Reviewer 2

This study combined satellite observations, model simulations, and ground measurements to assess the impact of long-term transport smoke from Canadian wildfires on U.S. PM2.5 levels during a 17-day study period. This topic is attractive but the manuscript is not well structed and the presentation of results/conclusion is confusing. The text is hard to read and the figures (too many figures) are in low quality. Detailed comments are listed below

We thank the reviewer for their time and effort in reviewing our work. Some responses may not display correctly here, so we have also attached a supplementary file with our detailed responses.

1.The structure of this manuscript needs to be carefully re-design. For example, the first paragraph of introduction talked about PM2.5, fire-related PM2.5, fire-related PM2.5 health effects and again PM2.5 health effects without a logical sequence.

Thank you for pointing this out. We have restructured the first paragraph of the introduction to establish a more logical flow. Additionally, we have reviewed and refined other sections of the manuscript to ensure a more coherent structure throughout.

2.The citation format needs to be double check throughout the text.

Thank you for pointing this out. We have thoroughly reviewed the manuscript to ensure that all citations adhere to the required format and are consistent throughout the text.

3.When averaging 1-km resolution AOD to 0.1-degree grid, how to deal with missing AOD data?

When averaging 1-km resolution AOD to a 0.1-degree grid, we calculate the average using only the valid AOD values within each grid box, excluding any missing values. If a grid box contains no valid AOD values, it is assigned a value of NaN.

4.Description of the AOD gap-filling method is not clear. In the third step, “calculating the ratio of pixels with valid MAIAC AOD over the total pixels”, are the total pixels satisfied the pixel selection conditions defined above? If the ratio is less than or equal to 80%, the AOD is filled by geo-weighted method including all the pixels within the pre-selected box or including only the pixels satisfied the pixel selection conditions defined above?

The total pixels in this context refer to all pixels within the selected box, regardless of whether they satisfy the filtering conditions. If the ratio of valid MAIAC AOD pixels to the total pixels in the box is less than 80%, the AOD value is estimated specifically for the target pixel, not for the entire box. Each pixel with missing AOD is evaluated individually using a moving box approach to ensure localized and precise gap-filling. We revised the text:

"To estimate the AOD value for a location without valid MAIAC AOD, we first select a 9×9-pixel box (spatial resolution of 0.1-degree) centered on the target pixel. Within this box, we identify pixels that satisfy two conditions: (a) having valid MAIAC AOD data, and (b) having a small AOD_{wrf} difference (<0.1) compared to the target pixel. If fewer than 50 pixels meet these criteria, the box radius is expanded, and the filtering process is repeated. Once at least 50 pixels are identified, we calculate the ratio of valid MAIAC AOD pixels to the total number of pixels in the box, where total pixels refer to all pixels in the selected box, regardless of filtering. If the ratio exceeds 80%, the AOD for the target pixel is estimated using the ordinary kriging (OK) method, based on the filtered pixels. However, if the ratio is 80% or lower, the AOD is calculated using a geographically weighted regression method that considers the neighboring ratio between MAIAC AOD and AOD_{wrf} . This process is performed individually for each target pixel using a moving box approach."

5.Regarding the hyperparameter of RF model, only the number of tree and number of parameters are considered and the model is over-fitted. The hyperparameter number of leaves (nodes) needs to be optimized to deal with the over fitting issue.

We have optimized the hyperparameters of the RF model, including the number of leaves (nodes), maximum depth of each tree, minimum samples required to split an internal node, and minimum samples required to be in a leaf node, to address the overfitting issue. Through this process, we observed that the GWR method outperformed RF under extreme conditions (e.g., during wildfire events) in our case. The discussion and related performance metrics have been updated accordingly.

6.In section 4.3, the control PM2.5 is calculated from simulated AOD, filled AOD and predicted PM2.5. I am not sure what is the contribution of AOD here. The control PM2.5 can be directly assessed from the estimated PM2.5 by calculating the Canadian fire contribution ratio as the simulated control PM2.5 divided by the simulated true PM2.5. Due to the non-linear relationship between AOD and PM2.5, the AOD-based calculation may introduce larger uncertainty. What's advantage of including AOD in this calculation?

Thank you for raising this point. We agree that the $\frac{PM_{2.5}}{AOD}$ ratio may introduce uncertainty in estimating the smoke contribution to surface PM2.5. To address this, we have revised our method to avoid directly using the $\frac{PM_{2.5}}{AOD}$ ratio. Instead, we now estimate the surface PM2.5 for the control case using the ratio between the WRF-Chem-simulated surface-level P2.5 concentrations of the experiment run (with Canadian smoke) and control run (without Canadian smoke), as follows:

$$PM_{2.5,control} = PM_{2.5} * \left(\frac{PM_{2.5,WRF,control}}{PM_{2.5,WRF,experiment}} \right)$$

The revised method uses the modeled surface P2.5 concentrations from WRF-Chem for both the control and experiment cases, which inherently accounts for the vertical distribution of smoke. By doing so, it avoids the assumption of directly using AOD to estimate surface-level

P2.5. We have updated the Section and equation 9, the results in the abstract have also been updated accordingly.

7.The result reported in line 295 directly compared AOD maps in Figure 2 and PM2.5 maps in Figure 4 and discussed the contribution of Canadian wildfires vs. US local fires to surface pollution. However, AOD and PM2.5 are not linearly correlated. This discussion here is with large uncertainty.

Thank you for pointing out the potential uncertainties in directly comparing AOD and PM2.5. To address this concern and enhance clarity, we have included surface PM2.5 dry mass difference maps for the four days in the revised manuscript. These maps directly represent the surface PM2.5 changes attributable to Canadian wildfire emissions, calculated from the difference between the Canadian wildfire case and the control case in the WRF-Chem simulations. By using these difference maps, we can more effectively illustrate the linkage between surface pollution changes and the associated pressure systems, reducing the uncertainties associated with comparing columnar AOD and surface PM2.5.

8.Section 5.3 discusses the PBL changes at hourly scale with diurnal patterns that is not consistent with other results at daily scale. This section can be removed.

Thank you for the suggestion. We agree that the discussion in Section 5.3 regarding PBL changes at the hourly scale is not fully consistent with the daily-scale analyses presented elsewhere in the manuscript. To maintain coherence and focus, we have removed this section from the revised manuscript.

9.Line 365-370, figure 10 appear early than figure 9 in the manuscript.

We have corrected the order of the figures in the manuscript so that Figure 9 now appears before Figure 10.

10.The different temporal trends reported by local stations, GWR and RF indicating that the model predictions are nor solid. What is the performance of the temporal trends of PM2.5 from EPA stations and from RF/GWR estimations at the same location? Please adjust the model prediction before further analyses.

The mean surface PM2.5 estimations for RF and GWR in Figure 14 are calculated across the entire study region, which inherently includes differences in the spatial distribution of sample points used for averaging. These differences largely explain the discrepancies between the estimations (RF and GWR) and EPA measurements.

As discussed in the text, the RF model tends to be more influenced by the spatial distribution of ground stations, resulting in regional mean values that closely follow the temporal trends of EPA station measurements. In contrast, the GWR model is better suited to capturing PM2.5

variations in areas with sparse station coverage, leading to differences in regional trends compared to RF and EPA measurements.

We believe these differences reflect the strengths and limitations of each model in different spatial contexts rather than an inherent lack of robustness in the predictions.

11.In Table 3, region 1 on Aug-22 with >12.2% Canadian wildfires' PM2.5 effects but only 3.3 $\mu gm^{-3}5$ concentrations? Please double check the number in table 2 and 3.

As mentioned above, we recalculated the surface PM2.5 for the control case. As a result, the percentage and PM2.5 concentration values have been updated in the revised manuscript.

12.Figure 1 What is "r"? Please label it in the figure or in the caption.

Thank you for pointing this out. The "r" in Figure 1 refers to the radius of the selected box used in the analysis. We have updated the figure caption to clarify this.

13.Figure 3 There are too many numbers in the figure.

Thank you for your feedback. To address this, we have changed the surface weather maps in Figure 3 for easier visualization and improved clarity. We hope the updated figure better conveys the key information.

14.Figure 4 Are these purple dots showing missing measurements or showing measurements of 0 $\mu g/m^3$?

The purple dots in Figure 4 represent measurements less than 2 $\mu g/m^3$. We have updated the plot to exclude all measurements below this threshold for clarity.

15.Figure 6 What date is this figure about?

Thank you for the comment. Since Section 5.3 has been removed, Figure 6 has also been removed from the revised manuscript.

16.Figure 11 (a), missing AOD should be shown as missing not as 0 in a continuous color scheme.

Thank you for your suggestion. We have updated Figure 11 to represent missing data in light gray, ensuring a clear distinction from areas of low AOD, which remain blue.

17.Figure 12 and Figure 13, the caption should be "100-fold CV" rather than "k-fold CV". Some key parameters, e.g. R and RMSE, should be added in the figures.

Corrected.

18.Figure 14 Why are some data points not on the line?

Thank you for pointing this out. To address your concern, we have remade Figure 14 with improved visualization, including adjustments to the data points and the addition of an EPA region map to provide spatial context.