Investigating the performance of satellite-based models in estimating the surface $PM_{2.5}$ over China

*Siwei Li[1], Lechao Dong[1], Jie Yang[1], Wenxuan Shi[1]

1. Wuhan University

Accurate estimation of atmospheric particulate matter with particle diameter less than 2.5 μm ($PM_{2.5}$) concentration is important because of its adverse impacts on public health. Due to the limited spatial coverage of ground monitoring stations, satellite products are often used to estimate surface $PM_{2.5}$ concentration by constructing a comprehensive relationship between satellite-retrieved aerosol optical depth (AOD) and ground-based in-situ measured $PM_{2.5}$ concentration. The accuracy of the satellite-based models in estimating $PM_{2.5}$ concentration is critical for its usage in health assessment, however, the performance of satellite-based models has not been well evaluated yet. In this study, we propose a new method to validate and evaluate the accuracy and performance of two popular machine learning models Random Forest (RF) and the BP Neural Network (BPNN) in estimating $PM_{2.5}$ concentration over China. The two models were trained and tested based on a whole year of satellite-retrieved hourly AOD from HIMAWARI, ground-based in-situ measured $PM_{2.5}$ from China National Environmental Monitoring Center, ERA5 meteorological conditions and other auxiliary variables in 2017. It is found that the traditional validation method may overestimate the performance of the models in estimating the $PM_{2.5}$ at the area with sparse in-situ measurements. Moreover, the spatial distribution of the training data will largely affect the performance of the models. A variable parameter retrieval algorithm based on RF model shows a better overall performance than the fixed parameter RF model. The R^2 is greater than 0.83 when the distance of the test station to the nearest training station is smaller than 100 km, but reduces to 0.51 when the distance is larger than 500 km . Such results suggest that the performance of satellite-based model in estimating the surface $PM_{2.5}$ need carefully examined, and still exhibits considerable uncertainties to be further improved.

Keywords: PM2.5, remote sensing, HIMAWARI-8, AOD, machine learning

