_1, \beta_2, \ldots, \beta_k$ are estimated effects of the independent variables
  - ε is the residual's coefficient
    - Differences between the "y" and the estimated effect of the predictors x
    - $\varepsilon = y - \alpha - \beta_1 x_1 - \beta_2 x_2 - \ldots - \beta_k x_k$

Analysis/Regression Model

- We tried different regression models with different combinations of the three independent variables
- The Final Model:
  - Included days of the year as a third order polynomial
  - Separate effects for days of the week

Regression Model in terms of days of the year and days of the week





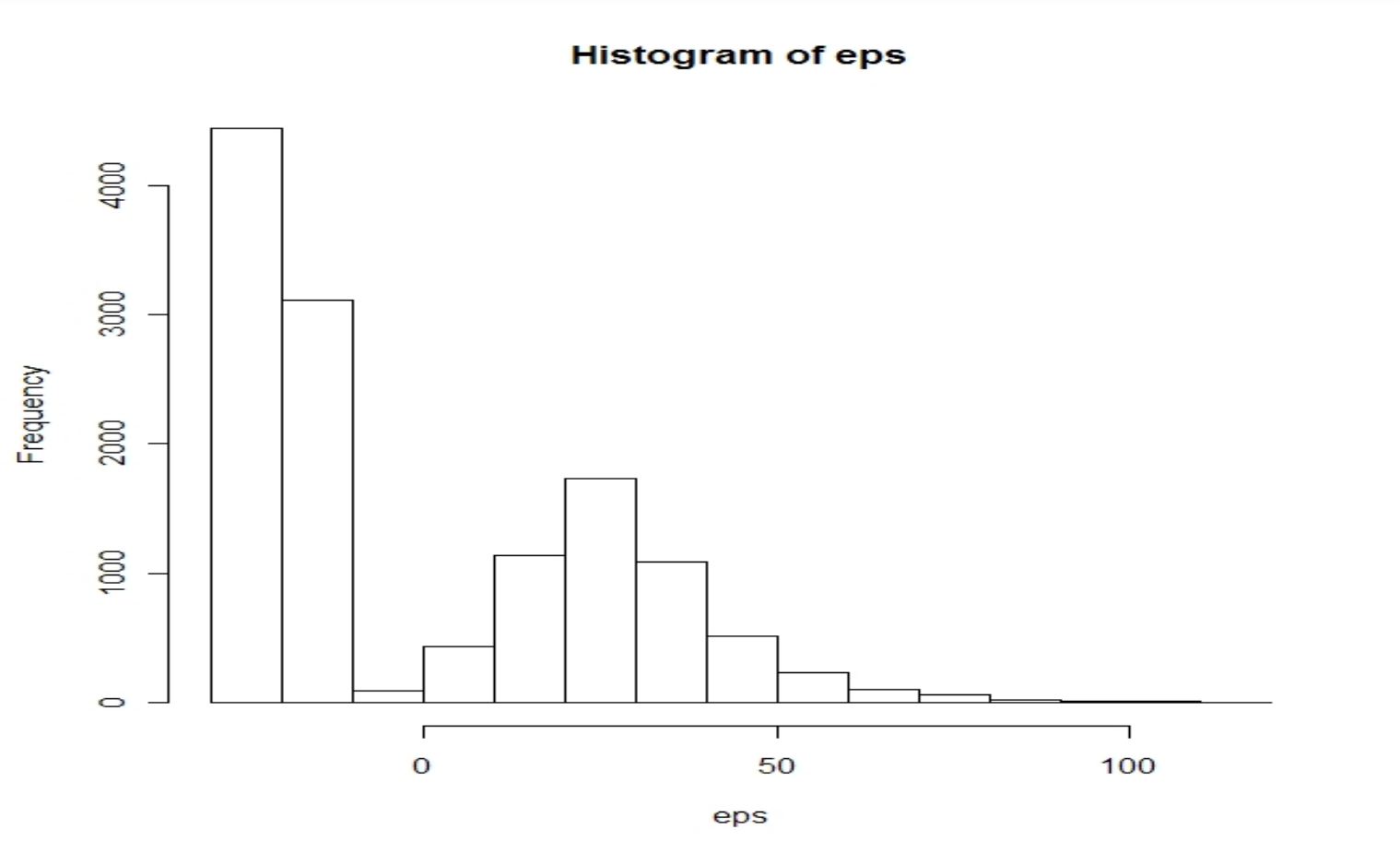
Geostatistical Analysis

- Spatial Analysis Goals
  - To look at spatial patterns in the residuals
    - Which are left-over variability from the regression on days of the year and days of the week

What are Residuals?

- The residuals for our data are the differences between the measured ozone values and the regressed (fitted) function values.
- The fitted function value is the value that the statistical model says the sample "should" have.

Histogram of Residuals



Spatial Interpolation Formula

- Let S_i be the spatial location of the i^{th} measurement
- $Z(S_i)$ denotes the i^{th} ozone measurement
- $Z(S_o)$ is the predicted measurement at a given site S_o estimated as the weighted average $\sum \lambda_i = 1$
  - $$\hat{Z}(S_o) = \sum_{i=1}^{N} \lambda_i Z(S_i)$$
  - Where the λ_i are the weights
  - λ_i takes on larger values for S_i closer to S_o
  - We will estimate λ_i from the data

Interpolation Weights

- Assume

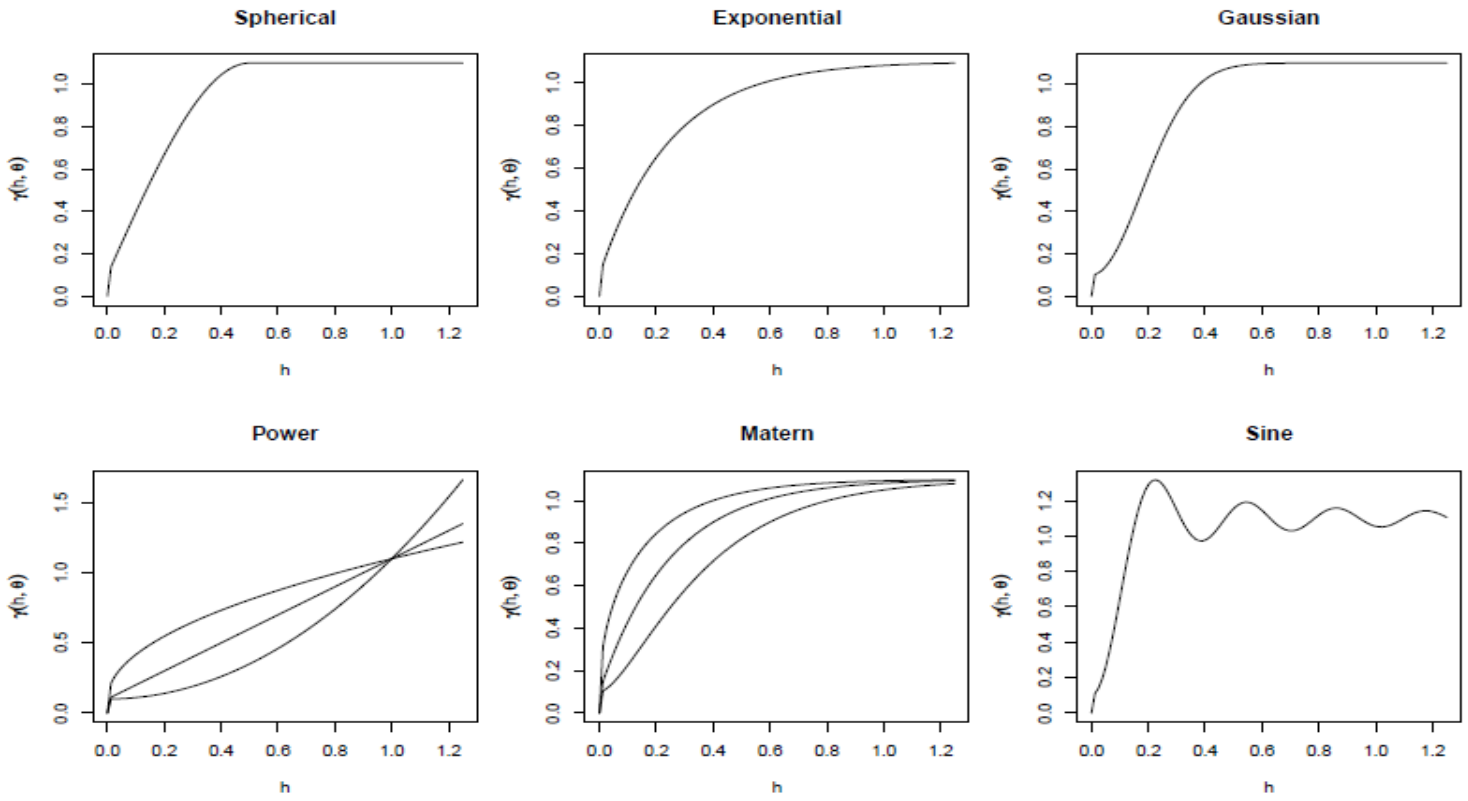
  λ_i = correlation between $Z(S_0)$ and $Z(S_i)$

  $\lambda_i = \rho(h)$

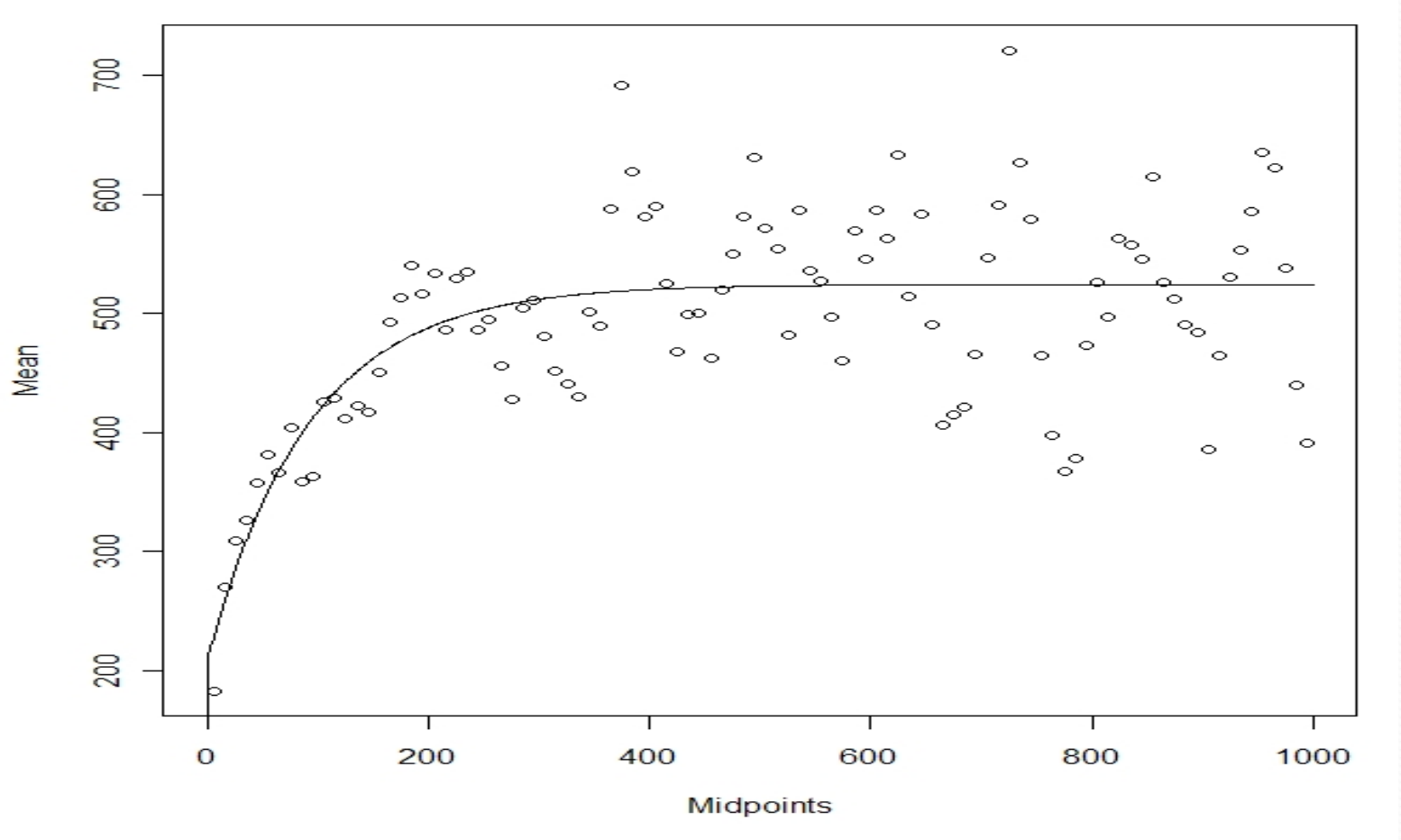
  Where $h = S_i - S_0$ is the distance between the prediction site S_0 and measured site S_i, and ρ is a correlation function
- Can be written as $\rho(h) = 1 - \frac{\gamma(h,\theta)}{\lim_{h\to\infty}\gamma(h,\theta)}$
- Where $\gamma(\cdot)$ is referred to as the semivariogram and θ are the parameters of the semivariogram.

Semivariogram

Figure: Examples of Parametric Semivariogram Functions.



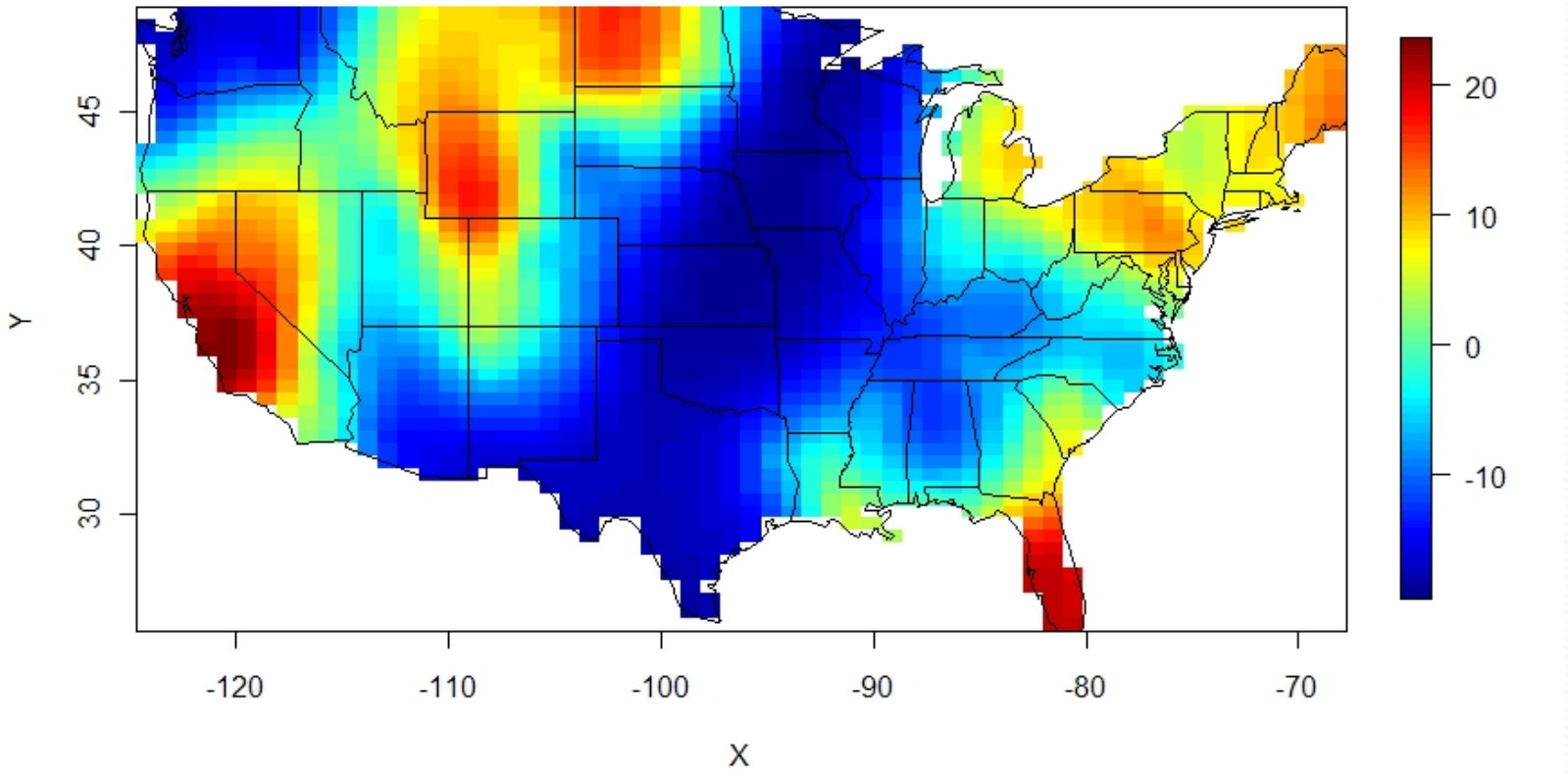
Empirical Semivariogram



Cont. of Empirical Semivariogram

- Data were plotted with circles
- Selected exponential semivariogram base on the pattern in the data
- Used numerical optimization to estimate $\hat{\theta}$ parameter
  - Which consisted of finding the best fitting line to the data
- Calculated weights λ_i from the semivariogram
- Predicted values across U.S

Results



Conclusion

- Ozone is a molecule that is harmless in the upper atmospheres of Earth, but has dangerous health effects as ground-level ozone such as chest pain and asthma. The objective of the research was to predict areas in the United States using data from the Air Quality System database that may have elevated levels of this dangerous molecule. Multiple regression and analysis using interpolation were used to created a spatial interpretation of the data which shows the predicted values of ozone across the country.

Future Work

- Since we did not account for temperature, or other variables, the work shows other causes of ozone other than the season and day of the week.
- Future research may involve investigating these other possible variables.

References

- (2009, May 20). Ground-level Ozone. Retrieved July 22, 2009, from U.S. Environmental Protection Agency Web site: http://www.epa.gov/air/ozonepollution/
- (2007, Dec 19). Air Quality System (AQS). Retrieved July 22, 2009, from U.S. Environmental Protection Agency Web site: http://www.epa.gov/ttn/airs/airsaqs/basic_info.html
- Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005) Applied Linear Statistical Models, Fifth Edition, New York: McGraw-Hill/Irwin Publishers.

Acknowledgements

- Thanks to our mentor Dr. Brian J. Smith PhD.
- Dr. Gideon J. Zamba PhD.
- Dr. Kathryn Chaloner PhD.
- Dr. Jeffrey D. Dawson PhD.
- Our T.A. Yu-Hui Chang
- Alliance/Vigre for a great summer!