1 Introduction

2 The Problem

3 The Approach

3.1 Identify U.S. $PM_{2.5}$ sites within or adjacent to AVHRR grids

As AVHRR satellites can only get the aerosol optical depth (AOD) from oceans and PM2.5 data are from lands, our strategy for identifying PM2.5 sites is to choose the ones that are as close to the coast as possible. Figure 1 shows all the PM2.5 sites and Figure 2 shows the sites that we chose. We chose 13 PM2.5 sites closest to the west coast.

Figure 1: All the PM2.5 sites

3.2 Generate relationships between AVHRR AOD and surface $PM_{2.5}$

We noticed that there is huge differences between the sites we identified in the last step and the sites located on Hawaii. The sites we identified in the last step are all close to the west coast, while the sites located on Hawaii are far away from lands. So we decided to build models separately for both the sites we chose and the sites on Hawaii.

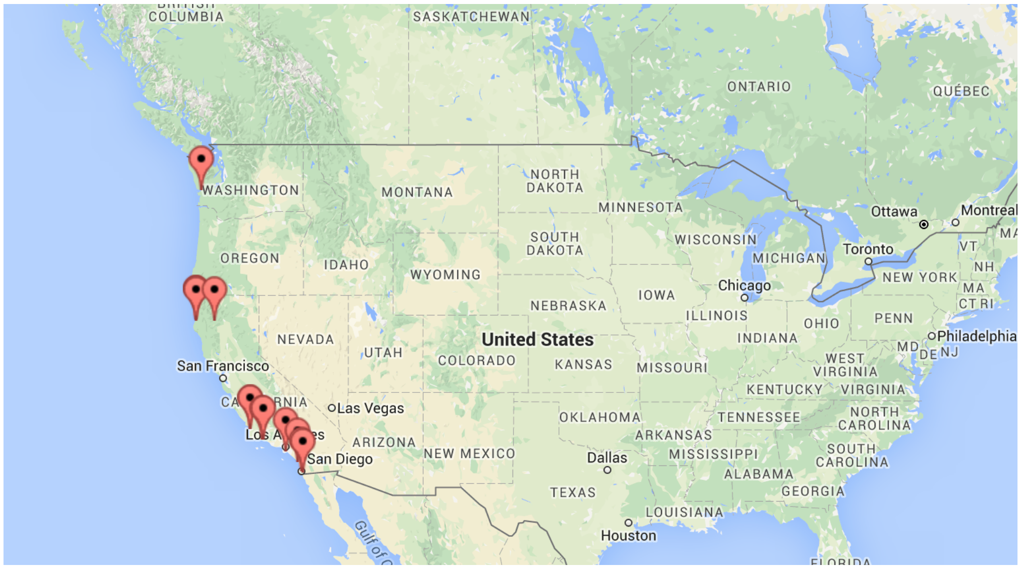

Figure 2: $PM_{2.5}$ sites chosen

3.2.1 Clean the AVHRR AOD data

No matter what $PM_{2.5}$ sites we use, we need to find the closest AOD coordinates to the $PM_{2.5}$ sites, and then to match the sites data for each day. This requires us to search all the observations in AOD dataset. The AOD data are very large and we cannot put all the data in hundreds of files into one matrix. Besides, AOD data have lots of missing data. So we need to clean the AOD data first.

Since the data we want to use are the data located on the west coast and Hawaii, first we used the longitude and latitude of the west coast and Hawaii to eliminate data from other locations that we won't use. Then, we dropped all the missing data and also changed the original format of the data into more readable one. Finally, our AOD data looks like Table 1.

Date	lat_aot	long_aot	aot
2000-01-25	30	-126.6	-0.0836133733391762
...	...	...	...

Table 1: AOD data

3.2.2 Match the AOD data with the $PM_{2.5}$ data

As we are trying to find the relationships between AOD and PM2.5, we need to keep all the date the same, and locations closest. After matching these two datasets, we got the following as in Table table3.2.

Date	lat_pm	long_pm	pm	lat_aot	long_aot	aot
2006-12-28	40.776944	-124.1775	17.8	40.9	-124.3	0.043815478682518
...	...	...	...	...	...	...

Table 2: AOD and PM2.5 match data

3.2.3 Relationships between AVHRR AOD and surface PM2.5 for sites closest to the west coast

For sites closest to the west coast, we finally found 4 sites that have common date information. Figure 3 is the plot regarding the AOD and PM2.5 by each site.

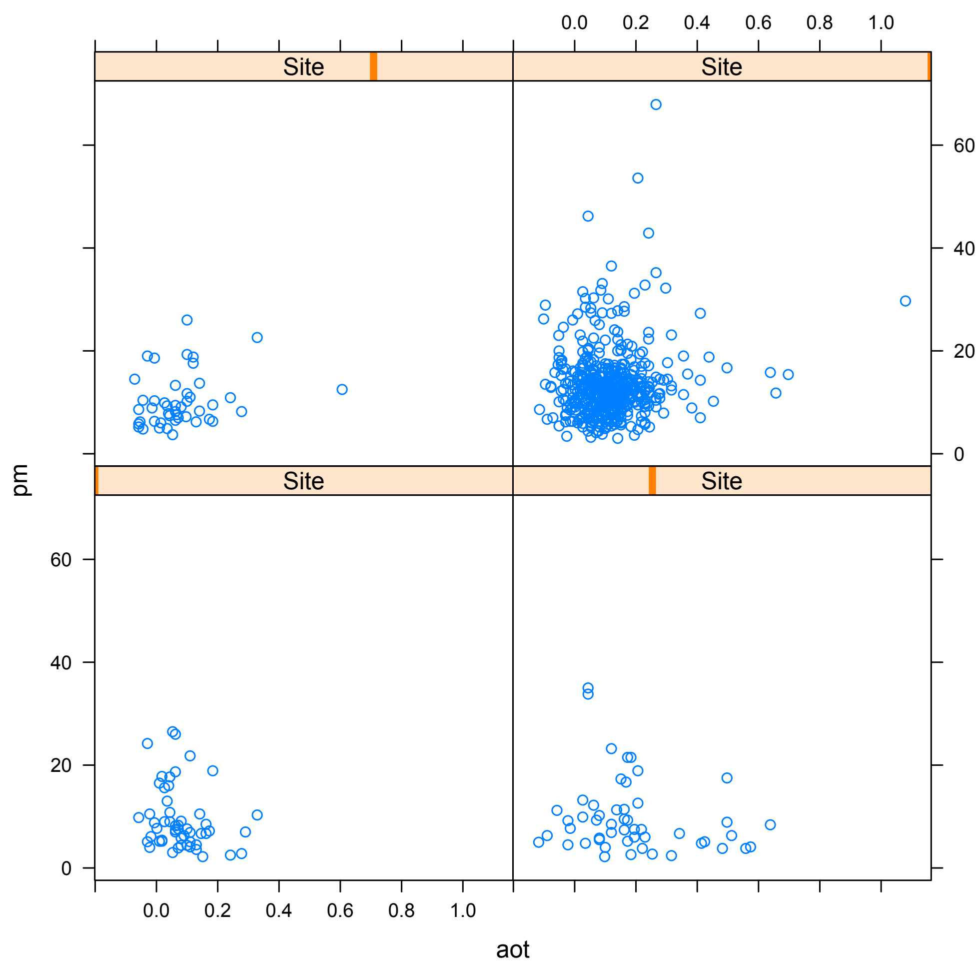

Figure 3: PM vs. AOD for each site close to the west coast

We can see from the above plot that the relationships between AOD and PM2.5 are different for different sites. So we consider treat sites as a factor variable to make sure there would be different models for different sites.

Inspired by the analysis in Lee et al., we built a mixed effects model to fit this relationship. Since there are lots of parameters varying each day, like humidity, clouds, etc., the *date* variable is treated as a random variable. Also, as we mentioned above, the *site* variable is treated as random too because of their different properties, like longitude, latitude, altitude, climate, etc. The model is as following:

$$PM_{ij} = \alpha + \beta \times AOD_{ij} + s_i + d_j + \epsilon_{ij},$$

where PM_{ij} is the $PM_{2.5}$ concentration at a spatial site i on a specific day j, α is the fixed intercept, β is the fixed slope, AOD_{ij} is the AOD value at a spatial site i on a specific day j, $s_i \sim N(0, \sigma_s^2)$ is the random intercept of site i, $d_j \sim N(0, \sigma_d^2)$ is the random intercept of a specific day j, and $\epsilon_{ij} \sim N(0, \sigma^2)$ is the error term at site i on a day j.

3.2.4 Relationships between AVHRR AOD and surface PM2.5 for Hawaii sites

For Hawaii sites, we finally found 9 sites that have common date information. Figure 4 is the plot regarding the AOD and PM2.5 by each site.

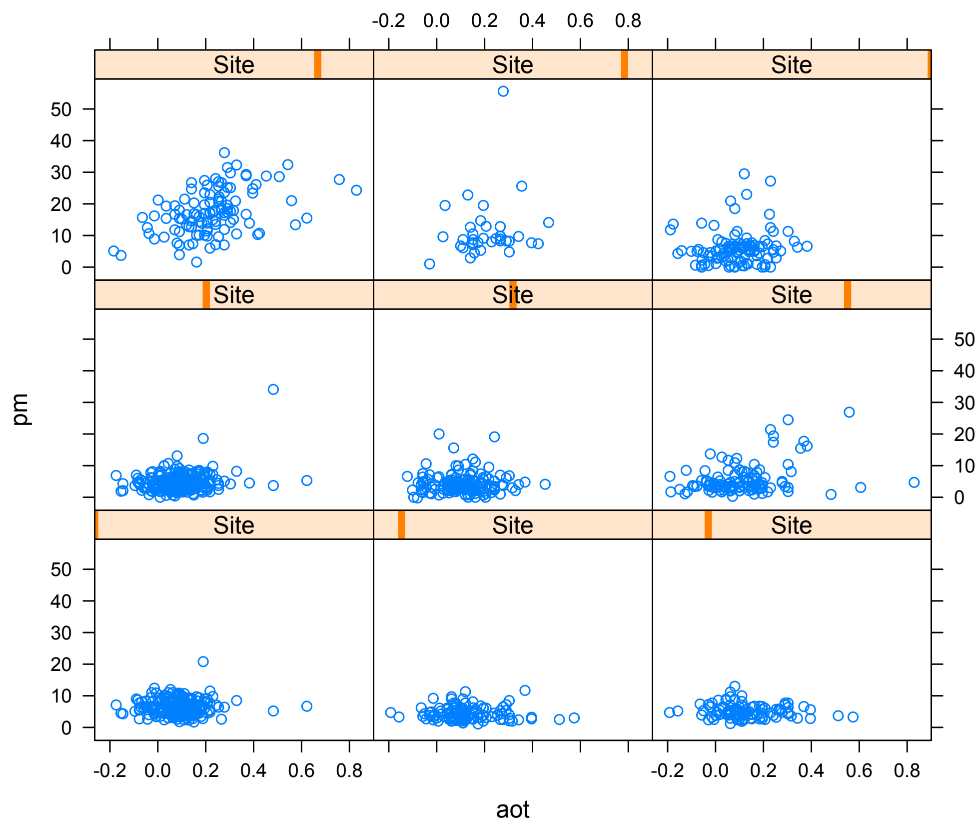

Figure 4: PM vs. AOD for each Hawaii site

Similar to what we have analyzed for the west coast, we used the same model structure to fit the relationship:

$$PM_{ij} = \alpha + \beta \times AOD_{ij} + s_i + d_j + \epsilon_{ij},$$

where PM_{ij} is the $PM_{2.5}$ concentration at a spatial site i on a specific day j, α is the fixed intercept, β is the fixed slope, AOD_{ij} is the AOD value at a spatial site i on a specific day j, $s_i \sim N(0, \sigma_s^2)$ is the random intercept of site i, $d_j \sim N(0, \sigma_d^2)$ is the random intercept of a specific day j, and $\epsilon_{ij} \sim N(0, \sigma^2)$ is the error term at site i on a day j.

The only difference between this model and the last model is that we used different datasets.

4 Computational Experiments

4.1 Relationships between AVHRR AOD and surface PM2.5

4.1.1 Sites closest to the west coast

Using the approach in the section 3.2.3, we fitted the mixed effects model as follows.

$$\hat{PM}_{ij} = 10.51 + 3.60 \times AOD_{ij} + \hat{s}_i + \hat{d}_j,$$

where $\hat{s}_i \sim N(0, 1.79^2)$ and $\hat{d}_j \sim N(0, 3.04^2)$.

The correlation between the fitted $PM_{2.5}$ data and the true $PM_{2.5}$ data is 0.802, which we agreed it is not a bad fit.

4.1.2 Hawaii sites

Similarly, we can get the mixed effects model for Hawaii sites.

$$\hat{PM}_{ij} = 6.205 + 7.780 \times AOD_{ij} + \hat{s}_i + \hat{d}_j,$$

where $\hat{s}_i \sim N(0, 4.05^2)$ and $\hat{d}_j \sim N(0, 2.83^2)$.

The correlation between the fitted $PM_{2.5}$ data and the true $PM_{2.5}$ data is 0.879, which is better than the west coast data.