atmosphere

MDPI

Supplementary Materials:

Aerosol Optical Properties and Type Retrieval via Machine Learning and an All-Sky Imager

Stavros-Andreas Logothetis [1], Christos-Panagiotis Giannaklis [1], Vasileios Salamalikis [2], Panagiotis Tzoumanikas [1], Panagiotis-Ioannis Raptis [3], Vassilis Amiridis [4], Kostas Eleftheratos [3,5] and Andreas Kazantzidis [1,*]

[1] Laboratory of Atmospheric Physics, Physics Department, University of Patras, GR-26500 Patras, Greece; logothetis_s@upnet.gr (S.-A.L.); up1057893@upnet.gr (C.-P.G.); tzumanik@ceid.upatras.gr (P.T.)
[2] NILU—Norwegian Institute for Air Research, P.O. Box 100, 2027 Kjeller, Norway; vsal@nilu.no
[3] Department of Geology and Geoenvironment, National and Kapodistrian University of Athens, GR-15784 Athens, Greece; piraptis@meteo.noa.gr (P.-I.R.); kelef@geol.uoa.gr (K.E.)
[4] Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, GR-15236 Athens, Greece; vamoir@noa.gr
[5] Center for Environmental Effects on Health, Biomedical Research Foundation of the Academy of Athens, GR-11527 Athens, Greece
* Correspondence: akaza@upatras.gr

Table S1. Machine Learning Architecture including their hyperparameters that are tuned during the training procedure.

MLA	Hyperparameters
GBM	**n_estimators** = [50, 100, 500, 1000, 1500, 2000] **learning_rate** = [0.3, 0.2, 0.1, 0.05, 0.01, 0.001, 0.0001] **max_depth** = [2, 5, 10, 12, 15, 20] **num_min_samples_leaf** = [2, 3, 4, 5, 6] **max_features =** [1,2,3]
XGBM	**n_estimators** = [50, 100, 500, 1000, 1500, 2000] **learning_rate** = [0.3, 0.2, 0.1, 0.05, 0.01, 0.001, 0.0001] **max_depth** = [2, 5, 10, 12, 15, 20] **colsample_bylevel =** [0, 1] **min_child_weight =** [1, 3, 5, 7] **gamma =** [0.0, 0.1, 0.2, 0.3, 0.4] **subsample =** [0.6, 0.7, 0.8, 0.9] **colsample_bytree =** [0, 1]
LGBM	**n_estimators** = [50, 100, 500, 1000, 1500, 2000] **learning_rate** = [0.3, 0.2, 0.1, 0.05, 0.01, 0.001, 0.0001] **max_depth** = [2, 5, 10, 12, 15, 20] **num_leaves** = [2, 5, 10, 12, 15, 20]
RF	**n_estimator** = [100, 300, 500, 700, 900, 1000, 1500, 2000] **max_features** = [1, 2, 3] **max_depth** = [None, 10, 15, 20, 25] **min_samples_split** = [2, 3, 4, 5, 6] **min_samples_leaf** = [1, 2, 3, 4, 5, 6]
SVM	**C =** [0.1, 1, 10, 100, 1000, 1500, 2000] **gamma =** [20, 10, 4, 2, 1, 0.1, 0.01, 0.001, 0.0001, 0.00001] **epsilon** = [0.1, 0.05, 0.01, 0.005, 0.001, 0.0001, 0.00001]
MARS	**max_terms** = [1, 2, 3] **max_degree** = [1, 2, 3]
KNN	**n_neighbors** = [neighbors from 1 to 30 per 1]
ANN	**Hidden layer 1** = [nodes from 8 to 64 per 4] **Hidden layer 2** = [nodes from 8 to 32 per 4]

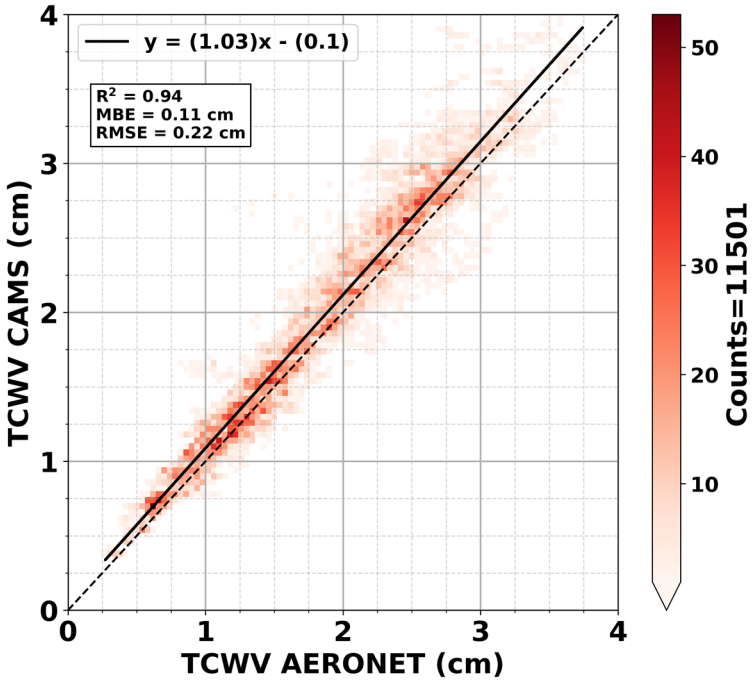

Figure S1. (a) Scatter plot between CAMS and AERONET total column water vapor retrievals.

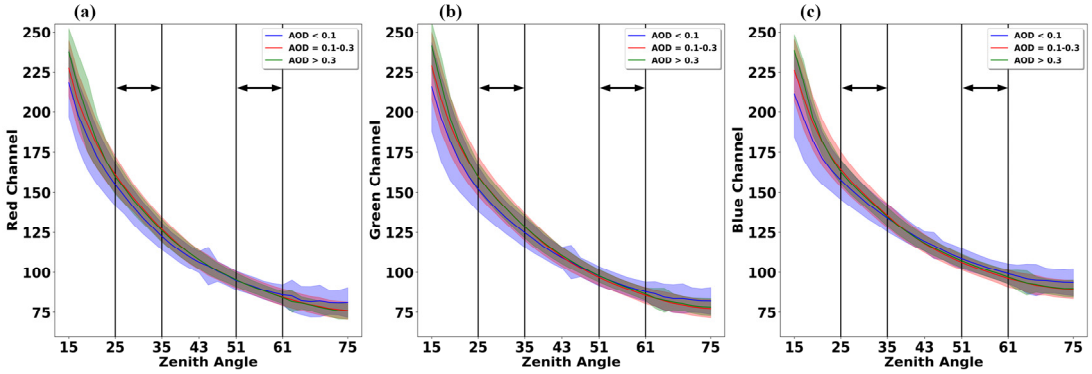

Figure S2. (a) Red, (b) Green, and (c) Blue channel intensity against zenith angle. The zenith angle points are shown in Fig. 1 for the principal plane (straight line 1), ranging from sun's center point (zenith angle=0) and to sun's area up to 75°. Different colors refer to three different AOD ranges. Blue, red, and green colors represent AOD range values that are relatively low (0.1), moderate (0.1–0.3), and high (> 0.3). The shaded areas around the lines correspond to ±1 standard deviation bands. The two rectangles refer to the specific zenith angles (25–35° & 51–61°) which are used in Sect. 4.1.

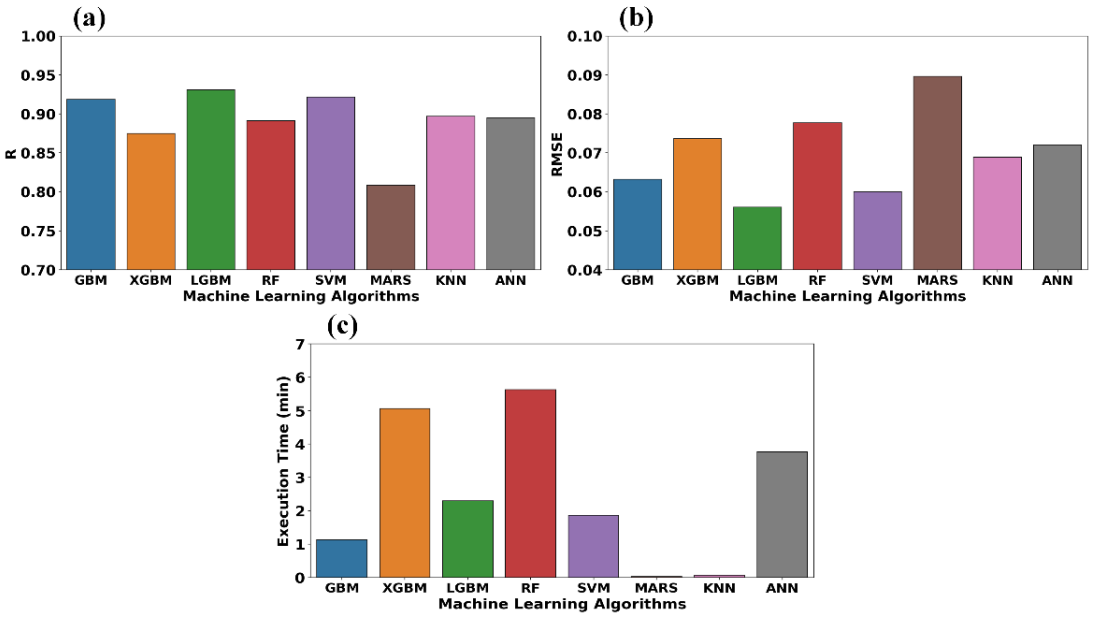

Figure S3. (a) R correlation coefficient and (b) RMSE for the eight different machine learning models by using the 60 pixels of Fig. 1 and the total column water vapor as model inputs to retrieve AOD_{440nm}. (c) Execution time for models training procedure including the tuning.