

The authors developed a random forest model to predict ground-level PM₁₀ and PM_{2.5} based on geostationary satellite observations and model-based emission and meteorological outputs over South Korea. Oversampling and subsampling strategies were used to balance the training samples in order to better estimate high-level PM concentrations. This study contributes to more accurate PM predictions by appropriately adjusting the reference PM data and the synergistic use of satellite AOD data and other environmental variables. The manuscript is well written and is suitable for publication in Atmospheric Chemistry and Physics after the following comments are addressed.

Comments:

1. Line 29, Page 2: The sentence starts with “more recent studies ...” needs references about these studies.
2. Line 7, Page 3: can you add references to support the statement that many studies have focused on PM prediction in the United States because of less cloud cover in satellite data?
3. Section 3.1: it is not clear how the authors chose the sizes of the sampling window (e.g., 3×3 and 5×5)? Was there any sensitivity analysis being conducted to determine appropriate window sizes? In addition, it was likely that an increased adjusted sample size (after oversampling and subsampling) contributed to better modeling performance. So, more clarifications are needed to better explain the effectiveness of the oversampling and subsampling strategies.
4. Section 3.2: Did the authors test the correlation among the independent variables? If two or more independent variables are correlated with each other, the model may include more variables than it is necessary, which could lead to bias/uncertainty in the interpretation of the results.
5. Table 5: I suggest adding a column showing the sample sizes (N) of the models.
6. Line 13, Page 15: For the word “congestion”, did the authors mean “advection” or “convection”?