



Proceeding Paper

PM_{2.5} Retrieval Using Aerosol Optical Depth, Meteorological Variables, and Artificial Intelligence [†]

Stavros-Andreas Logothetis ^{1,*}, Georgios Kosmopoulos ¹, Vasileios Salamalakis ² and Andreas Kazantzidis ¹

¹ Laboratory of Atmospheric Physics, Physics Department, University of Patras, GR-26500 Patras, Greece; giokosmopoulos@gmail.com (G.K.); akaza@upatras.gr (A.K.)

² NILU—Norwegian Institute for Air Research, 2027 Kjeller, Norway; vsal@nilu.no

* 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: Particulate matter (PM) is one of the major air pollutants that has adverse impacts on human health. The aim of this study is to present an alternative approach for retrieving fine PM (particles with an aerodynamic diameter less than 2.5 μm , PM_{2.5}) using artificial intelligence. Ground-based instruments, including a hand-held Microtops II sun photometer (for aerosol optical depth), a PurpleAir sensor (for PM_{2.5}), and Rotronic sensors (for temperature and relative humidity), are used for the machine learning algorithm training. The retrieved PM_{2.5} reveals an adequate performance with an error of 0.08 $\mu\text{g m}^{-3}$ and a Pearson correlation coefficient of 0.84.

Keywords: particular matter; aerosol optical depth; artificial intelligence; machine learning; PM_{2.5} retrieval

1. Introduction

Particulate matter (PM)-related air pollution is a major environmental risk affecting human health and the environment [1]. Thus, precise knowledge of PM mass concentration spatiotemporal distribution is vital to quantitatively assessing its impact on the environment and investigating the health risks for the public [2]. Current conventional reference grade instruments face several limitations, mainly due to their increased installation and operation costs. Therefore, regulatory monitoring sites' density is impeded, and they are unable to capture the small-scale variations of PM concentrations across complex environments. Recent advance in electronics facilitates the assessment of PM monitoring techniques using low-cost and portable sensing modules. Low-cost sensor technologies constitute a promising tool to supplement and enhance the spatiotemporal resolution of existing PM monitoring networks.

During the last two decades, new alternative techniques for retrieving the spatiotemporal distribution of PM_{2.5}, have rapidly increased, using the relationship between satellite-based AOD and PM_{2.5} in conjunction with advanced mathematical methods [3]. Some of the most frequently implemented methods are multiple linear regression models [4] and machine learning (ML) algorithms such as artificial neural networks [5], support vector machines [6], and random forest [7,8]. The accuracy of PM_{2.5} estimations is related to the uncertainties that are induced by satellite AOD products. In addition, since AOD measurements from satellites are available 1–2 times per day, PM_{2.5} retrievals are provided solely on a daily basis.

In this work, an alternative machine learning methodology for retrieving PM_{2.5} is proposed, taking into account for the first time the importance of applying the AOD to various spectral channels along with several meteorological variables using quality-assured data from ground-based instruments.



Citation: Logothetis, S.-A.; Kosmopoulos, G.; Salamalakis, V.; Kazantzidis, A. PM_{2.5} Retrieval Using Aerosol Optical Depth, Meteorological Variables, and Artificial Intelligence. *Environ. Sci. Proc.* **2023**, *26*, 136. <https://doi.org/10.3390/environsciproc2023026136>

Academic Editors: Konstantinos Moustiris and Panagiotis Nastos

Published: 31 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/>).

2. Data

The data used in this study were collected at the Laboratory of Atmospheric Physics at the University of Patras (38.291° N, 21.789° E) and were divided into three main categories. The first category includes aerosol optical properties such as aerosol optical depth (AOD) at four spectral channels (e.g., 440, 500, 675, and 870 nm) as collected from a hand-held Microtops II (MII) sun photometer. MII retrieves columnar AOD using the Bouguer-Lambert-Beer law [9]. All the MII measurements were acquired under cloud-free conditions at a 30 min resolution.

The second category includes calibrated PM_{2.5} measurements from a PurpleAir-II low-cost particle concentration sensor (PAir). PAir monitors integrate a set of PMS 5003 sensors (Plantower Co., Ltd., Beijing, China) and conduct simultaneous PM concentration measurements at approximately 2 min temporal resolution. PMS sensors' operation is based on particle light scattering principles and reports the size distribution of particles, with a diameter ranging between 0.3 and 10 µm, and the mass concentration of PM₁, PM_{2.5}, and PM₁₀. They are equipped with a built-in fan that draws ambient air (flow rate: 0.1 L min⁻¹), and a laser at 680 nm wavelength that is used as the light source. Particles pass through the laser beam and the scattered light is collected by a photodetector; a proprietary algorithm is used to determine PM mass concentrations based on the output signal. PAir sensors' sensitivity and reliability have been widely investigated during the last few years, exhibiting good performance and long-term performance stability [10–12]. Low-cost sensors, however, require site-specific calibration to assure good data quality [13,14]. In this work, PAir PM_{2.5} values were corrected by implementing a calibration method proposed by [15] that is appropriate for the examined area.

The third data category contains meteorological data, ambient temperature (T), and relative humidity (RH) obtained from Rotronic sensors (MP101A-T7-W4W) at the automatic weather station located at the University campus in Patras, Greece. Within the study period, 1767 measurements were acquired, spanning from 04/2021 to 10/2022. The meteorological and PM_{2.5} data were temporally aggregated within the time window of 2 min (±1 min) centered over the MII timestamp.

3. Methodology

The PM_{2.5} is retrieved based on the following parameters: (1) AOD at four spectral channels (440, 500, 670, and 870 nm), (2) T, and (3) RH. AOD is an adequate variable in terms of capturing the intra-day variations of PM_{2.5} mass concentrations since aerosol emissions, dynamical transport, etc., will affect both parameters. The whole dataset, which consists of the previous parameters, has initially been separated into two datasets: the train and the test, which include 70% and 30% of the whole dataset, respectively. For the sake of this study, an ensemble technique, the random forest (RF), is adapted. RF presents a very effective supervised machine learning algorithm that can produce very accurate predictions in large datasets, either for classification or regression tasks. In this study, the RF is used for regression. Thus, the train dataset is applied to train the RF algorithm.

In order to achieve optimal accuracy, a randomized search procedure was performed during the training in order 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 RF algorithm, the RF scheme with the highest performance, including the best combination of hyperparameters, is implemented to evaluate the test dataset.

4. Results

4.1. Descriptive Statistics

Based on Table 1, the minimum and maximum values of PM_{2.5} ranged from 0.37 to 18.76 µg m⁻³, with a mean of 4.72 µg m⁻³, highlighting the modest level of pollution across the study station. During the same period, the mean AOD values ranged between 0.11 and 0.21. The city of Patras, located in southern Europe, is frequently affected by dust particles transported from the Sahara Desert, recording high levels of AOD (maximum

values 0.93–1.10). Nevertheless, fine particles are dominant across the area revealing a mean $AE_{440-870nm}$ (Ångström Exponent between 440 and 870 nm) of 1.41. The $AE_{440-870nm}$ from MII is computed using the Ångström power formula from the corresponding AOD channels. T and RH values ranged between 4.40–39.70 °C and 11.80–89.80% with average values of 24.26 °C and 45.36%, respectively.

Table 1. Minimum, maximum, and average values of ML algorithm input parameters.

| Variables | Minimum | Maximum | Mean |
|--|---------|---------|-------|
| PM _{2.5} (µgm ⁻³) | 0.37 | 18.76 | 4.72 |
| AOD _{440nm} | 0.031 | 1.10 | 0.25 |
| AOD _{500nm} | 0.027 | 1.02 | 0.21 |
| AOD _{675nm} | 0.021 | 0.97 | 0.15 |
| AOD _{870nm} | 0.013 | 0.93 | 0.11 |
| $AE_{440-870nm}$ | 0.15 | 2.21 | 1.41 |
| T (°C) | 4.40 | 39.70 | 24.26 |
| RH | 11.80 | 89.80 | 45.36 |

4.2. Machine Learning Algorithm Performance

In order to investigate the different effects of spectral AOD and meteorological variables on model retrieval performance, a sensitivity analysis of the input parameters was performed during the training of the RF algorithm. In total, 15 different cases were applied, with the aerosol optical properties as a baseline (Table 2). The first scenario (Scenario 1) consisted of five different sub-scenarios. Scenario 1.1 included solely the AOD_{440nm} as an input parameter for the RF algorithm training, whereas for scenario 1.2 the AOD_{500nm} was included, and so on for the rest of the sub-scenarios. Thus, scenario 1.5 included the AOD at four MII spectral channels and $AE_{440-870nm}$. The cases in Scenarios 2 and 3 are similar to Scenario 1 but included T and RH, respectively, as input parameters.

Table 2. Scenarios applied during the RF algorithm training procedure.

| Scenario 1: Only aerosol optical properties | | | | |
|---|-------------------------------|-------------------------------------|---|--|
| 1.1 | 1.2 | 1.3 | 1.4 | 1.5 |
| AOD _{440nm} | AOD _{440 and 500 nm} | AOD _{440, 500, and 675 nm} | AOD _{440, 500, 675 and 870 nm} | AOD _{440, 500, 675, and 870 nm and $AE_{440-870nm}$} |
| Scenario 2: Aerosol optical properties and ambient temperature | | | | |
| 2.1 | 2.2 | 2.3 | 2.4 | 2.5 |
| 1.1 and T | 1.2 and T | 1.3 and T | 1.4 and T | 1.5 and T |
| Scenario 3: Aerosol optical properties, ambient temperature and relative humidity | | | | |
| 3.1 | 3.2 | 3.3 | 3.4 | 3.5 |
| 1.1, T and RH | 1.2, T and RH | 1.3, T and RH | 1.4, T and RH | 1.5, T and RH |

Figure 1 illustrates the findings of the sensitivity analysis for the 15 different training scenarios. In the literature, the majority of the studies dedicated to PM_{2.5} retrieval via ML use satellite based AOD at a specific channel. In this study, firstly the effect of spectral AOD information on ML algorithm performance (Scenario 1) is investigated, and it is apparent that the performance of the ML algorithm increases as more spectral channels of AOD are included. In particular, the MAE (RMSE) values range from 1.76 µg m⁻³ (2.25 µg m⁻³) to 1.10 µg m⁻³ (1.53 µg m⁻³). In terms of correlation coefficient (R), the ML algorithm performance increased substantially by including all four spectral channels of AOD (from 0.45 to 0.78). The effect of $AE_{440-870nm}$ was marginal for all scenarios. In total, including all spectral AOD channels, the Mean Absolute Error (MAE) (Root Mean Square Error (RMSE)) was suppressed by ~38% (~32) compared to when using only AOD_{440nm}.

Secondly, the effect of two meteorological parameters on ML performance was investigated together with AOD (Table 1). By including T (Scenario 2) in ML training, an increase in the model’s performance was revealed, reducing the MAE (RMSE) from

1.46 $\mu\text{g m}^{-3}$ (1.90 $\mu\text{g m}^{-3}$) to 0.97 $\mu\text{g m}^{-3}$ (1.38 $\mu\text{g m}^{-3}$). In addition, R improved from 0.62 to 0.82. For scenario 3, RH was also included on ML training in addition to AOD and T, leading to a further improvement of the model's performance from 1.31 $\mu\text{g m}^{-3}$ (1.72 $\mu\text{g m}^{-3}$) to 0.91 $\mu\text{g m}^{-3}$ (1.30 $\mu\text{g m}^{-3}$) for MAE and RMSE, respectively, and from 0.70 to 0.84 for R. Including the two meteorological parameters, MAE (RMSE) was decreased by ~20% (~15%), compared to using the parameters of scenario 1.5. Figure 2a shows the linear relationship between the ML-based (estimations) and ground-based (measurements) $\text{PM}_{2.5}$ for the scenario with the highest accuracy (Scenario 3.5). The findings revealed a dispersion of 26.9%.

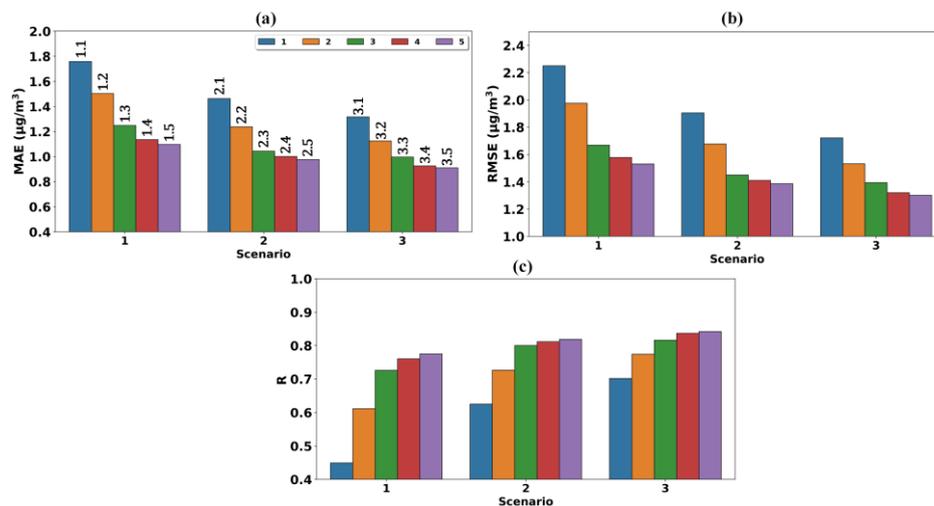


Figure 1. (a) MAE, (b) RMSE, and (c) R for the 15 scenarios. The description of each scenario is presented in Table 1.

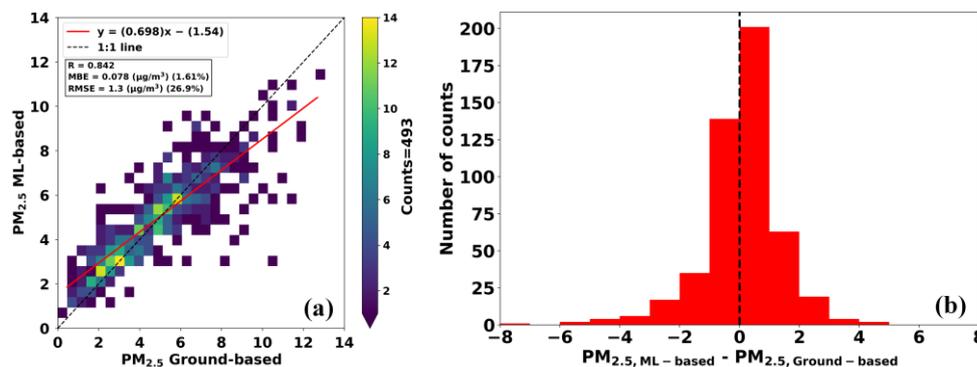


Figure 2. (a) Linear relationship and (b) frequency distribution of differences between the ML-based (estimations) and ground-based (measurements) $\text{PM}_{2.5}$ for scenario 3.5 (see Table 2).

Figure 2b depicts the frequency distribution of differences between the ML-based (estimations) and ground-based (measurements) $\text{PM}_{2.5}$ for the scenario with the highest accuracy (Scenario 3.5). For the 69% (89%) of the test dataset, the differences between the $\text{PM}_{2.5}$ estimations and measurements were lower than $1 \mu\text{g m}^{-3}$ ($2 \mu\text{g m}^{-3}$).

5. Conclusions

Quantitative and qualitative information on surface $\text{PM}_{2.5}$ mass concentration is vital for monitoring and regulating air quality. In this work, an alternative ML-based methodology relying on the synergy of ground-based AOD and meteorological measurements is proposed for retrieving $\text{PM}_{2.5}$. The most interesting finding of this study is the great improvement in ML algorithm's performance by including AOD spectral information. Moreover, the addition of two meteorological parameters, T and RH, increased the retrieval

performance of the ML algorithm. The results of the proposed methodology, due to their high temporal resolution, could be used to fill and extend either existing or missing PM_{2.5} time series derived from ground-based measurements. In addition, the retrieved PM_{2.5} can be used as a reference measurement for the validation of retrieval algorithms based on satellite measurements.

Author Contributions: Conceptualization, S.-A.L. and A.K.; methodology, S.-A.L.; software, S.-A.L.; validation, S.-A.L.; formal analysis, S.-A.L.; investigation, S.-A.L.; data curation, S.-A.L. and G.K.; writing—original draft preparation, S.-A.L. and G.K.; writing—review and editing, S.-A.L., G.K. and V.S.; visualization, S.-A.L.; 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. Lelieveld, J.; Pozzer, A.; Pöschl, U.; Fnais, M.; Haines, A.; Münzel, T. Loss of life expectancy from air pollution compared to other risk factors: A worldwide perspective. *Cardiovasc. Res.* **2020**, *116*, 1910–1917. [[CrossRef](#)]
2. Sorek-Hamer, M.; Just, A.C.; Kloog, I. Satellite remote sensing in epidemiological studies. *Curr. Opin. Pediatr.* **2016**, *28*, 228–234. [[CrossRef](#)]
3. Li, Y.; Yuan, S.; Fan, S.; Song, Y.; Wang, Z.; Yu, Z.; Yu, Q.; Liu, Y. Satellite Remote Sensing for Estimating PM_{2.5} and Its Components. *Curr. Pollut. Rep.* **2021**, *7*, 72–87. [[CrossRef](#)]
4. Gupta, P.; Christopher, S.A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *J. Geophys. Res. Atmos.* **2009**, *114*, D14205. [[CrossRef](#)]
5. Gupta, P.; Christopher, S.A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *J. Geophys. Res. Atmos.* **2009**, *114*, D20205. [[CrossRef](#)]
6. de Hoogh, K.; Héritier, H.; Stafoggia, M.; Künzli, N.; Kloog, I. Modelling daily PM_{2.5} concentrations at high spatio-temporal resolution across Switzerland. *Environ. Pollut.* **2017**, *233*, 1147–1154. [[CrossRef](#)] [[PubMed](#)]
7. Park, S.; Lee, J.; Im, J.; Song, C.-K.; Choi, M.; Kim, J.; Lee, S.; Park, R.; Kim, S.-M.; Yoon, J.; et al. Estimation of spatially continuous daytime particulate matter concentrations under all sky conditions through the synergistic use of satellite-based AOD and numerical models. *Sci. Total. Environ.* **2020**, *713*, 136516. [[CrossRef](#)] [[PubMed](#)]
8. Ghahremanloo, M.; Choi, Y.; Sayeed, A.; Salman, A.K.; Pan, S.; Amani, M. Estimating daily high-resolution PM_{2.5} concentrations over Texas: Machine Learning approach. *Atmos. Environ.* **2021**, *247*, 118209. [[CrossRef](#)]
9. Hadjimitsis, D.-G.; Mamouri, R.-E.; Nisantzi, A.; Kouremerti, N.; Retalis, A.; Paronis, D.; Tymvios, F.; Perdikou, S.; Achileos, S.; Hadjicharalambous, M.; et al. Air Pollution from Space. In *Remote Sensing of Environment*; IntechOpen: Rijeka, Croatia, 2013.
10. Sayahi, T.; Butterfield, A.; Kelly, K.E. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* **2019**, *245*, 932–940. [[CrossRef](#)] [[PubMed](#)]
11. Wallace, L.; Bi, J.; Ott, W.R.; Sarnat, J.; Liu, Y. Calibration of low-cost PurpleAir outdoor monitors using an improved method of calculating PM. *Atmos. Environ.* **2021**, *256*, 118432. [[CrossRef](#)]
12. Kosmopoulos, G.; Salamalikis, V.; Matralli, A.; Pandis, S.N.; Kazantzidis, A. Insights about the Sources of PM_{2.5} in an Urban Area from Measurements of a Low-Cost Sensor Network. *Atmosphere* **2022**, *13*, 440. [[CrossRef](#)]
13. Stavroulas, I.; Grivas, G.; Michalopoulos, P.; Liakakou, E.; Bougiatioti, A.; Kalkavouras, P.; Fameli, K.M.; Hatzianastassiou, N.; Mihalopoulos, N.; Gerasopoulos, E. Field Evaluation of Low-Cost PM Sensors (Purple Air PA-II) Under Variable Urban Air Quality Conditions, in Greece. *Atmosphere* **2020**, *11*, 926. [[CrossRef](#)]

14. Giordano, M.R.; Malings, C.; Pandis, S.N.; Presto, A.A.; McNeill, V.; Westervelt, D.M.; Beekmann, M.; Subramanian, R. From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors. *J. Aerosol. Sci.* **2021**, *158*, 105833. [[CrossRef](#)]
15. Kosmopoulos, G.; Salamalikis, V.; Pandis, S.; Yannopoulos, P.; Bloutsos, A.; Kazantzidis, A. Low-cost sensors for measuring airborne particulate matter: Field evaluation and calibration at a South-Eastern European site. *Sci. Total. Environ.* **2020**, *748*, 141396. [[CrossRef](#)] [[PubMed](#)]

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.