Response to Reviewer #1:

The manuscript" Global aerosol typing classification using a new hybrid algorithm utilizing Aerosol Robotic Network data" aims to develop a new aerosol-type classification model using an innovative hybrid algorithm to improve the precision and efficiency of aerosol-type identification. The study shows good consistency between the new method and traditional Gaussian density cluster method, with consistency rates of 90%, 85%, 84%, 84%, and 100% for dust, mixed-coarse, mixed-fine, urban/industrial, and biomass burning aerosols, respectively. Overall, the manuscript provides a well-structured and clear overview of the study design, methodology, and results. The authors communicate the significance of their findings in addressing the issue of classifying aerosol type accurately and efficiently in global scale, which has important implications for aerosol inversion and aerosol pollution study. However, there are a few areas where the manuscript could be improved.

1. The authors should elaborate on why they chose Mie scattering model to build an aerosol optical database for classifying the aerosol type.

Response: Thank you for your constructive advice. The Mie scattering model, known for its simplicity and practicality, provides an analytic solution to Maxwell's equations for light scattering by ideal spherical particles. It efficiently depicts the scattering and absorption properties of aerosols in the atmosphere, serving as fundamental basis of radiative transfer, Lidar, and optical particle characterization (Ma et al.,2007; Bian et al., 2017; Michael et al., 1994) (see line 240-245).

There are two reasons why the Mie scattering model is used to calculate aerosol optical parameters instead of directly using observational data: First, the optical parameters from AERONET site, such as AOD and complex refractive index, are seriously missing, and the effective information provided is limited, particularly, the amount of observation data of AERONET site dominated by biomass combustion can no longer meet the requirement of machine learning algorithms. Second, the calculation of aerosol optical properties by fixing the refractive index of the Mie scattering model

allows for precise determination of aerosol types (see line 231-237).

2. It would be helpful to provide more context on the limitations of their approach and future directions for research in this area.

Response: Thanks for your advice. On page 32, the last paragraph, we further enriched the limitations and pointed out the possible future directions for the aerosol-type identification study.

However, marine aerosols were not considered for fewer valid data after site screening in this study. The dust aerosol in the Mie scattering model is assumed to be spherical, and the actual natural environment is not spherical, which inevitably brings errors, and the accuracy of the optical database needs to be further improved. In the future, with the development of machine learning, random forest algorithms can be replaced with stronger learning algorithms. Meanwhile, multi-source satellite data and reanalysis products can be incorporated into aerosol-type identification. This study will provide support for the identification and control of air pollution sources. (see line 656-661).

3. The manuscript needs to have more information in the result about the improvements in calculation time efficiency of aerosol type classification with a specific scale.

Response: Thanks for the advice.

Additionally, in this study, the number of 326400 data points from optical parameters database and 98000 observed data for calculation spans from Jan.1st,1993 to Dec.31st,2021, passing through Gaussian kernel density clustering algorithm and new hybrid algorithm Python progresses, which is archived on the personal Windows system computer (Intel® Core™ i7-10710U,16G DDR4 2666MHz, 512G PCIE SSD). The computational time for the two algorithms indicates the new hybrid algorithm runs faster than the Gaussian kernel density clustering algorithm with huge quantities of data and trained in advance, which can obtain aerosol type in 20 seconds, in contrast, it will take 30 to 40 seconds to obtain aerosol type in one site by using the Gaussian algorithm (see line 449-458).

4. The manuscript requires a clearer explanation of how the random forest model was implemented and any potential biases associated with the model.

Response: Thanks. On page 15, line 321-336, and page 17, line 371-379, we implement the further assessment of the random forest algorithm in the new hybrid algorithm and analyze the potential biases associated with the model.

In this study, the evaluation matrix was brought into this study, and it further quantitatively assesses the performance of the Gaussian density clustering algorithm and the new hybrid algorithm. The metric indexes include accuracy, recall, precision, and F-scores (Reddy et al., 2022). Here, the indexes are adjusted to micro-precision, micro-recall, micro-F1-score, and accuracy to solve the multi-classification problem. Micro refers to the weighted average of the five aerosol types rather than the arithmetic mean, due to the large difference in sample size among the five aerosol types, the arithmetic mean is highly susceptible to the influence of very large or very few sample size aerosol types.

Table 5 shows the metric index value of the random forest algorithm in the new hybrid algorithm. The micro-precision, micro-recall, micro-F1 score, and accuracy are 0.95, 0.89, 0.91, and 0.89, respectively. These metrics are derived from the core values of the window, as determined by the Gaussian density clustering algorithm. Consequently, the strong performance of these indicators further confirms the efficacy and reliability of the newly developed hybrid algorithm.

Table 5 Matrix evaluation between hybrid classification algorithm and Gaussian density clustering algorithm

	Micro-Precision	Micro-Recall	Micro-F1-Score	Accuracy
New Hybrid algorithm	0.95	0.89	0.91	0.89

Line 99-106 describe the limitation of the machine learning algorithm, which also the potential bias of random forest. The content is as follows:

However, some challenges remain in identifying aerosol types through machine learning. First, the amount of valid ground aerosol property data that can be used for training is less due to cloud removal and quality control. Second, the accuracy of machine learning depends on the labeled aerosol typing dataset, and finding a suitable

classification method to classify the dataset is challenging. Third, evaluating the accuracy of the final trained model is also tedious (Zhang & Li, 2019; Siomos et al., 2020; Choi, et al., 2021a,b).

In addition, the random forest algorithm associated with learning ability is insufficient and can be replaced by other stronger learning algorithms in the future (see line 656-658).

5. More relevant literature review should be included, especially those from the last three years.

Response: Thank you for your patience. We added several recent three years of study in this field to support our references. Please, refer to these references:

Elham Ghasemifar: Climatology of aerosol types and their vertical distribution over Iran using CALIOP dataset during 2007–2021,Remote Sensing Applications: Society and Environment,32, 101053, 2352-9385,https://doi.org/10.1016/j.rsase.2023.101053.2023.

Nandan, R., Ratnam, M.V., Kiran, V.R., Madhavan, B.L., & Naik, D.N.: Estimation of Aerosol Complex Refractive Index over a tropical atmosphere using a synergy of in-situ measurements. Atmospheric Research, 257, 105625, https://doi.org/10.1016/J.ATMOSRES.2021.105625, 2021

Reddy LA, Glover TA, Dudek CM, Alperin A, Wiggs NB, Bronstein B.: A randomized trial examining the effects of paraprofessional behavior support coaching for elementary students with disruptive behavior disorders: Paraprofessional and student outcomes. J Sch Psychol. 2022 Jun;92:227-245. https://doi.org/10.1016/j.jsp.2022.04.002, 2022.

Wang J, Liu Y, Chen L, Liu Y, Mi K, Gao S, Mao J, Zhang H, Sun Y, Ma Z.: Validation and calibration of aerosol optical depth and classification of aerosol types based on multi-source data over China. Sci Total Environ. 2023 Dec 10;903:166603. doi: 10.1016/j.scitotenv.2023.