Geophysical Research Abstracts
Vol. 20, EGU2018-5077, 2018
EGU General Assembly 2018

Estimating PM2.5 concentration based on random forest regression method with remote sensing and monitoring station data

Qi Shao, Yunfei Zhang, Jing Li, and Yunhao Chen

Beijing Normal University, faculty of geographical science, school of geography, China (zhangyunfei_gis@yahoo.com)

With the acceleration of economic development and urbanization, haze phenomenon in cities happens frequently and the haze influencing area is constantly expanding, showing the characteristics of long duration and regionalization . Not only Haze reduces atmospheric visibility, but also harms people's physical and mental health. Beijing, the capital of China, has suffered the haze many times. PM2.5 is the major pollution component of haze. Therefore, it is of great significance to retrieve PM2.5 concentration in Beijing and analyze its temporal and spatial distribution. The traditional method of monitoring PM2.5 is to set up many ground monitoring stations, which can obtain accurate and real-time data. The defect is that the monitoring points are discrete and sparse and the spatial distribution of PM2.5 data cannot be obtained continuously. However, satellite remote sensing technique has the advantages of wide coverage, continuous observation and accurate data acquisition, which can be used to retrieve the spatially continuous PM2.5 concentration. In addition, the formation, transformation and sedimentation of PM2.5 are complicated and changeable, having obvious spatial heterogeneity and nonlinear characteristics. The traditional multiply linear regression model is not suitable. To overcome the defects of the traditional method, we use random forest regression, based on MODIS 3km aerosol optical thickness product, ECMWF-ERA5 meteorological reanalysis data and the PM2.5 concentration data of the pollutant monitoring station, to establish the model to estimate the near-surface PM2.5 concentration. The model considers the nonlinear characteristic and utilizes both monitoring station and remote sensing data to obtain the spatially continuous PM2.5 concentration. The results have the complete spatial coverage. Then we analyze the temporal and spatial variation of PM2.5 concentration in Beijing. The results demonstrate the potential of remote sensing data and random forest regression method to predict PM2.5 concentration.