Response to review

In the manuscript, the authors compared the satellite-derived PM2.5 in two different periods to see the impacts of wildfires on air quality in the US. Although the study presented some valuable results, it is relatively simple which lacks in-depth analysis, and the scientific innovation is not clear. In addition, I am mainly concerned about the used method for PM2.5 retrievals, and also a lot of important information is missing.

We thank the reviewer for the helpful comments. The suggestions given by the reviewer helped to clarify our arguments and to improve the quality of the paper. We have made significant revisions to the paper based on comments from all reviewers.

Below are my specific comments: Line 54-70: The authors should carefully summarize the methods of PM2.5 estimations according to different categories, and the cited reference is too old and need to be updated by adding more recent studies.

As suggested by the reviewer, we added more recent references in the introduction section.

Line 86: The authors need clearly clarify the novelty of the study and the difference with previous related studies.

This study is novel in: 1) applying PM2.5 estimation methods on wildfire events and calculate the prediction error at high pollution concentration condition; 2) analyzing predictors' different influences in estimating PM2.5 under various conditions; 3) quantify the air pollution from fires by states and EPA regions.

Line 107: What's the accuracy of MAIAC AOD products in your study region? I suggest adding a preliminary validation by comparing the AERONET groundbased measurements.

We have revised the reference for MAIAC AOD performance. Over North America, MAIAC AOD has a very small bias of -0.01 compared to AERONET AOD (Superczynski et al., 2017). The typical error is usually around ±0.05 during times of high aerosol loading, and the bias slightly increases as AOD increases.

Line 109-110: How do the authors deal with such a big cloud missing situation in such a short study period in summer? In this way, ground-based observations could be more suitable than satellite retrievals due to a large number of missing data. In addition, cloud and smoke are difficult to be distinguished during the AOD retrieval, resulting in the smoke areas are often masked as clouds?

Cloud contamination is indeed an important limitation on estimating surface PM2.5 using satellite data, and that is also the reason for performing GWR in an aggregate sense in this study.

By aggregating satellite data for 17-days, we are able to predict PM2.5 using a reasonable amount of satellite observations. We also added one paragraph in model uncertainties and limitations section to provide one possible solution for these problems (both missing AOD data due to clouds and smoke misidentification). We plan to use chemistry transport models to fill in all the AOD gaps in the future work.

Line 117: Why not use the ERA5-Land meteorological data at a finer resolution of 0.1 degrees?

We agree with the reviewer that finer resolution of ERA-5 would be better for analysis, but ERA5-Land meteorological data does not have boundary layer height, which is very important for assessing surface PM2.5.

Line 146: 0.1◦ or 0.01◦? MAIAC AOD is 1 km.

The surface PM2.5 data we generate is 0.1-degree resolution due to the coarse resolution of meteorological datasets.

Section 3.3: The reviewer doesn't know why the authors choose the GWR model since there are many existed more accurate statistical regression (e.g., GTWR) or machine learning (e.g., random forest) models that have been proved in previous studies. The author should clearly clarify this.

We agree that other statistical regression especially GTWR would improve the accuracy on PM2.5 predictions. However, the limitation on AOD data (missing data due to cloud and other gaps) is one problem for conducting the model since it needs more training data. This is also the reason for conducting this analysis on a 17-day aggregate sense.

Line 177: What is the LOOCV method and how does it work?

LOOCV is actually an extreme version of k-fold cross validation, and it requires maximum computational cost because, for each data in the dataset, we will create one model and evaluate this data. Therefore, the results are reliable and unbiased though computationally expansive.

Table 2: Should be improved (a line or bar chart might be better), in addition, state abbreviations are hard to read. The result analysis is very simple, which seems like an article about the PM2.5 retrieval algorithm. More in-depth analysis of the impacts of wildfires on air quality is needed.

Thank you for pointing this out. We changed the state abbreviations to full names and added a plot to show only some most influenced states. We also add more analysis in the discussion section.