Fusing surface and satellite-derived PM observations to determine the impact of international transport on coastal $PM_{2.5}$ concentrations in the western U.S.

Neha Bora Tuo Chen Dana Cochran Kelly Dougan Gautam Sabnis Chuanping Yu

Industrial Math/Stat Modeling Workshop
Environmental Protection Agency

May 8, 2017

Outline

Health Impacts

Figure: Donora, PA at noon on October 29, 1948

- 20 people died, thousands were sickened by smog from a steel mill.
- $PM_{2.5}$ one of the most harmful pollutants, can get lodged in lungs and cause respiratory problems

PM and AOD

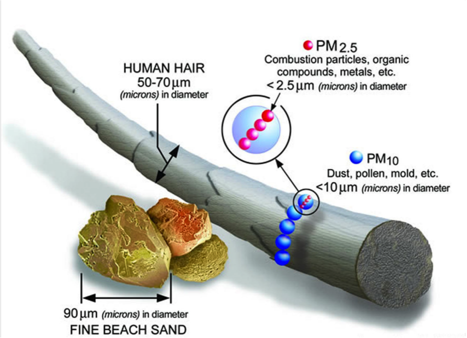


Figure: Size of $PM_{2.5}$ compared to other particulate matter

- $PM_{2.5}$ is a particulate matter that is less than 2.5 micrometers in diameter.
- Aerosol Optical Depth (AOD) measures the amount of light from the sun blocked by dust and pollutants.

Pollution in Western U.S.

- The Clean Air Act imposes a national standard of permissible amount of $PM_{2.5}$.
- Sites in California, such as in the Los Angeles area consistently violate this standard.
- Want to investigate the possibility of external impact by international air travel of $PM_{2.5}$.

Outline

1. Introduction
2. Data Sources and methods
3. Experiments
   - $PM_{2.5}$ time series
   - AOD and $PM_{2.5}$ fitting
4. Conclusions

$PM_{2.5}$

- Ground sites of $PM_{2.5}$ measurements, as collected by EPA
- Data collected either daily, every three days or every six days.

AOD

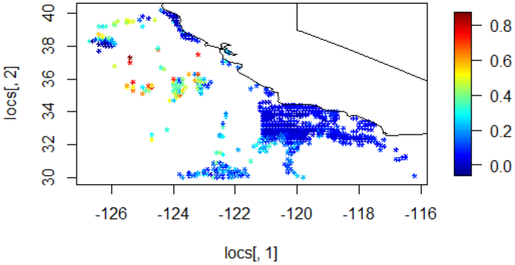


Figure: AOD readings on Jun 6, 2009

- AOD data collected over Pacific Ocean, near coast of California.
- Satellites travel around the Earth once every 16 days.
- AOD data is collected at each location least twice a month.

Data Cleaning

Figure: (left) CA $PM_{2.5}$ sites we compared with AOD data, (right) Hawaii $PM_{2.5}$ sites we compared with AOD data

- Picked PM sites closest to coast (13 in California, 9 in Hawaii)
- Found closest locations of AOD measurements to these sites
- Matched data by date of AOD/PM readings

Using additional data

Challenges

- AOD and $PM_{2.5}$ datasets are quite sparse in space-time.
- Need additional information to improve fits and predictions.

Covariates used in this analysis

- wind speed and direction
- humidity
- planetary boundary layer height
- air temperature
- added values of these measures at each location at each date

Overall big picture

Fuse AOD and $PM_{2.5}$ datasets to

- Understand their quantitative relationships,
- Predict PM measurements from years 1991 – 2000.

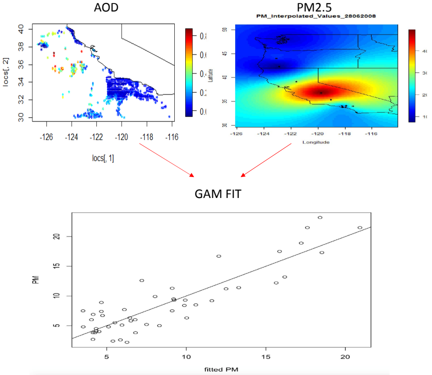
Overall Goal

Methods:

1. Time Series

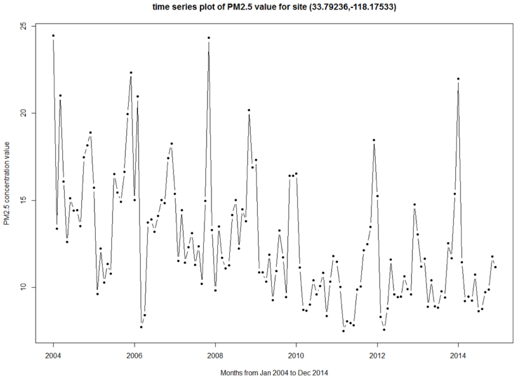
2. Spatial Interpolation

3. Multivariate Regression



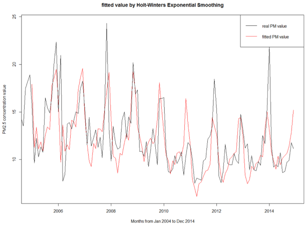
Outline

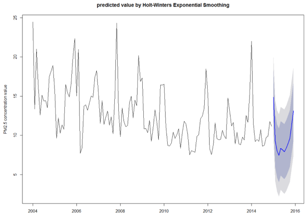
$PM_{2.5}$ Time Series



- Goal is to make predictions about $PM_{2.5}$ in 2015.
- Overall pollution appears to be decreasing.
- Seasonal variations are observed.
- Two methods used – Holt-Winter Exponential smoothing and ARIMA model

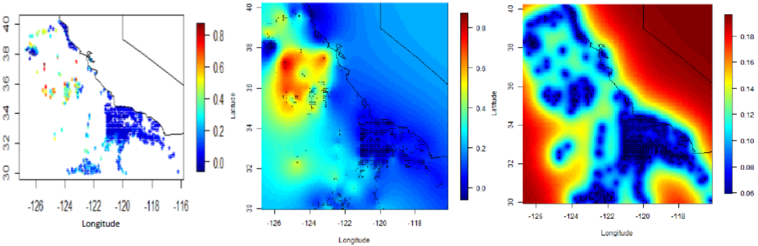
Holt-Winters Exponential Smoothing





- Mean square error is 5.5 but the mean of $PM_{2.5}$ values is 12.6
- Similar result observed using different method - ARIMA.
- $PM_{2.5}$ dataset itself not sufficient to make predictions
- This motivates using covariates and AOD data.

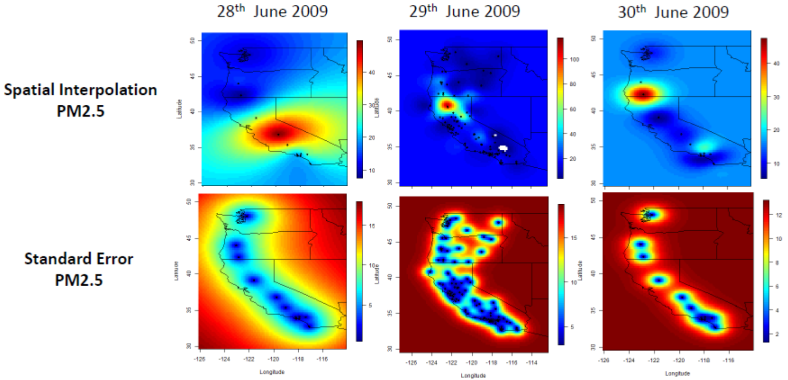
Spatial interpolation of AOD data



AOD values from AVHRR | Interpolated AOD values | Standard Error

- Interpolation using Ordinary Kriging.
- The quality of interpolation depends on availability of data.
- Understand the transport of pollutants to inform regression models.

Spatial interpolation of $PM_{2.5}$ data



Need further validation with satellite date to confirm plume tracking.

$PM_{2.5}$ vs AOD

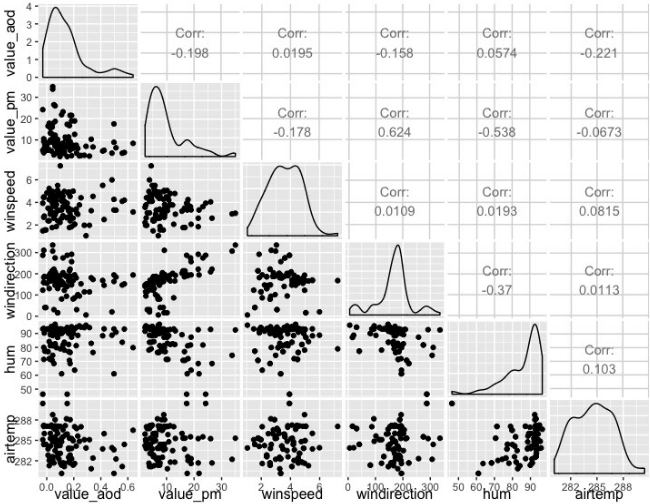
- Goal - relationship between $PM_{2.5}$ and AOD in presence of other meteorological variables.
- Response variable - $PM_{2.5}$ - "What we want to predict"
- Covariates - "What we use to predict"
  - continuous variables: wind direction, wind velocity, relative humidity, air temperature
  - factors: season – a discrete variable that models seasonality.
- Model the relationship at a particular $PM_{2.5}$ site or combine data from two nearby $PM_{2.5}$ sites along the west coast of United States.

Snapshot of the dataset

date	value_aod	value_pm	winspeed	windirection	hum	airtemp	season
2006-12-28	0.01794	17.8	4.25	193.17	71.9	283	4
2007-01-12	0.10983	21.8	3.33	194.59	61.1	281	4
2007-01-21	0.04382	17.7	3.83	190.70	81.1	283	4
2007-03-28	0.17270	7.2	4.53	173.31	76.3	283	1
2007-04-18	0.13014	3.6	4.97	122.07	78.0	282	1
2007-05-18	0.01794	5.2	2.38	130.32	93.2	283	1
2007-06-05	0.24137	2.5	3.78	95.29	95.4	285	2
2007-06-11	0.10983	5.1	4.74	163.40	92.0	285	2
2007-08-10	-0.02195	4.0	4.44	158.96	93.4	287	2
2007-09-21	0.06187	8.2	2.91	188.68	92.6	288	3

- $PM_{2.5}$ sites' coordinates near Eureka, CA - (40.8, -124) and (40.9, -123)
- This site was chosen since it provides best fits.

Preliminary Analysis I



- Shows nonlinear relationships between covariates and responses.
- skewed data, outliers.

Multiple Linear Regression Model

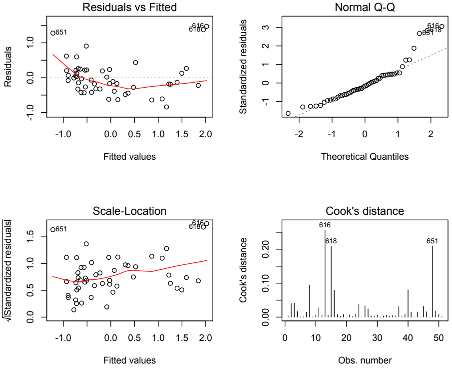
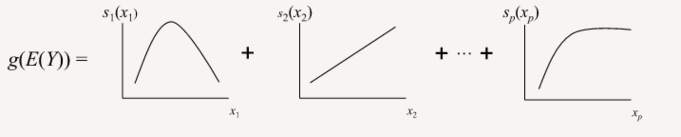


Figure: Some model diagnostics

Multiple linear regression does not provide good fits.

Generalized Additive Model



- Regression splines:

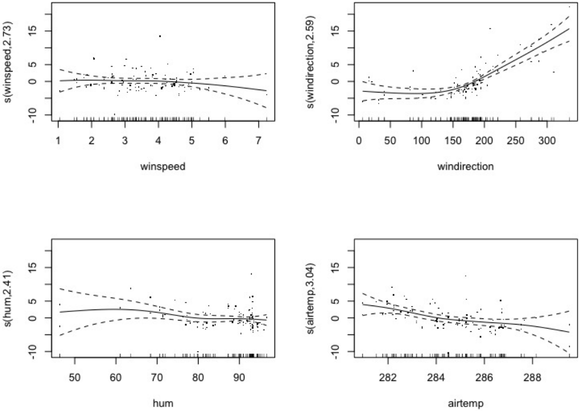
$$s_j(x) = \sum_{j,k} \beta_{jk}\phi_{jk}(x) = \boldsymbol{\beta}'\phi_j$$

- Estimate β by minimizing the penalized least squares objective function.

Results I

- Formula: $\mathrm{PM}_{2.5} \sim$ AOD + s(windspeed) $+$ s(winddirection) $+$ s(hum) $+$ s(airtemp) $+$ season
- Coefficient = 4.085, β_{AOD} = -2.283 (t = 3.84, p-value = 0.0023).
- R^2 = 0.782, Deviance = 81.4%
- We compared various models using a subset of covariates and chose the model with the smallest AIC (Akaike's Information Criterion).

Results II



- Equivalent degrees of freedom = (2.69, 2.54, 2.37, 3.00)

Outline

Conclusions and future work

Conclusions

- AOD has a linear relationship with the $PM_{2.5}$ measurements.
- Meteorological information is also important in predicting $PM_{2.5}$ (particularly wind speed).
- GAM model better than multivariate linear regression model better than mixed effects model.

Future work

- Perform analysis on Hawaii sites.
- Use data from more years to train our model.
- Performing spatio-temporal analysis.

Acknowledgments

- Faculty mentors
- EPA – Brett Gantt and Elizabeth Mannshardt
- NOAA, specifically Jessica Matthews for help with satellite data
- IMSM
- NSF

References

- Lee, H. J., et al. “A novel calibration approach of MODIS AOD data to predict PM2. 5 concentrations.” Atmos. Chem. Phys 11.15 (2011): 7991-8002.
- van Donkelaar, Aaron, et al. “Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter.” Diss. University of British Columbia, 2015.
- Li, Jing, Barbara E. Carlson, and Andrew A. Lacis. “How well do satellite AOD observations represent the spatial and temporal variability of PM 2.5 concentration for the United States?.” Atmospheric Environment 102 (2015): 260-273.
- Liu, Yang, Christopher J. Paciorek, and Petros Koutrakis. “Estimating Regional Spatial and Temporal Variability of $PM_{2.5}$ Concentrations Using Satellite Data, Meteorology, and Land Use Information.” Environmental health perspectives 117.6 (2009): 886.
- National Centers for Environmental Information. National Oceanic and Atmospheric Administration. Department of Commerce, n.d. Web. 23 July 2016. *https://www.ncdc.noaa.gov/cdr/atmospheric/avhrr-aerosol-optical-thickness*.
- United States Environmental Protection Agency. AirData. EPA, 5 July 2016. Web. 23 July 2016. *https://www3.epa.gov/airdata/*.