Final Project Proposal

Beijing and a substantial part of China are experiencing chronic air pollution. The main pollutants are fine particulate matter, and PM2.5 in particular. PM2.5 consists of airborne particles with aerodynamic diameters of less than 2.5 μm. They are known to influence visibility, human health and even climate. Epidemiological evidence shows that exposure to PM2.5 can cause lung morbidity, serious respiratory and cardiovascular diseases, and even death.

Given that there are 22 million inhabitants in Beijing, and 300 million immediately to the south in the North China Plain (NCP), it is vital to measure the severity of the PM2.5 pollution in Beijing. This is particularly relevant since China's State Council has a target of reducing PM2.5 by at least 25% from the 2012 level by 2017 for Beijing.

For this project, I decide to divide it into three parts. For the first part, by learning the PM2.5 readings and meteorological records from 2010–2015, the severity of PM2.5 pollution in Beijing can be visualized by a set of statistical graphs. For the second part, I decide to use the linear regression to research the correlation between PM2.5 and some meteorological factors like temperature, pressure, wind speed and wind direction. For the last part, I decide to use the time series to forest the future PM 2.5 value. The result can help us to test if the PM2.5 levels in Beijing have been lowered since China’s State Council set up a pollution reduction target.

Our analysis uses hourly PM2.5 readings taken at the US Embassy in Beijing located at (116.47 E, 39.95 N), in conjunction with hourly meteorological measurements at Beijing Capital International Airport (BCIA), obtained from weather.nocrew.org. Both data series run from 1 January 2010 to 31 December 2014. Although the embassy and the airport are 17 km apart, they experience very much the same weather.

Beijing PM2.5 Data Set

Abstract: This hourly data set contains the PM2.5 data of US Embassy in Beijing. Meanwhile, meteorological data from Beijing Capital International Airport are also included

Source:

Song Xi Chen, csx@gsm.pku.edu.cn, Guanghua School of Management, Center for Statistical Science, Peking University.

Data Set Information:

The data set time period is between Jan 1st, 2010 to Dec 31st, 2014. Missing data are denoted as NA

Attribute Information:

No: row number

year: year of data in this row

month: month of data in this row

day: day of data in this row

hour: hour of data in this row

pm2.5: PM2.5 concentration (ug/m^3)

DEWP: Dew Point (℉)

TEMP: Temperature (℉)

PRES: Pressure (hPa)

cbwd: Combined wind direction

Iws: Cumulated wind speed (m/s)

Is: Cumulated hours of snow

Ir: Cumulated hours of rain

Data Set Link

https://www.kaggle.com/djhavera/beijing-pm25-data-data-set