Kaipeng Zhong

Overview of AOD-PM2.5 Spatial Pattern in United State

Introduction

Particulate matter (PM) has increasingly become a hot research topic in the literature because of its significant role in environmental, climate and epidemiological studies. In environmental field, PM is classified based on the diameter of the particle such as PM10 and PM2.5. PM2.5 refers to particles with diameters smaller than 2.5 micrometers. (Cai et al., 2014) In this size, PM can penetrate into alveolus through the lungs and affect other organs. In climate science, PM is also named aerosol which impacts the Earth's climate by altering the radiation budget.(Li, Carlson, & Lacis, 2014)

Historically, PM concentration and speciation were studied spatially and temporally using ground-based observation. However, ground observation method has the drawback of limited spatial coverage and high operational cost that making holistic study of PM a difficult task. Today, satellite-based observation of aerosol has been widely used to estimate and predict ground-level PM by optical depth, which measures the fraction of radiation that is not scattered or absorbed on its path.

Because of its extensive spatial coverage, AOD-PM2.5 relationship has been well-studied in the literature from different parts of the world. (Chudnovsky et al., 2014; Guo et al., 2009; Lee, Coull, Bell, & Koutrakis, 2012; Li et al., 2014; Xu, Ceamanos, Roujean, Carrer, & Xue, 2014; You, Zang, Pan, Zhang, & Chen, 2015) The concentration of PM2.5 was estimated from satellite AODs obtained from MODIS, MISR, SeaWiFS, OMI and etc. (Xu et al., 2014) In general, empirical statistical correlations were developed from time series between AOD and PM2.5 in designated study sites.

However, the relationship between AOD and PM2.5 can vary in time and space. A good statistical model may not hold true in all study sites. This study aims to provide an overview on the degree of disagreement between AOD and PM2.5 in the US Continent using the best-fitted regression models.

The AOD and PM2.5 were fitted using both linear and nonlinear methods. The Gaussian Generalized Linear Model (GLM) assumes the expected value of Y has a linear form of

$$E(Y) = f(X_1, ... X_p) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p$$

where Y is a response random variable and $X_1, ... , X_p$ is a set of predictor variables. Normal distribution is assumed for the response variable and the only predictor variable is AOD. Given a sample of values for Y and X, estimates of $\beta_0, \beta_1, ..., \beta_p$ are usually obtained by the least squares method.

Generalized additive models (GAM) extend traditional linear models in another way, namely by allowing for a link between $f(X_1, ... , X_p)$ and the expected value of Y. The additive model generalizes the linear model by modeling the expected value of Y as

$$E(Y) = f(X_1, ... X_p) = s_0 + s_1(X_1) + ... + s_p(X_p)$$

where $s_i(X)$, $i = 1, ... , p$ are smooth functions. These functions are estimated in a nonparametric fashion. A combination of backfitting and local scoring algorithms are used in the actual fitting of the model. (James, Witten, Hastie, & Tibshirani, 2013)

GAM and GLM can be applied in similar situations, but they serve different analytic purposes. Generalized linear models emphasize estimation and inference for the parameters of the model, while generalized additive models focus on exploring data nonparametrically.

Kaipeng Zhong

Method

The annual mean of AOD-550nm in the US Continent used in this study was obtained from monthly product of Satellite MODIS-Terra in 2013. The ground-based PM2.5 data for the same year was obtained from USEPA AirData. To ensure the purity of the data, only Gravimetric PM2.5 measured with Met One SASS Teflon and IMPROVE Module A with Cyclone Inlet-Teflon Filter were used.

The output area of AOD data was set to 23.5N - 54.5N and 129.5W - 60.5W, which covers most of the PM2.5 ground monitoring stations in the US Continent. The spatial resolution of MODIS-Terra is $1°$ x $1°$.

In order to match ground observation to satellite observation, grid-averaging of ground observation data in accordance to satellite grids was performed using statistical packages in R. Grids with no ground observation were omitted. The AOD and PM2.5 pairs were fitted using GLM and GAM. GLM was fitted first and RMSE, R Squared were calculated from that fit. For GAM, smooth spline function with different degrees of freedom was used to produce the best fit. Model training of GAM was accomplished in R based on in-sample Root Mean Square Error (RMSE), which is calculated by the following:

$$RMSE = \sqrt{(obs - pred)^2}$$

Model with the smallest value of RMSE and larger value of R Squared indicates good fit. The goodness-of-fit was also evaluated graphically using residual plot.The residuals from the best-fitted model was finally plotted on the map of US Continent.

Kaipeng Zhong

Result and Discussion

   The PM monitoring sites were reduced to 344 by grid-averaging and each grid represents the
average concentration of monitoring sites in that grid. The tentative fit of GLM shows poor
goodness-of-fit with R-squared of 0.009544 and the p-value from F-statistic between null model
and fitted model is 0.07035 indicating the current model is no better than a null model. The
residual plot shows biased trend.
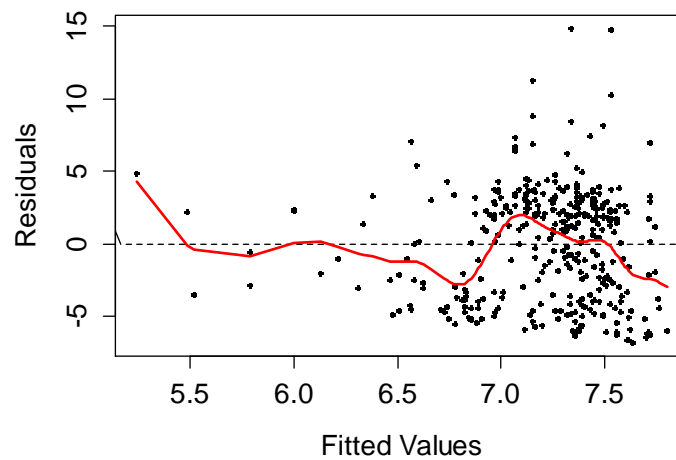
   The nonlinear model (GAM) however, produced a good fit of the data with smooth spline of
7 degrees of freedom. The p-value from Chi-square Test between null model and current model
is nearly 0, indicating the current model fits very well. The residual plot also shows steady trend
of residuals.
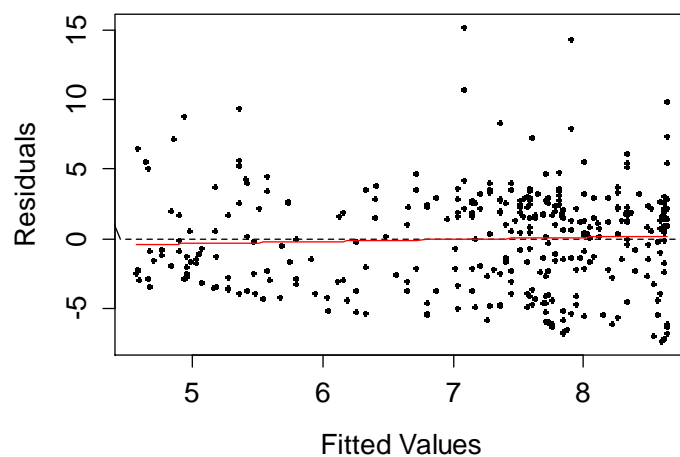
The spline smoothing function plot indicates that most of the AOD values fell between the range of $0.04 - 0.15$, where the 95% confident interval is the narrowest.
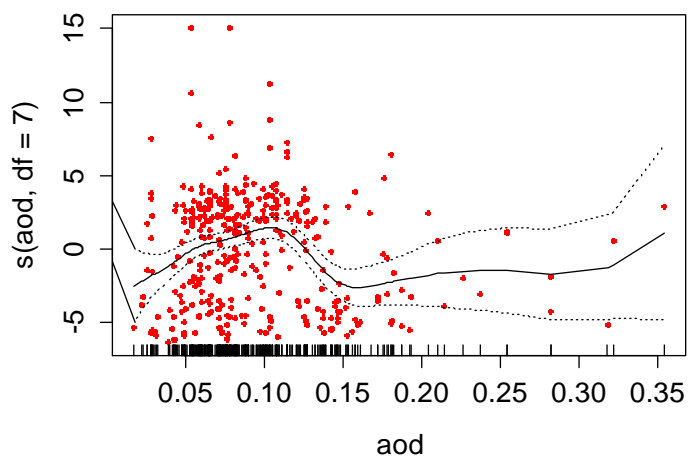
The residuals and absolute value of residual from the GAM model were plotted on the map of United State.

Kaipeng Zhong
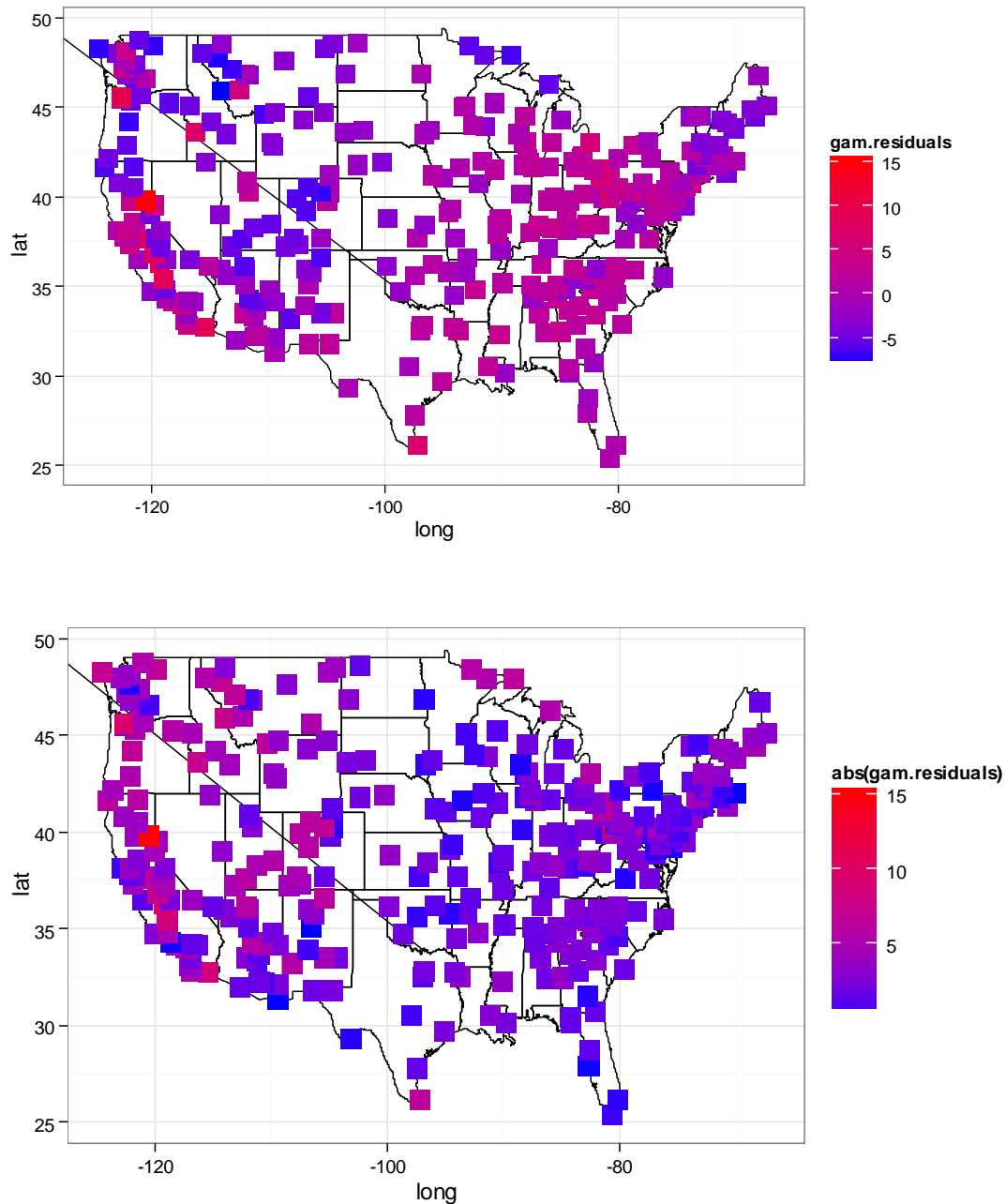
Since the residual is calculated by subtacting the observed value by the model predicted

value, the residual map of the GAM model shows that there are more overprediction occurred in

the West United State than the East United State. Also, prediction error is more likely to appear

in West United State than East United State. In coastal region near to Pacific Ocean and

boundary between US and Canada, the model seems to more susceptible to error as comapred to inland area.

Conclusion

   This study summarized the spatial pattern of residual obtained from empirical statistical models. The result indicates some degree of spatial dependence on prediction errors. In general, West United State is more likely to be affected by error. One possible reason is that the AOD-550nm does not work well in snow-covered and mountainous regions in States such as Idaho and Colorado. Deep Blue AOD in 550nm can be used to address bright surface but the availability of data is lower than general AOD-550nm. This study can go further by performing spatial statistics on the residual map to quantify the spatial dependence of the model residuals to better understand the pattern. Also, one can apply spatial krigging to enrich the gridded data rather than grid-averaging.

Annual Average of PM2.5 in 2013

Kaipeng Zhong
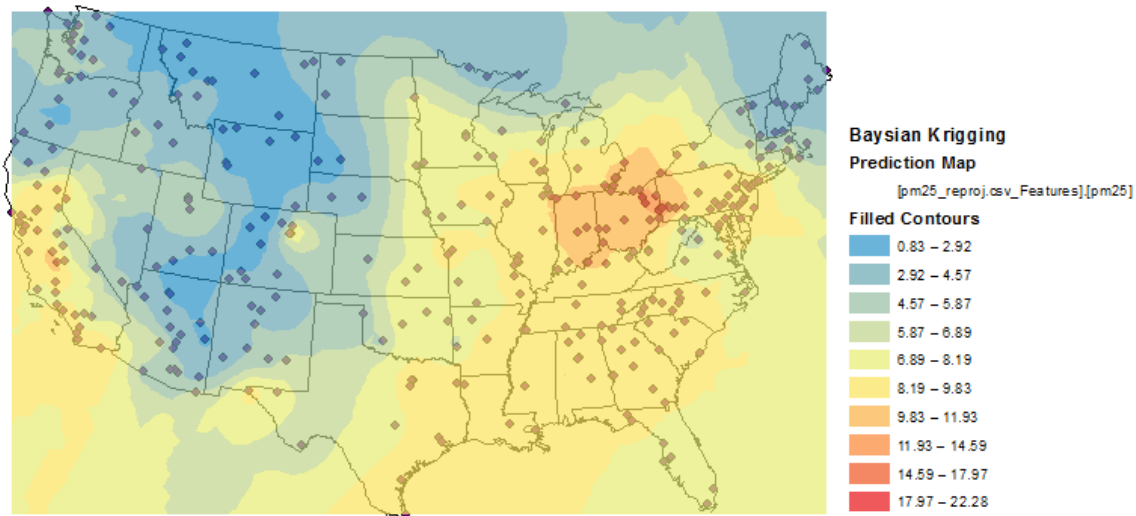
Reference

Cai, J., Zhao, A., Zhao, J., Chen, R., Wang, W., Ha, S., … Kan, H. (2014). Acute effects of air pollution on asthma hospitalization in Shanghai, China. *Environmental Pollution (Barking, Essex : 1987)*, *191*, 139–44. doi:10.1016/j.envpol.2014.04.028

Chudnovsky, A. A., Koutrakis, P., Kloog, I., Melly, S., Nordio, F., Lyapustin, A., … Schwartz, J. (2014). Fine particulate matter predictions using high resolution Aerosol Optical Depth (AOD) retrievals. *Atmospheric Environment*, *89*, 189–198. doi:10.1016/j.atmosenv.2014.02.019

Guo, J.-P., Zhang, X.-Y., Che, H.-Z., Gong, S.-L., An, X., Cao, C.-X., … Li, X.-W. (2009). Correlation between PM concentrations and aerosol optical depth in eastern China. *Atmospheric Environment*, *43*(37), 5876–5886. doi:10.1016/j.atmosenv.2009.08.026

Lee, H. J., Coull, B. A., Bell, M. L., & Koutrakis, P. (2012). Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM2.5 concentrations. *Environmental Research*, *118*, 8–15. doi:10.1016/j.envres.2012.06.011

Li, J., Carlson, B. E., & Lacis, A. A. (2014). How Well Do Satellite AOD Observations Represent the Spatial and Temporal Variability of PM2.5 Concentration for the United States? *Atmospheric Environment*. doi:10.1016/j.atmosenv.2014.12.010

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R.* New York: Springer.

Xu, H., Ceamanos, X., Roujean, J.-L., Carrer, D., & Xue, Y. (2014). Can satellite-derived aerosol optical depth quantify the surface aerosol radiative forcing? *Atmospheric Research*, *150*, 151–167. doi:10.1016/j.atmosres.2014.07.008

You, W., Zang, Z., Pan, X., Zhang, L., & Chen, D. (2015). Estimating PM2.5 in Xi'an, China using aerosol optical depth: A comparison between the MODIS and MISR retrieval models. *The Science of the Total Environment*, *505*, 1156–65. doi:10.1016/j.scitotenv.2014.11.024