



Proceeding Paper

Retrieval of Aerosol Optical Properties via an All-Sky Imager and Machine Learning: Uncertainty in Direct Normal Irradiance Estimations [†]

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; up1057893@upnet.gr (C.-P.G.); tzumanik@ceid.upatras.gr (P.T.); akaza@upatras.gr (A.K.)

² NILU—Norwegian Institute for Air Research, 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: logothetis_s@upnet.gr

[†] Presented at the 16th International Conference on Meteorology, Climatology and Atmospheric Physics—COMECAP 2023, Athens, Greece, 25–29 September 2023.

Abstract: Quality-assured aerosol optical properties (AOP) with high spatiotemporal resolution are vital for the accurate estimation of direct aerosol radiative forcing and solar irradiance under clear skies. In this study, the sky information from an all-sky imager (ASI) is used with machine learning (ML) synergy to estimate aerosol optical depth (AOD) and the Ångström Exponent (AE). The retrieved AODs (AE) revealed good accuracy, with a dispersion error lower than 0.07 (0.15). The retrieved ML AOPs are used to estimate the DNI by applying radiative transfer modeling. The estimated ML DNI calculations revealed adequate accuracy to reproduce reference measurements with relatively low uncertainties.

Keywords: all-sky imager; aerosol optical properties; machine learning; aerosol retrieval; AERONET; DNI estimations



Citation: Logothetis, S.-A.; Giannaklis, C.-P.; Salamalikis, V.; Tzoumanikas, P.; Raptis, P.-I.; Amiridis, V.; Eleftheratos, K.; Kazantzidis, A. Retrieval of Aerosol Optical Properties via an All-Sky Imager and Machine Learning: Uncertainty in Direct Normal Irradiance Estimations. *Environ. Sci. Proc.* **2023**, *4*, 133. <https://doi.org/10.3390/environsciproc2023026133>

Academic Editors: Konstantinos Moustiris and Panagiotis Nastos

Published: 30 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Aerosol particles constitute the highest source of uncertainty in direct aerosol radiative forcing and solar irradiance estimations under cloud-free conditions [1]. So, accurate aerosol optical and microphysical properties are vital to alleviating the uncertainties in solar resource and climate change studies. Currently, several monitoring techniques are available for AOP retrieval, based on satellite and ground-based measurements. The most widely used ground-based network is AERONET (AERosol RObotic NETwork). AERONET delivers AOP at the highest temporal resolution (5–15 min), using several inversion algorithms and CIMEL sun-photometers [2–4]. From space, several passive sensors aboard satellites are available, such as MODIS (MODerate resolution Imaging Spectroradiometer) [5–7] and MISR (Multi-angle Imaging SpectroRadiometer) [8]. Despite the great performance of those aerosol remote sensing data sets, there are still some limitations.

Although ground-based measurements are optimal in terms of accuracy, a limitation is that their global coverage is sparse. On the other hand, satellite-based AOP retrievals provide approximate global coverage; a limitation is that their temporal variability is up to two measurements per day. For this reason, several alternative techniques for AOP

retrieval have emerged in the last few years to fill the gap and extend the availability of accurate AOP. In this study, a supervised ML algorithm is applied to retrieve AOPs such as the spectral AOD and AE by using the sky information captured from an all-sky imager (ASI) and the aerosol information from AERONET. In addition, the potentiality of using the ML retrieved AOPs to estimate the direct normal irradiance (DNI) is presented using radiative transfer modeling, analyzing their levels of uncertainty.

2. Data

The data applied in this work were acquired from an ASI and a CIMEL sun-photometer installed at the National Observatory of Athens (NOA) in Thissio, Greece. The CE318 sun-sky photometer installed in the station is the standard instrument of AERONET. For the sake of this study, Level 2.0 data from the latest AERONET Version 3 were used, including pre-field and post-field calibrations [3], providing quality-assured data that are optimal for the application of the proposed methodology. Moreover, the used ASI is a commercial Mobotix Q24M model, which captures images from the entire upper hemisphere every 640 μ s and stores them in 24-bit JPEG format with a spatial resolution of 1024 \times 768 pixels. In particular, the sensor has a red–green–blue (RGB) filter. So, images on those three spectral channels (440, 500, and 675 nm) were used, including color intensities ranging from 0 to 255.

The AERONET measurements were temporally synchronized with the closest ASI image, spanning a time frame between 1 January 2021 and 18 November 2021. All the images are examined to be under cloud-free conditions by visual inspection. In addition, measurements with a solar zenith angle (SZA) greater than 70° are eliminated from the analysis to prevent cases with the sun at the edges of the image, where many obstacles interfere. In total, 3212 images were retained to apply the proposed methodology.

3. Methodology

In this study, a machine learning algorithm, the Light Gradient Boosting Machine (LGBM) [9], is applied to retrieve AOD and AE. The data used as input parameters to the ML algorithm are the following: (1) RGB information from the ASI, (2) SZA, and (3) total column water vapor (TCWM). TCWV data were obtained through the CAMS (Copernicus Atmosphere Monitoring Service) reanalysis product [10]. The RGB ASI information are extracted from the following two different aspects: Firstly, 60 pixels were selected from the RGB images outside the sun saturated area including (1) 30 pixels with a 2° step across the principle plane and (2) 30 pixels with a 2° step across almucantar. Secondly, the saturation area (SAT in %) is applied, which is defined as the ratio between the number of pixels around the sun that includes sunlight and the total number of image pixels. SAT is related to Mie theory [11], providing valuable information for ML algorithm training.

The dataset has initially been separated into two datasets, i.e., the train and the test, which include 70% and 30% of the whole dataset, respectively. 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, while for the AE, all RGB values are applied. All four models included the SZA and TCWV. Before the training process, all input parameters are normalized to include values solely between 0 and 1. In particular, the Min–Max normalization process is applied to all parameters except SZA, where the cosine function is used. To achieve optimal accuracy, a randomized search procedure was performed during the LGBM algorithm training to find the best combination of hyperparameters, including a 10-fold cross-validation process using the mean square error as a loss function. After the training of the LGBM algorithm, the LGBM scheme with the highest performance, including the best combination of hyperparameters, is implemented to evaluate the test dataset.

4. Results

4.1. Aerosol Optical Properties Retrieval Performance

In the first section of the results, the ML model’s AOP retrieval performance is investigated. Figure 1 depicts the ML model’s performance against the reference measurements (AERONET). In general, AOD values range between 0.02 and 1.2. Athens, Greece, is frequently affected by dust events from the Sahara Desert, leading to quite high aerosol burden levels (>0.8). For AODs, the MBE (RMSE) values range between -0.71 and 0.033% (23.1 and 33.9%) (Table 1). Regarding $AE_{440-675nm}$, rMBE (rRMSE) equals 1.4% (12.0%) (Table 1).

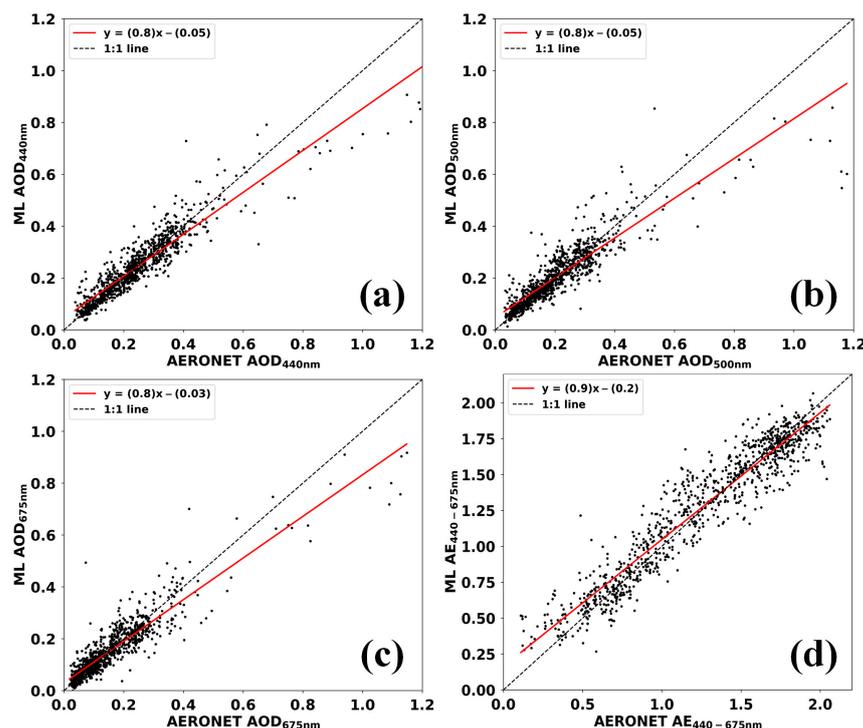


Figure 1. Linear relationship of ML retrieved (a) AOD_{440nm} , (b) AOD_{440nm} , (c) AOD_{675nm} , and (d) $AE_{440-675nm}$ as a function of AERONET.

Table 1. Mean Bias Error (MBE), Root Mean Square Errors (RMSE) and coefficient of determination (R^2) for AOD_{440nm} , AOD_{440nm} , AOD_{675nm} , and $AE_{440-675nm}$. In the parenthesis are presented the relative MBE and RMSE.

Variable	MBE (rMBE)	RMSE (rRMSE)	R^2
AOD_{440nm}	-0.0012 (-0.51%)	0.056 (23.1%)	0.86
AOD_{500nm}	0.000069 (0.033%)	0.066 (32.4%)	0.79
AOD_{675nm}	-0.0011 (-0.71%)	0.053 (33.9%)	0.85
$AE_{440-675nm}$	0.017 (1.4%)	0.15 (12.0%)	0.85

The coefficient of determination (R^2) ranged between 0.79 and 0.86 for AODs and equaled 0.85 for $AE_{440-675nm}$, revealing an adequately good linear relationship between the ML-based AOPs from ASI and AERONET (Table 1).

4.2. DNI Estimations and Uncertainties

In the second section of the results, the DNI is calculated using the libRadtran software package and the UVSPEC radiative transfer code [12,13]. In the libRadtran simulations, the radiative transfer equation is numerically solved using the SDISORT radiative transfer solver, which includes a pseudospherical approximation [14]. The libRadtran simulations

were calculated using a band parameterization based on an optimized version of the correlated-k approximation [15], KATO2. Regarding aerosol, the default background aerosol model of Shettle (1989) [16] is used, considering constant climatological values of AERONET single scattering albedo, asymmetry parameter, and ozone across the study station, while the AE and AOD are applied from AERONET measurements and ML retrievals. The DNI estimations using the retrieved ML and AERONET AOPs are hereafter abbreviated as “DNI ML” and “DNI AERONET,” respectively.

Figure 2a illustrates the linear relationship between the ML and AERONET DNI. An R^2 of 0.90 is found, revealing an adequately good linear relationship. The MBE statistic revealed a modest bias (0.54 W m^{-2}), indicating a small overestimation of the ML DNI compared to AERONET. The highest discrepancies between ML and AERONET DNI are observed for high AOD cases (Figure 2a), providing a relatively high DNI overestimation due to a relatively high underestimation of AOD values.

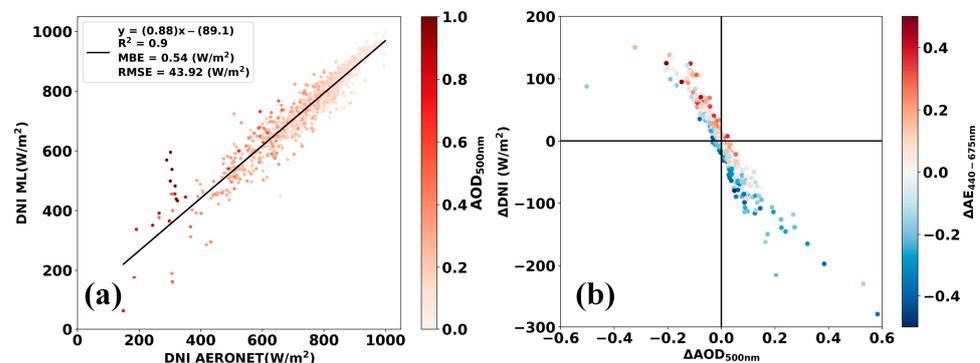


Figure 2. (a) Linear relationship between the retrieved and AERONET DNI. The color bar presents the AERONET AOD at 500 nm. (b) Linear relationship between DNI discrepancies ($\Delta\text{DNI} = \text{DNI}_{\text{AERONET}} - \text{DNI}_{\text{ML}}$) and AOD_{500nm} discrepancies ($\Delta\text{AOD} = \text{AOD}_{\text{AERONET}}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

Figure 2b depicts the linear relationship between the discrepancies of DNI ($\Delta\text{DNI} = \text{DNI}_{\text{AERONET}} - \text{DNI}_{\text{ML}}$) and AOD_{500nm} ($\Delta\text{AOD} = \text{AOD}_{\text{AERONET}} - \text{AOD}_{\text{ML}}$). Figure 2b depicts the differences between AERONET and ML, AOD, and AE and the resulting differences in DNI when using the simulated AOPs instead of AERONET. It is apparent that the AOD retrieval performance regulates the ΔDNI . In particular, ML AOD underestimation ($\Delta\text{AOD} > 0$) or overestimation ($\Delta\text{AOD} < 0$) reflects negative or positive ΔDNI up to 300 and 150 W m^{-2} , respectively. Regarding AE, a negative $\Delta\text{AE}_{440-675\text{nm}}$ leads to a more negative ΔDNI , and vice versa.

5. Conclusions

In this study, an ML-based approach is evaluated to simulate AOPs using the sky information captured from an ASI. Auxiliary parameters were also used, such as the water vapor column and solar geometry. Considering the adequate ML AOP simulation performance, the potential use of them for DNI estimations was investigated. ML DNI estimations revealed quite good results with a Pearson correlation coefficient of 0.95 and biases around 0.54 W m^{-2} , presenting the applicability of ML AOPs for solar resource estimation with small uncertainties.

Author Contributions: Conceptualization, S.-A.L., A.K. and V.S.; methodology, S.-A.L. and V.S.; software, S.-A.L.; validation, S.-A.L.; formal analysis, S.-A.L.; investigation, S.-A.L., V.S. and A.K.; data curation, S.-A.L., P.T. and C.-P.G.; writing—original draft preparation, S.-A.L.; writing—review and editing, S.-A.L., C.-P.G., V.S., K.E. and A.K.; visualization, S.-A.L.; resources, P.-I.R., K.E. and V.A.; supervision, A.K.; funding acquisition, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: The publication of this article has been 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).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. IPCC, 2021: Climate Change 2021: The Physical Science Basis. In *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S.L.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; Gomis, M.I.; et al. (Eds.) Cambridge University Press: Cambridge, UK; New York, NY, USA, 2021; *in press*. [[CrossRef](#)]
2. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanré, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [[CrossRef](#)]
3. Giles, D.M.; Sinyuk, A.; Sorokin, M.G.; Schafer, J.S.; Smirnov, A.; Slutsker, I.; Eck, T.F.; Holben, B.N.; Lewis, J.R.; Campbell, J.R.; et al. Advancements in the Aerosol Robotic Network (AERONET) Version 3 database—Automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Meas. Tech.* **2019**, *12*, 169–209. [[CrossRef](#)]
4. Sinyuk, A.; Holben, B.N.; Eck, T.F.; Giles, D.M.; Slutsker, I.; Korokin, S.; Schafer, J.S.; Smirnov, A.; Sorokin, M.; Lyapustin, A. The AERONET Version 3 aerosol retrieval algorithm, associated uncertainties and comparisons to Version 2. *Atmospheric Meas. Tech.* **2020**, *13*, 3375–3411. [[CrossRef](#)]
5. Remer, L.A.; Kleidman, R.G.; Levy, R.C.; Kaufman, Y.J.; Tanré, D.; Mattoo, S.; Martins, J.V.; Ichoku, C.; Koren, I.; Yu, H.; et al. Global aerosol climatology from the MODIS satellite sensors. *J. Geophys. Res. Atmos.* **2008**, *113*. [[CrossRef](#)]
6. Levy, R.C.; Mattoo, S.; Munchak, L.A.; Remer, L.A.; Sayer, A.M.; Patadia, F.; Hsu, N.C. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* **2013**, *6*, 2989–3034. [[CrossRef](#)]
7. Sayer, A.M.; Hsu, N.C.; Bettenhausen, C.; Jeong, M.J. Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data. *J. Geophys. Res. Atmos.* **2013**, *118*, 7864–7872. [[CrossRef](#)]
8. Kahn, R.A.; Gaitley, B.J.; Garay, M.J.; Diner, D.J.; Eck, T.F.; Smirnov, A.; Holben, B.N. Multiangle Imaging Spectroradiometer global aerosol product assessment by comparison with the Aerosol Robotic Network. *J. Geophys. Res. Earth Surf.* **2010**, *115*, D23209. [[CrossRef](#)]
9. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3147–3155.
10. Inness, A.; Ades, M.; Agustí-Panareda, A.; Barré, J.; Benedictow, A.; Blechschmidt, A.-M.; Dominguez, J.J.; Engelen, R.; Eskes, H.; Flemming, J.; et al. The CAMS reanalysis of atmospheric composition. *Atmospheric Meas. Tech.* **2019**, *19*, 3515–3556. [[CrossRef](#)]
11. Wriedt, T. Mie Theory: A Review. In *The Mie Theory. Springer Series in Optical Sciences*; Hergert, W., Wriedt, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 169. [[CrossRef](#)]
12. Mayer, B.; Kylling, A. Technical note: The libRadtran software package for radiative transfer calculations—Description and examples of use. *Atmos. Chem. Phys.* **2005**, *5*, 1855–1877. [[CrossRef](#)]
13. Emde, C.; Buras-Schnell, R.; Kylling, A.; Mayer, B.; Gasteiger, J.; Hamann, U.; Kylling, J.; Richter, B.; Pause, C.; Dowling, T.; et al. The libRadtran software package for radiative transfer calculations (version 2.0.1). *Geosci. Model Dev.* **2016**, *9*, 1647–1672. [[CrossRef](#)]
14. Dahlback, A.; Stamnes, K. A new spherical model for computing the radiation field available for photolysis and heating at twilight. *Planet. Space Sci.* **1991**, *39*, 671–683. [[CrossRef](#)]

15. Kato, S.; Ackerman, T.P.; Mather, J.H.; Clothiaux, E. The k-distribution method and correlated-k approximation for a shortwave radiative transfer model. *J. Quant. Spectrosc. Radiat. Transf.* **1999**, *62*, 109–121. [[CrossRef](#)]
16. Shettle, E.P. Models of aerosols, clouds and precipitation for atmospheric propagation studies. In Proceedings of the AGARD Conference on Atmospheric Propagation in the UV, Visible, IR and mm-region and Related System Aspects, Copenhagen, Denmark, 9–13 October 1989.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.