$PM_{2.5}$ retrieval using aerosol optical depth, meteorological variables, and artificial intelligence

Logothetis S.A.[1]*, Kosmopoulos G.[1], Salamalikis V.[2], and Kazantzidis A.[1]

[1]Laboratory of Atmospheric Physics, Physics Department, University of Patras | [2]NILU-Norwegian Institute for Air Research, P.O. Box 100, 2027 Kjeller, Norway | *Correspondence: logothetis_s@upnet.gr

Objective

The aim of the present study is to present an alternative machine learning methodology for retrieving $PM_{2.5}$ taking into account the aerosol optical depth (AOD) to various spectral channels, Ångström Exponent ($AE_{440\text{-}870nm}$), and several meteorological variables.

Data

Aerosol burden and size	Meteorological Variables	Particulate Matter
AOD_{440nm}, AOD_{500nm}, AOD_{670nm}, AOD_{870nm}, $AE_{440\text{-}870nm}$	Temperature (T), Relative Humidity (RH)	$PM_{2.5}$
Hand-held Microtops II (MII) sun photometer	Rotronic sensors (MP101A-T7-W4W)	PurpleAir-II low-cost sensor (PAir)

Data Availability: 1767 measurements, spanning from 04/2021 to 10/2022 at Laboratory of Atmospheric Physics, University of Patras.

Methodology

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

[2] Supervised Machine Learning Algorithm: The Random Forest (RF) 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 hyperparameters, including a 10-fold cross-validation process using the mean square error as a loss function. The RF scheme with the highest performance, including the best combination of hy-perparameters, is implemented to evaluate the test dataset.

[3] Data Importance: A sensitivity analysis of the input parameters is performed during the training of the RF algorithm. (Table 1)

Table 1. Scenarios applied during the RF algorithm training procedure in order to retrieve $PM_{2.5}$.

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

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).

Co-financed by Greece and the European Union

Results: Sensitivity Analysis

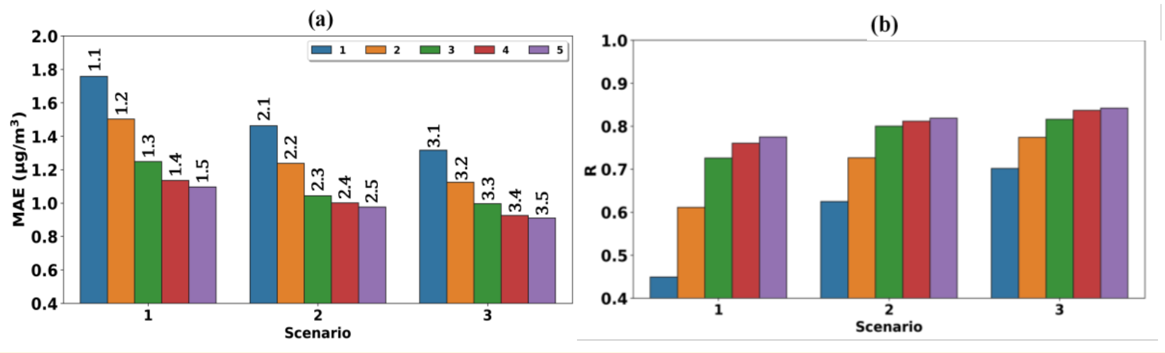

Figure 1: **(a)** Mean Absolute Error, MAE, and **(b)** correlation coefficient, R, for the 15 scenarios. The description of each scenario is presented in Table 1.

- ✓ Including all spectral AOD channels, the MAE was suppressed by ~38% compared to using only AOD_{440nm}.
- ✓ In terms of R, the ML algorithm performance increased substantially by including all four spectral channels of AOD (from 0.45 to 0.78). The effect of $AE_{440\text{–}870nm}$ was marginal for all scenarios.

Results: Machine Learning Performance

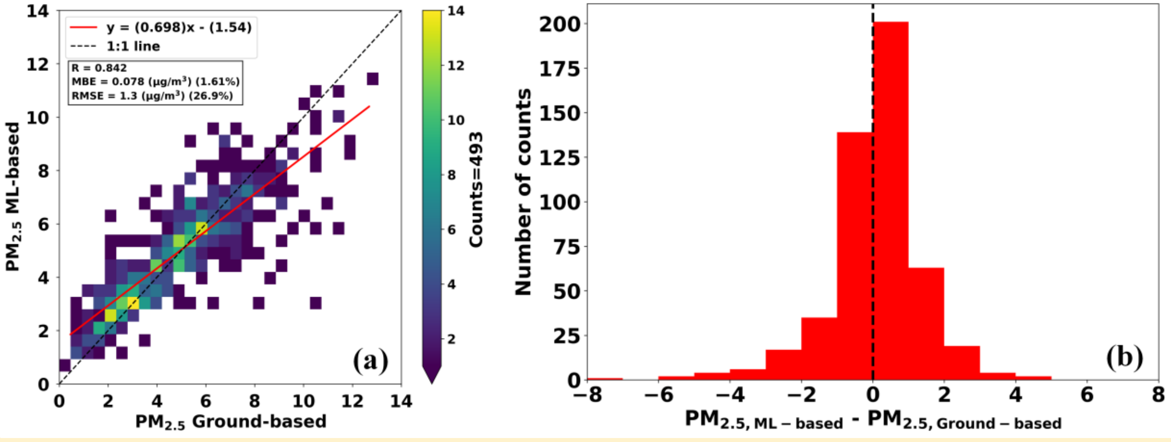

Figure 2: **(a)** Linear relationship and **(b)** frequency distribution of differences between the ML-based (estimations) and ground-based (measurements) $PM_{2.5}$ for scenario 3.5 (see Table 1).