We would like to thank the reviewer again for their availability and willingness to review our study and for the valuable contributions provided by their recommendations. Our replies to the comments and recommendations of both reviewers are organized below.

Reply to the reviewer #01

The authors addressed my previous comments satisfactorily. I have a few minor comments still and suggest the paper to be accepted once these are addressed

Ln 48: "two critical information" rephrase to "critical information"
R: Thank you, the sentence was rephrased as suggested.

Ln 65-66: "which idealized internal mixing assumptions" is not fully clear, is there a word missing here?
R: The reviewer is right, the sentence was adjusted now.

Adjusted version: *"Brown et al. (2021) findings indicate that biomass-burning aerosols in most climate models are too absorbing, mainly due to treatments of aerosol mixing state. They found the internal mixing assumptions used in climate models to overestimate Black Carbon(BC) absorption when compared to the observations.* "

Please, see description in the manuscript 2nd paragraph of the Introduction.

Ln 209-210: "Online modelling … is calculated as" maybe rephrase to e.g. "The computation of optical properties for radiative transfer computations is usually based on a mass-weighted …"

R: The sentence was rephrased as suggested.

Ln 242: "indicated to separate" rephrase to "that separate"
R : Thank you, done.

Equation 1: I still find 'minimize' confusing here. WCSS is the "cost function" and try to find an optimum between WCSS and the number of clusters. As far as I understand, the computation of WCSS is just the sum of the within-cluster variances over the clusters, and does not contain any minimization.

R: The reviewer is right, 'minimize' is misplaced, the expression does not minimize, indeed it just performs the sum of the within-cluster variances over the clusters for a range of clusters number (k), from which the optimal number of clusters (k) is selected using the Elbow method. The WCSS equation was adjusted in a way that expresses the real meaning of WCSS. For that, we needed to adjust the text that described it.

Please, see all the adjustments highlighted in the topic 2.4 of the marked manuscript.

Ln 558: "which can be inferred from .." I might have misunderstood this sentence, but I would suggest to rephrase to e.g. "which indicate a potential growth of …"

R: Thank you, the sentence was rephrased as suggested.

Ln 816-817: 3D modelling of optical properties does not (necessarily) involve performing Mie calculations for each 3D grid cell. The outcomes of Mie computations are often tabulated and interpolated, which is much faster.
R: Yes, the reviewer is correct, most of the models nowadays use look-up-table instead of online Mie calculations. Since we did not perform an objective evaluation of the performance of these different approaches, the sentence was removed. We understand that it is not critical to our discussion and conclusions.

Reply to the reviewer #02

1. The authors have made several useful corrections and clarifications (terminology fixes; specification of CMIP6; expanded explanation of LR/LDR derivation and AERONET v3 improvements; explicit statement that Table 1 variables are AERONET inversion products; clarification of temporal synchronization and of use of MERRA-2 ADP at 550 nm; and additional figures in the supplement showing predictor importance).

However, several key methodological and robustness issues remain insufficiently addressed and must be corrected before acceptance:

Quantify or test the impact of AERONET inversion uncertainties on K-means cluster identification (sensitivity test recommended). Authors should have either:

(i) performed and reported a simple sensitivity analysis (recommended), e.g., perturb each input variable by its published retrieval uncertainty (or by $\pm1\sigma$) and rerun clustering to show cluster stability/membership changes and report metrics (e.g., fraction of points changing cluster, confusion matrix, silhouette change),

or

(ii) if not possible, explicitly quantify typical uncertainties for each inversion variable (cite AERONET v3 uncertainty estimates), discuss how these uncertainties could bias particular clusters (qualitatively but concretely), and state how future work will address it.

Instead, they acknowledge uncertainties and state they did not consider how differing uncertainties affect clustering by adding a discussion noting as a limitation they claim consistency with previous works. Without making any attempts (i or ii), the robustness of the five regimes is unclear.

R: We thank the reviewer for his comments on this aspect and for the guidance provided, as recommended, a sensitivity analysis was performed by perturbing the input data by ±1 standard deviation(SD) and reapplying the k-means clustering. Cluster stability was assessed using the fraction of points changing cluster membership relative to the original classification. The fraction of points changing cluster assignment was 15.3% for the +1SD perturbation and 17.6% for the −1 SD perturbation, yielding a mean sensitivity of 16.5%. Only 0.5% of points changed cluster membership under both perturbations, indicating that the observed sensitivity is largely confined to boundary points, while the cluster cores remain robust.

Please, see the description in the topic Methods 2.4 (last paragraph) and in the Results topic 3.1 (2nd paragraph) of the marked manuscript.

2. Explicitly report preprocessing (scaling/normalization) used for K-means and Random Forest, and, if not performed, rerun analyses with appropriate normalization.

R: The process of scaling data through normalization was done for clustering based on K-means (check), below is displayed, the code name and the part of the script where the scaling process is done.

(Code_01_zenodo_paper_kmeans_clusters_number_criteria_for_aerosol_mixtures_piberica_basedon_elbow_and_stability_method.py)

"#Standardize features by removing the mean and scaling to unit variance.
scaler = StandardScaler()
scaled_data = scaler.fit_transform(data)"

Please, see the description in the topic Methods 2.4 (4th paragraph) of the marked manuscript.

Regarding Random Forest application, unless we misunderstood the reviewer recommendation, we understand that there is literature supporting no requirement for data normalization, since tree-based algorithms split nodes based on feature thresholds rather than absolute magnitudes. These splitting criteria are invariant to monotonic transformations and the scale of input variables (Breiman, 2001; Hastie et al., 2009). Pedregosa et al.(2011) also states that Random Forest is robust to varying scales as it partitions the feature space without computing distances.

References:

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

3. Clearly state the spatial collocation procedure between AERONET sites and MERRA-2 grid cells.

R: To proceed with the collocation between AERONET retrievals and MERRA-2, for a spatial matching we considered the nearest neighbor by taking the MERRA-2 grid cell that contains the AERONET station location, and for temporal collocation MERRA-2 hourly aerosol diagnostics were matched to the closest AERONET observation time.

This description was included in the manuscript (in the 3rd paragraph of the topic 2.5) of the marked manuscript.

4. Clarify whether MERRA-2 optical properties were used directly or mass fields only; if optical fields are used, provide a brief validation vs AERONET (basic error metrics) or properly justify omission.

R: Only MERRA-2 mass fields were used in the Random Forest training process, optical fields from MERRA-2 were not used at all. MERRA-2 optical fields(SSA and AE) were only used at the final stage to compare the difference that would be observed between the present study trained model and MERRA-2 optical module in the prescription of these aerosol intensive optical properties fields.

A description clarifying that only mass fields were used in the training was included in the manuscript (in the 3rd paragraph of the topic 2.5) of the marked manuscript

5. Provide fuller ML methodological details (train/test split or CV strategy, handling of class imbalance, hyperparameter tuning, per-class metrics and confusion matrices, and method for computing feature importance).

R: In the previous version of the manuscript we already pointed out the CV and hyperparameter tuning strategies adopted, and now we try to extend the description related to the ML methodology that we used. These descriptions can be seen in the highlighted parts of marked manuscript (methods topic 2.5), nevertheless we include the most relevant description below:

a. train/test split

"Therefore, the dataset was randomly divided into training and testing subsets, with 70% of the data used for model training and 30% reserved for independent evaluation. This was done using the train_test_split utility from the Scikit-learn library (Abraham et al. 2014)."

b. CV strategy
hyperparameter tuning
handling of class imbalance

"We used stratified k-fold cross-validation integrated within a RandomizedSearchCV hyperparameter optimization process as a strategy. Our hyperparameter optimization was performed using RandomizedSearchCV with five-fold cross-validation. The search space included the number of trees (n_estimators) sampled uniformly between 50 and 500 and the maximum tree depth (max_depth) sampled between 1 and 20. A total of five random hyperparameter combinations were evaluated, and the best-performing model was refitted on the full training dataset. The random search methodology was used to find parameter combinations inside the parameter space without the processing demands of grid search and with the stratified k-fold cross-validation we search to ensure that each fold has approximately the same class proportions as the full dataset, which allows a fair evaluation, since every validation set includes samples from all classes. This also contributes to the meaning of the performance metrics in relation to the minority classes, for example strong absorbing aerosol regimes. The stratified k-fold also favours a more stable training, given that in every training split the less frequent aerosol property regimes are also seen, which helps to reduce variance in model performance across folds. This strategy contributes to improving the hyperparameter tuning, once the RandomizedSearchCV won't select parameters based on misleading folds. So, by preserving class distribution in every fold and preventing biased results, the strategy based on stratified k-fold cross-validation helps to handle class imbalance, which in turn improves the model reliability and generalization. Class imbalance is typical in atmospheric aerosol characterization, where extreme but radiatively important aerosol regimes, like intense smoke episodes, are rare compared to more common background conditions. Therefore, with the strategy described, we also aimed to address the issues of class imbalance of aerosol regime classification in our study. "

c. method for computing feature importance

"To identify which aerosol types most influence aerosol regime prescription and to understand whether meaningful types are driving predictions, we used the best_rfc.feature_importances_, an utility in scikit-learn's RandomForestClassifier, to calculate the scores indicating the importance of each aerosol types in the training dataset. The importance of an aerosol type in making predictions was based on how much it reduces impurity across all trees. Each decision tree in the forest splits data using different features, and each split reduces impurity. The reduction in impurity is attributed to the feature used at that split. An aerosol type importance is based on the frequency that it is used to split nodes. "

6. Either quantify the radiative impact of misclassification (radiative transfer sensitivity test) or soften/remove the unquantified claim.

R: Since a new article is planned to focus exclusively on the radiative impact of the application of the prescription approach developed in this current article, we followed the review recommendation and the mentioned claim has been removed. For the sentence where radiative impact is invoked we replace *"radiative impact (or error)"* for *"optical properties prescription error"*, which is directly aligned with our results.

For instance, from this:

"The classifier's confusion in this case is between the two dust-regime models; therefore, the induced error in radiative transfer calculations would be lower than that if the confusion was between a dust and a non-dust regime, especially like C2, which is substantially different from any of the dust regimes. Rarely does the classifier take either C0 or C1 as C2, C3, and C4, a case where substantial error in the radiative effect would be expected."

To this:

"The classifier's confusion in this case is between the two dust-regime models; therefore, the induced error in optical properties prescription would be lower than that if the confusion was between a dust and a non-dust regime, especially like C2, which is substantially different from any of the dust regimes. Rarely does the classifier take either C0 or C1 as C2, C3, and C4, a case where substantial error in the optical properties prescription would be expected."

7. Either quantify or rephrase the "lower computational cost" claim.

R: The sentence has been rephrased.

8. Address the minor repetition flagged by the reviewer by editing the text.

R: We tried to go through all repetition in the text.

Once the authors address the methodological gaps and provide the requested quantitative sensitivity and validation, as well as ML details, the manuscript will be much stronger and suitable for reconsideration.

We thank the reviewer again and hope that all the important requirements have been answered.