Review of Xue et al., Satellite-based Estimation of the Impacts of Summertime Wildfires on Particulate Matter Air Quality in United States

This study uses the GWR method to predict surface PM2.5 concentrations in the US based on satellite AOD and meteorological variables. The statistic method is robust and is well referenced from previous studies, and the prediction results show good agreement with the in-situ measurements. However, the study is still lack of adequate scientific expansion from the results, and the conclusions are similar to the studies on satellite AOD products or ground measurements only, making this study less meaningful.

Before the consideration of publishing, the authors need to further explore the prediction results, make good findings or quantifications that simple AOD and scattering ground measurements cannot show. The authors also need to clean up the minor typos, formats, the potential figure-caption disagreements and misleading journal names in the page head.

The major difference between the science of AOD and surface PM2.5 is that, AOD is showing the vertical column conditions instead of surface only. Even not considering the aerosol chemistry and secondary formation in clouds, the convection conditions, atmospheric stability or vertical profiles of other meteorological conditions should contribute very much to the difference of AOD and surface PM. Especially for fire plumes, the long-term transport of fire smoke can be at a high altitude, and the vertical pattern of PM will be very different from the no-fire patterns. However, in the GWR model used in this study, only near-surface data are used. Also, noticing the AOD coefficient is much higher that all the other predictors (Table 4). It is doubtful how good the model is, compared to the agreement between AOD and surface PM. Therefore, the authors need to:

1. Show the improvement of the model from using AOD as the only factor, and discuss how the model predict the surface PM out of a column variable.
2. Estimate the model performance only looking at fire region, compare to Figure 3, and discuss the performance and potential bias.

Except for the main concerns, there are some minor suggestions and questions listed below:

1. Line 141-143: Since all the regions in the US are evaluated (in Figure 6), FRP in the other regions should also be verified, to make sure the 2011-2018 difference over the regions other than NW US is not affected by regional fire.
2. Line 269-270: as discussed in 1.2, further discussions e.g. calculation R^2 for high PM values may be useful.

3. Line 274-278: Is there any logic about the box selecting? For example, how to decide the size of the box? Can the box be larger? For each type of region, the authors can also show a regional mean with standard deviations of each coefficient. Also, can the box/region selection be more quantified, for example, by classifying using the background PM concentrations or FRP?
4. Line 612-614: The caption of Figure 4 seems not agree with the figure it self. "*PM2.5 values equal or larger than 30 $\mu g\ m$-3 are shown as the same color (red)*…", but the color label is ~-5 to 60 ug/m3.