Comparing the Spatial Qualities from Two Measurements of Air Quality over Chicago

Marie Ozanne
Mount Holyoke College
Alliance (DMS 0502354) – University of Iowa
Department of Biostatistics

Outline

- Background
- Linear Regression
- Variograms
- Spatial Models
- Results
- Conclusions
- Challenges

Important Terms

- Particulate matter (particle pollution, PM): mix of solid particles and liquid droplets found in the air
- Aerosol Optical Depth (AOD): measure of the degree to which airborne particles absorb light
- Variogram: function describing degree of spatial dependence of a spatial random field or stochastic process

Particulate Matter

- Two classes of particulate matter:
  - PM_{10}: "inhalable coarse particles" 2.5μm -10μm in diameter
    - "Primary particles"
  - $PM_{2.5}$: "fine particle" air pollution of diameters less than or equal to 2.5μm
    - "Secondary particles"
- Sources:
  - PM_{10}: construction sites, fields, unpaved roads, smokestacks, fires
  - $PM_{2.5}$: industrial processes, automobiles, power plants

Why $PM_{2.5}$? Health

- Health Concerns:
  - Smaller particles pose a greater threat
  - According to the EPA, effects of exposure include:
    - increased respiratory symptoms
    - decreased lung function
    - aggravated asthma
    - development of chronic bronchitis
    - irregular heartbeat
    - nonfatal heart attacks
    - premature death in people with heart or lung disease

      - Information Source: http://www.epa.gov/air/particlepollution/health.html

Why $PM_{2.5}$? Environment

- Major cause of "haze" in the United States
- Fine particles can be carried long distances by the wind and can settle:
  - Making lakes and streams acidic
  - Changing nutrient balance along the coasts and large river basins
  - Leaching nutrients from the soil
  - Damaging forests and crops
  - Affecting ecosystems

Why $PM_{2.5}$? Aesthetic Damage

- Buildings and Monuments:
  - Staining
  - Decay of stone
  - Corrosion of metal
  - Deterioration of paint
- Automotive coatings
  - Permanently etched surfaces
- Reduces societal value:
  - Bridges
  - Tombstones
  - Statues
  - Monuments

AOD: A Predictor of $PM_{2.5}$ Levels

- Advantages:
  - Read by satellites at consistent time intervals
  - Easy to obtain for any location
  - Contain predictors of $PM_{2.5}$ levels
- Challenges:
  - Contain information on entire atmosphere
  - Need to create spatial models with appropriate predictor variables

EPA Special Trends Network (STN)

- Established in 2000 to measure $PM_{2.5}$ composition
- Originally had 13 sites
- Now has 54 trends sites and 150 state and local sites following protocol
- Sampling schedules: 1 in 3 days

IMPROVE (Interagency Monitoring of Protected Visual Environments)

- Sampling network primarily focused on preserving visibility in parklands
- Most of the ~150 IMPROVE sites are in rural areas
- Sampling schedules are 1 in 3 days

MODIS (MODerate Imaging Spectroradiometer)

- Onboard NASA Terra satellite
- Provides near global coverage daily
- Measures satellite aerosol optical depth
- Data available for download from NASA

Air Quality In the Chicago Area

- Ground Sites:
  - Fifteen sites
  - Only nine sites had sufficient data to be statistically useful in this analysis
- Satellite Data Collection
  - AOD readings are taken every hour of every day
- Project Intent:
  - Find statistically significant predictors of $PM_{2.5}$ ground level values AOD measurements
  - Provide continuous $PM_{2.5}$ readings for the region

Simplification: Removing Time

- Methods:
  - Measurements averaged over time at each geographic location
  - Locations with sufficient data collection included in averaged data
- Reasons:
  - Including time makes computations much more difficult
  - Running data in its raw form takes a long time

Linear Regression: $PM_{2.5}$ Data

- $PM_{2.5} = \beta_0 + \beta_1 * lat + \beta_2 * lon + \varepsilon$
  - "lat": latitude
  - "lon": longitude
  - ε: error

$PM_{2.5}$ Regression Summary

- Residuals:
-     Min   1Q Median   3Q   Max
- -0.9234 -0.6891 -0.0714 0.3225 2.3666

- Coefficients:
-             Estimate Std. Error t value Pr(>|t|)
- (Intercept) 267.045   127.274   2.098   0.0807 .
- lat.avg.new   -2.426   2.566 -0.945   0.3810
- lon.avg.new   1.751   1.912   0.916   0.3952
- ---
- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Residual standard error: 1.168 on 6 degrees of freedom
- Multiple R-squared: 0.4534,   Adjusted R-squared: 0.2712
- F-statistic: 2.488 on 2 and 6 DF, p-value: 0.1633

Linear Regression: AOD Data

- AOD = $\beta_0 + \beta_1$*lat + β_2*lon + β_3*temp + β_4*wind + ε
  - "lat": latitude
  - "lon": longitude
  - "temp": temperature
  - "wind": wind speed
  - ε: error

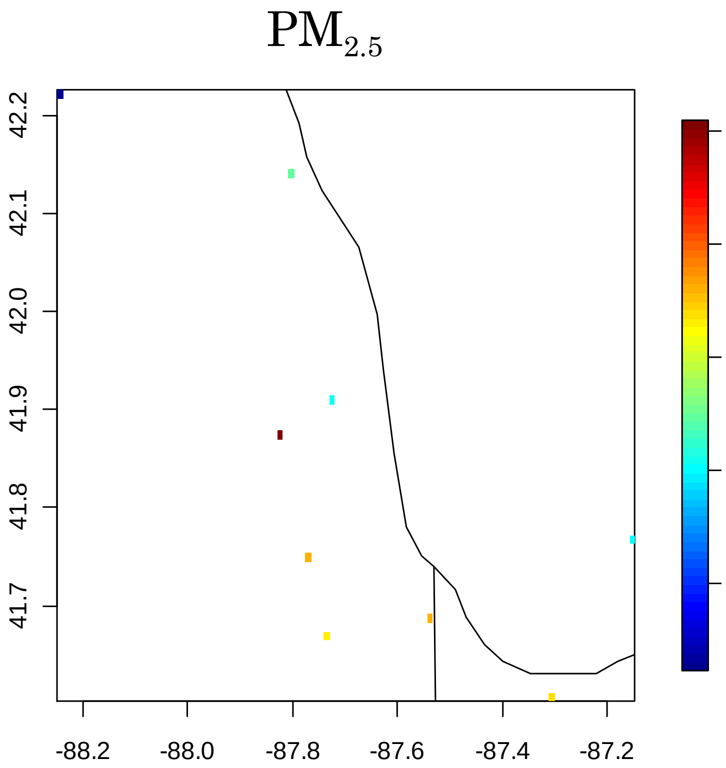
AOD Regression Summary

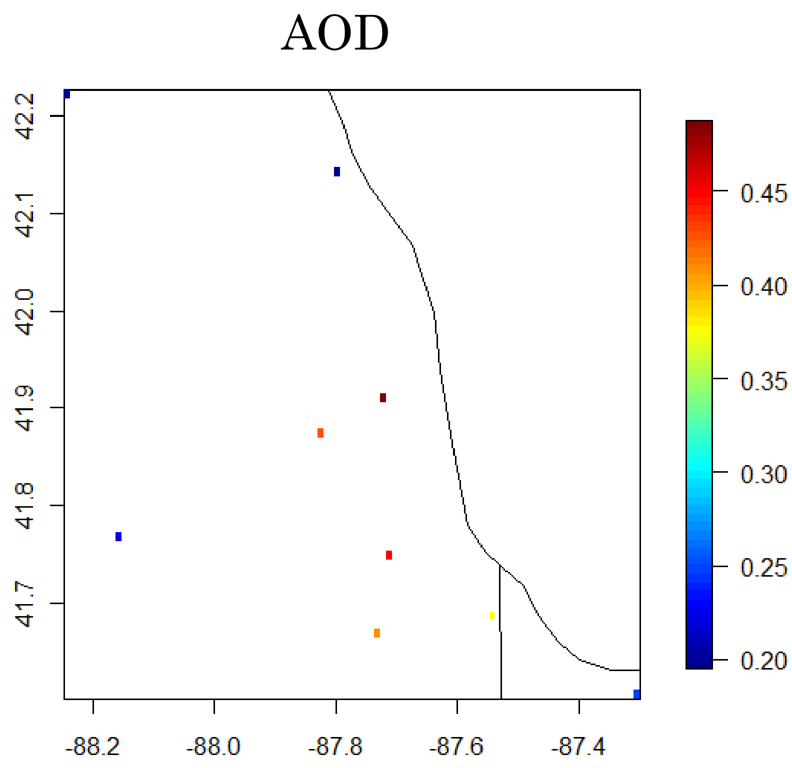
- Residuals:
- 1 2 3 4 5 6 7 8
- 0.05072 -0.04852 -0.04412 -0.09615 -0.01128 0.10277 0.03580 0.04843
- 9
- -0.03764

- Coefficients:
- Estimate Std. Error t value Pr(>|t|)
- (Intercept) -3.42650 12.17277 -0.281 0.7923
- lat.aodmeans 0.85452 0.42528 2.009 0.1149
- lon.aodmeans 0.37042 0.19677 1.882 0.1329
- aodtemp -0.09398 0.03876 -2.424 0.0724 .
- aodwnd 0.48757 0.19167 2.544 0.0637 .
- ---
- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

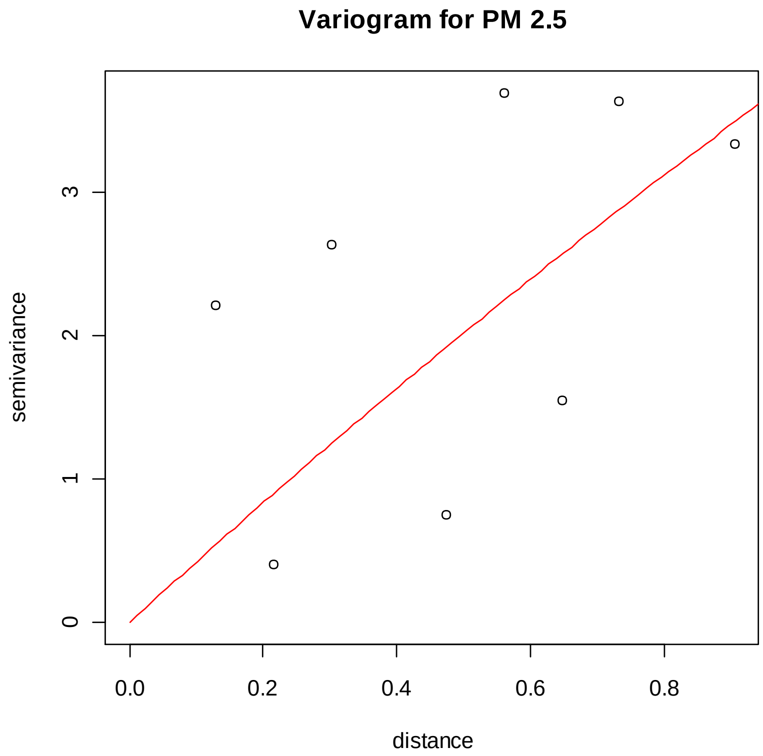
- Residual standard error: 0.08924 on 4 degrees of freedom
- Multiple R-squared: 0.713, Adjusted R-squared: 0.426
- F-statistic: 2.484 on 4 and 4 DF, p-value: 0.1998

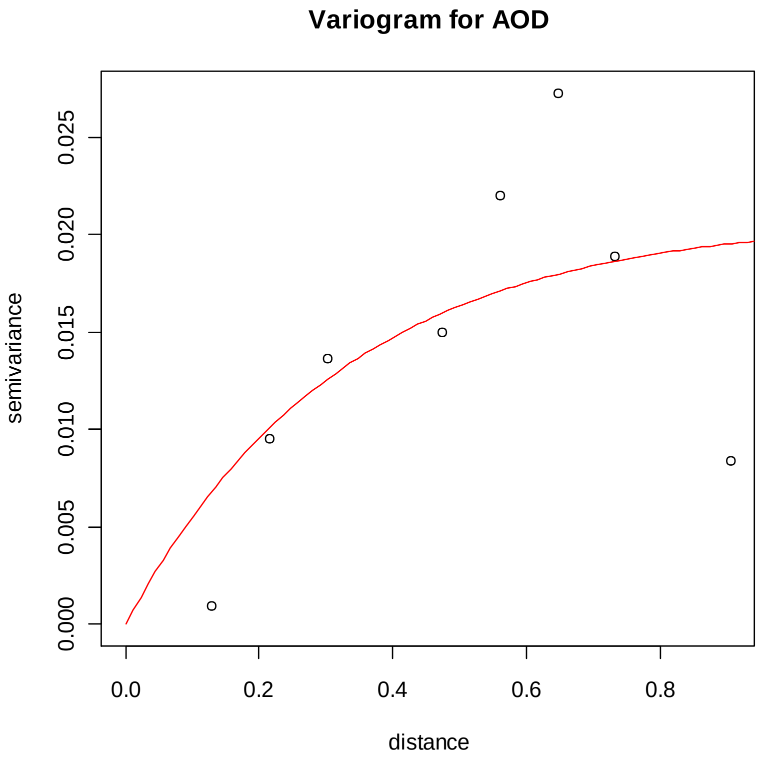
LOCATIONS


PM$_{2.5}$
42.2
42.1
42.0
41.9
41.8
41.7
-88.2
-88.0
-87.8
-87.6
-87.4
-87.2


AOD
42.2
42.1
42.0
41.9
41.8
41.7
-88.2
-88.0
-87.8
-87.6
-87.4
0.45
0.40
0.35
0.30
0.25
0.20

Variograms


Variogram for PM 2.5
semivariance
distance
0
1
2
3
0.0
0.2
0.4
0.6
0.8


Variogram for AOD
semivariance
distance
0.000
0.005
0.010
0.015
0.020
0.025
0.0
0.2
0.4
0.6
0.8

Spatial Models

- Phi (ϕ): range
- Sigma2.z (σ_z^2): variance due to spatial error
- Sigma2.e (σ_e^2): variance of measurement error
- Beta (β): fixed effects coefficient

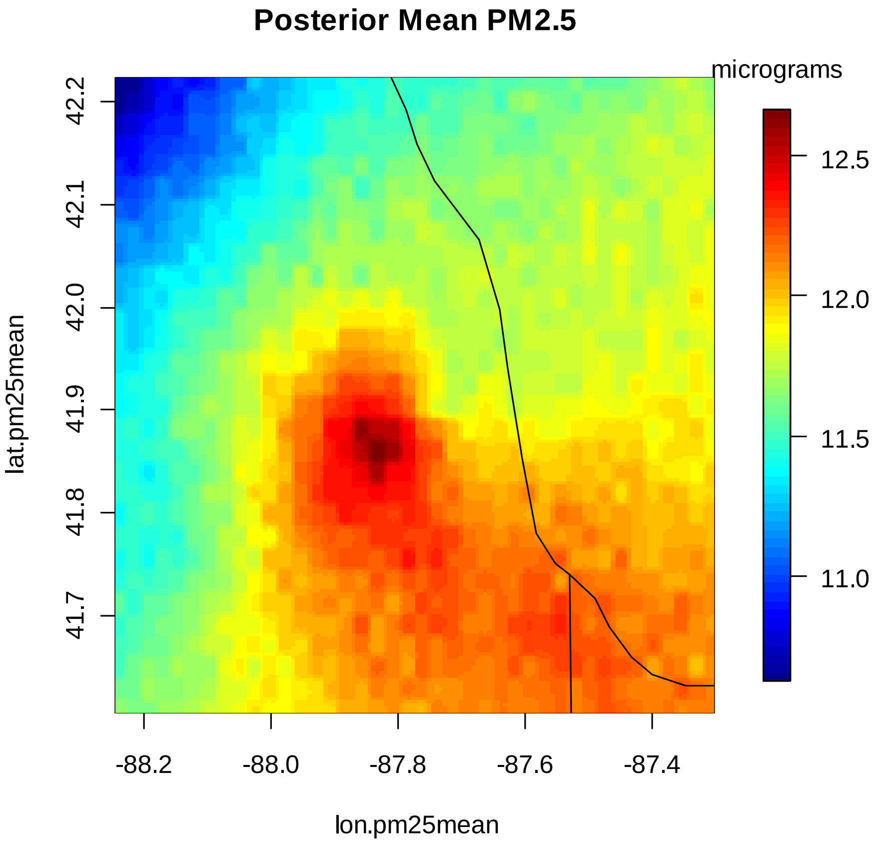
Spatial Models

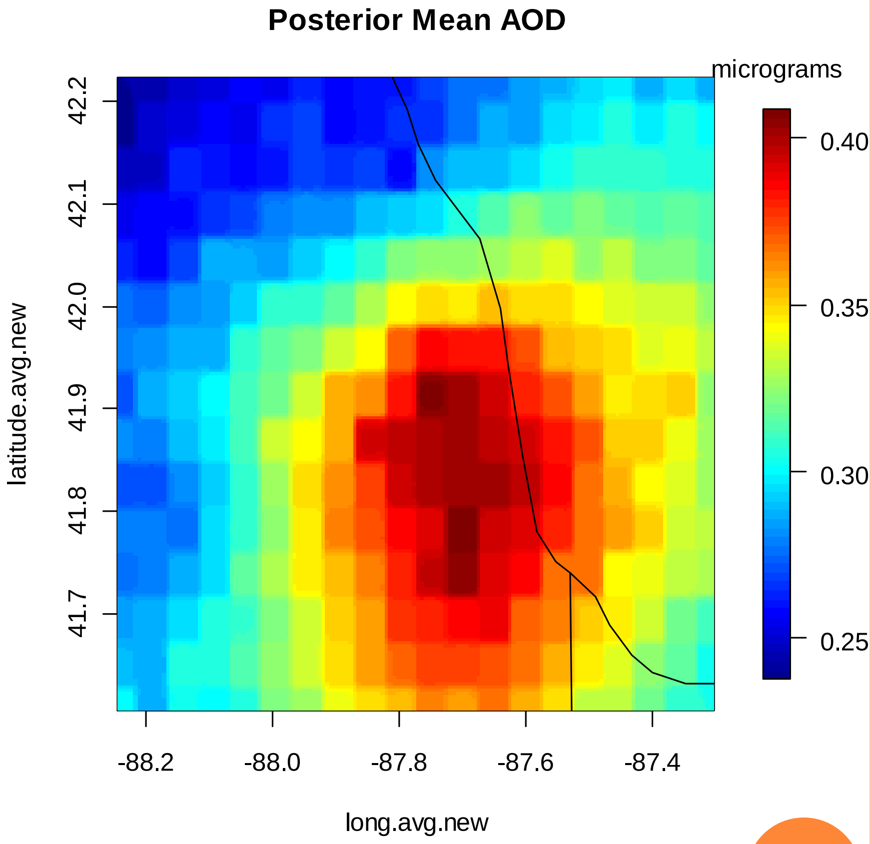
- library(ramps)

- control.pm25 <- ramps.control(iter = 1100,
- phi = param(NA, "uniform", min = 1, max = 60, tuning = 0.5),
- sigma2.z = param(NA, "invgamma", shape = 0.01, scale = 0.01),
- sigma2.e = param(NA, "invgamma", shape = 0.01, scale = 0.01),
- beta = param(rep(0, 1), "flat"),
- file = c("pm25params.txt", "z.txt"))

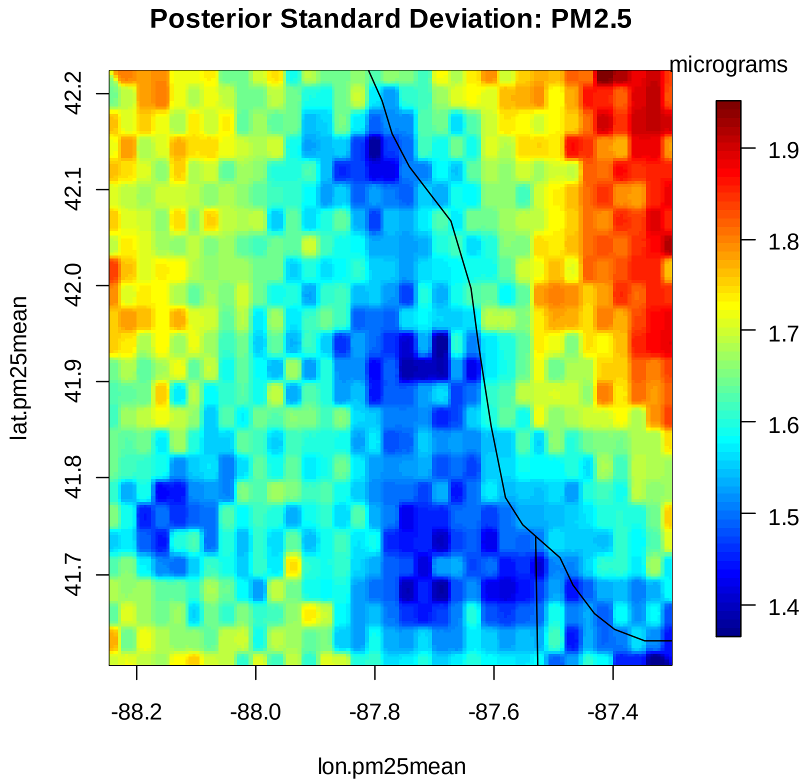
- fit.pm25 <- georamps(fixed = pm25mean.new ~ 1,
- correlation = corRExp(form = ~lon.pm25mean + lat.pm25mean, metric = "haversine"),
- control = control.pm25)

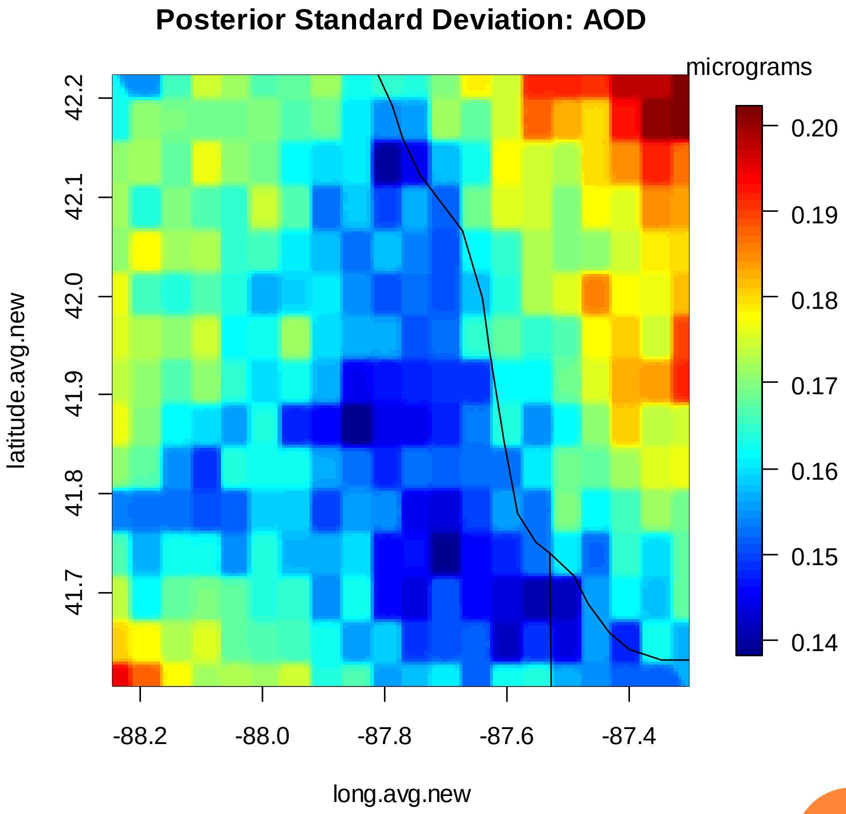
Posterior Mean: $PM_{2.5}$ and AOD





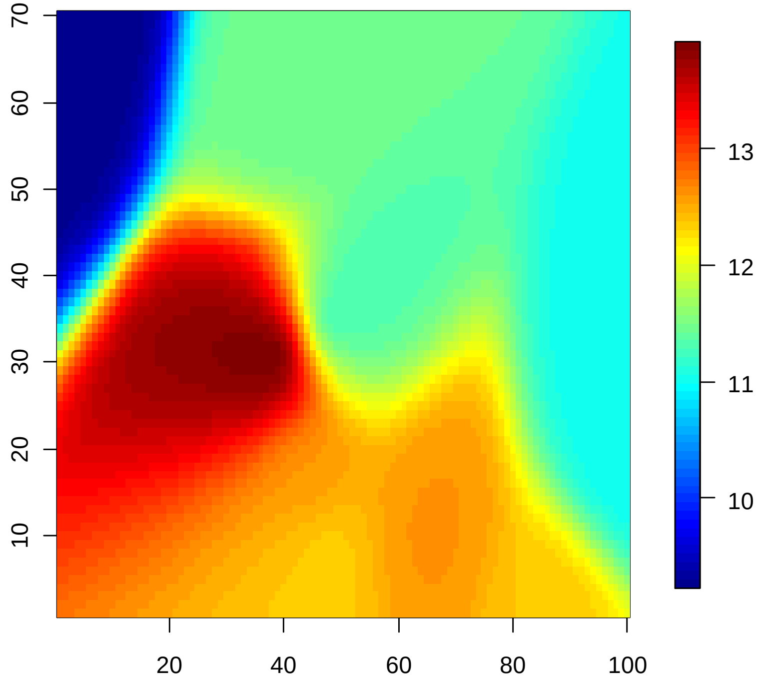
Posterior Standard Deviation: $PM_{2.5}$ and AOD



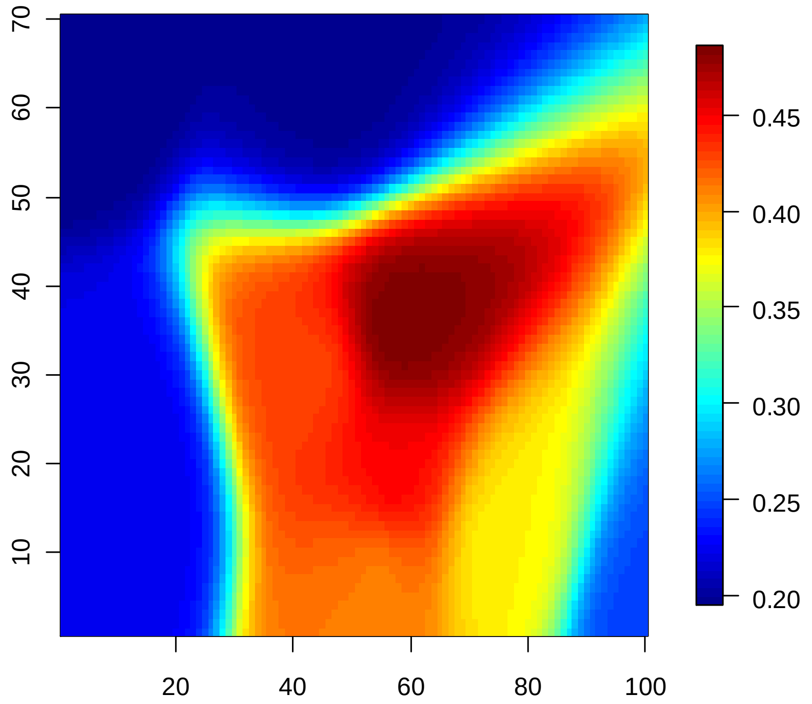


Raw Data

$PM_{2.5}$ Raw Data



AOD Raw Data



Conclusions

- Location is not the only important factor
- AOD itself is not appropriate for predicting health problems
- AOD is related to $PM_{2.5}$, which is appropriate for such predictions
- Bayesian statistical methods are used
- Simplified model

Challenges

- Large data sets
  - Computing time
- Spatial correlation
- Interpretation
  - Averages

Further Study

- Include land use as a predictor variable
- Include time as a predictor variable
  - Seasonal variability
  - Day to day changes
- Link $PM_{2.5}$ readings to health records in Chicago area
- Use AOD predictions to study health effects
- Pinpoint causes of $PM_{2.5}$ pollution

CITATIONS

- “Particulate Matter,” *EPA*. 25 February 2010. U.S. Environmental Protection Agency. 20 July 2010. http://www.epa.gov/air/particlepollution/index.html
- Smith, B. J., Yan, J., and Cowles, M. K. (2008) Unified Geostatistical Modeling for Data Fusion and Spatial Heteroskedasticity with R Package ramps, Journal of Statistical Software, 25(10), 1-21

Thanks

- Dr. Jacob Oleson, PhD
- Aaron Porter
- University of Iowa, Department of Biostatistics
- Environmental Protection Agency