Retrieval of aerosol optical properties via an all-sky imager and machine learning: Uncertainty in direct normal irradiance estimations

Logothetis S.A.[1*], Giannaklis C.P.[1], Salamalikis V.[2], Tzoumanikas P.[1], Raptis P.I.[3], Amiridis V.[4], Eleftheratos K.[3,5] and Kazantzidis A.[1]

[1] Laboratory of Atmospheric Physics, Physics Department, University of Patras, Patras, Greece
[2] NILU-Norwegian Institute for Air Research, P.O. Box 100, 2027 Kjeller, Norway
[3] Department of Geology and Geoenvironment, National and Kapodistrian University of Athens, Athens, Greece
[4] Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens, Greece
[5] Center for Environmental Effects on Health, Biomedical Research Foundation of the Academy of Athens, Athens, Greece
*Correspondence: logothetis_s@upnet.gr

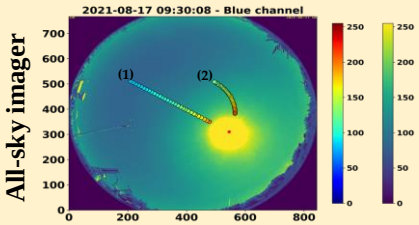
Objective

The aim of the presented study is to use the sky information from an all-sky imager (ASI) to train a machine learning (ML) algorithm to retrieve the aerosol optical depth (AOD) and Ångström Exponent ($AE_{440\text{-}670nm}$) in order to estimate the direct normal irradiance (DNI).

Data

Measurement site: National Observatory of Athens (NOA), Thissio, Greece
Reference instrument: CE318 sun-sky photometer (CIMEL Electronique)
All-sky imager: Mobotix Q24M model
Campaign: ASPIRE project (https://aspire.geol.uoa.gr/)
Temporal coverage: 01/01/2021 – 18/11/2021

All-sky imager



AERONET

Input parameters

Input parameters from the images:
1) 30 RGB pixels with a 2° step across the principal plane (straight line).
2) 30 RGB pixels with a 2° step across almucantar (curved line).
3) the saturation area (SAT in %), which is defined as the ratio between the number of pixels around the sun that includes sunlight and the total number of image pixels.

Auxiliary input parameters:
1) solar zenith angle, 2) total column water vapor

Reference parameters

Input parameters from the AERONET station:
1) AOD at 440nm
2) AOD at 500nm
3) AOD at 670nm
4) AE between 440 and 675nm

Level 2.0 data from AERONET **Version 3.0**.

Methodology

[1] Data splitting: The dataset is separated into **train** and **test** dataset, including the **70%** and **30%** of respectively.

[2] Machine Learning Algorithm: The Light Gradient Boosting Machine (LGBM) for regression is applied using the train dataset.
- A randomized search procedure was performed during the training in order to find the best combination of hyperparametes, including a 10-fold cross-validation. The LGBM scheme with the highest performance is implemented to evaluate the test dataset.

[3] LGBM distinct models: In total, four LGBM models are trained. Three for retrieving the spectral AOD and one for AE.
- For the AOD_{440nm}, AOD_{500nm}, and AOD_{675nm} ML models, the corresponding images for the blue, green, and red channels are used.
- For the AE, all RGB values are applied.

Acknowledgments

We acknowledge support of this work by the project DeepSky co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code: T2EDK - 00681). Kostas Eleftheratos acknowledges support from the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Atmospheric parameters affecting Spectral solar IRradiance and solar Energy (ASPIRE), project number 300).

Co-financed by Greece and the European Union

Results: Aerosol optical properties retrieval performance

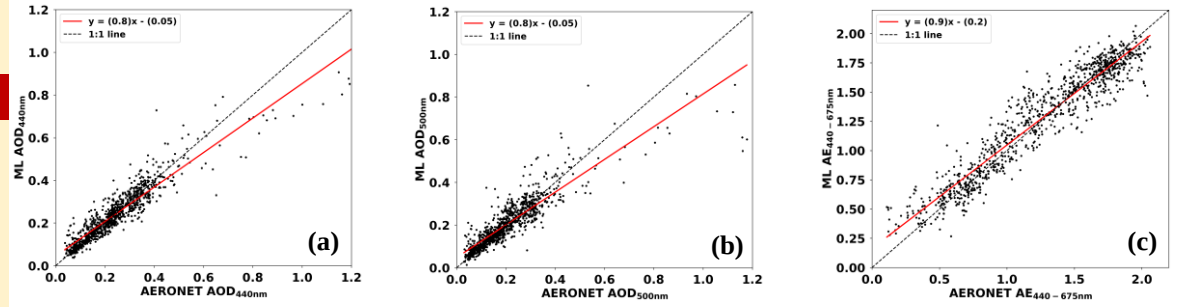


Figure 1: Linear relationship of ML retrieved **(a)** AOD_{440nm}, **(b)** AOD_{500nm}, and **(c)** $AE_{440–675nm}$ as a function of AERONET

- ✓ The coefficient of determination (R^2) ranged between 0.79 and 0.86 for AODs and $AE_{440\text{-}675nm}$.
- ✓ The relative biases ranged between −0.71% and 1.4% for all aerosol optical properties.

Results: DNI estimations and uncertainties

DNI estimation: the DNI is calculated using libRadtran and the UVSPEC radiative transfer code with SDISORT as radiative transfer equation. AOD and AE are used from AERONET measurements and ML retrievals. The DNI estimations using the retrieved ML and AERONET aerosol optical properties are abbreviated as "DNI ML" and "DNI AERONET," respectively.

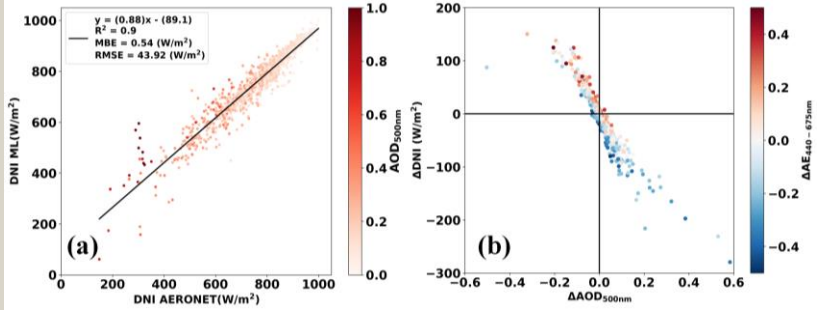


Figure 2: (a) Linear relationship between the retrieved and AERONET DNI. The color bar presents the AERONET AOD at 500nm. **(b)** Linear relationship between DNI discrepancies ($\Delta DNI=DNI_{AERONET}-DNI_{ML}$) and AOD_{500nm} discrepancies ($\Delta AOD=AOD_{AERONET}-AOD_{ML}$). The color bar presents the $AE_{440–675nm}$ discrepancies ($\Delta AE_{440–675nm}=AE_{440–675nm;AERONET}-AE_{440–675nm;ML}$).

- ✓ A R^2 of 0.90 is found, revealing an adequately good linear relationship between the ML and AERONET DNI (Fig. 2a).
- ✓ It is apparent that the AOD retrieval performance regulates the ΔDNI (Fig. 2b).