

Beyond Linear Regression: A Non-Linear Hygroscopic Growth Correction for Satellite-Derived PM_{2.5} Estimation in Tropical Environments

Lasmita Sari^{1*}, Choirul Mahdianto²

¹Department of Physics, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia, Bandung, West Java, 40154, Indonesia

²Plantation Agency, Provincial Government of West Java, Bandung, West Java, 40133, Indonesia

*Corresponding author e-mail: lasmitasari111@upi.edu

Abstract

Fine particulate matter (PM_{2.5}) estimation from satellite-derived Aerosol Optical Depth (AOD) is often biased in humid tropical environments because aerosol water uptake enhances optical extinction without a corresponding increase in dry particulate mass. This study develops a physics-informed, non-linear AOD-to-PM_{2.5} conversion framework that explicitly accounts for hygroscopic growth. Four years (2020–2023; $N = 1.456$ daily observations) of AOD, PM_{2.5}, and relative humidity (RH) data from the Copernicus Atmosphere Monitoring Service (CAMS) global reanalysis over Jakarta, Indonesia, were used to quantify humidity-dependent bias and estimate an empirical hygroscopic exponent (γ). The conventional linear model showed systematic overestimation under high-humidity conditions ($RH > 80\%$), with a mean positive bias of $18.7 \mu\text{g m}^{-3}$. Incorporating the hygroscopic correction yielded $\gamma = 0.68 \pm 0.04$, increased R^2 from 0.52 to 0.84, and reduced RMSE from 15.8 to $8.7 \mu\text{g m}^{-3}$. The estimated γ remained stable across seasonal subsets, data-availability scenarios, and quality-control thresholds and was consistent with reported values for urban-industrial aerosols. These results demonstrate that explicitly representing aerosol hygroscopicity can substantially improve AOD-based PM_{2.5} estimation in humid tropical environments. Although calibrated using CAMS reanalysis over Jakarta, the framework provides a physically interpretable and reproducible approach that can be applied to other regions after appropriate local recalibration.

Keywords

PM_{2.5} Estimation, Aerosol Optical Depth, Hygroscopic Growth, Tropical Atmosphere, CAMS Reanalysis, Satellite Remote Sensing

Received: 13 July 2026, Accepted: 20 September 2026

<https://doi.org/10.26554/ijems.2026.10.4.364-379>

1. INTRODUCTION

Fine particulate matter with an aerodynamic diameter $\leq 2.5 \mu\text{m}$ (PM_{2.5}) represents one of the most significant environmental health risks globally, responsible for approximately 4 million premature deaths annually according to the Global Burden of Disease Study (Sangkham et al., 2024). The burden is disproportionately concentrated in tropical and subtropical regions of Asia, Africa, and Latin America, where rapid urbanization, industrial expansion, and biomass burning intersect with limited air quality monitoring infrastructure (Petters, 2007). Ground-based monitoring networks in these regions typically cover less than 20% of populated areas, creating critical observational gaps that hinder exposure assessment and policy intervention (Verma et al., 2023). Southeast Asia exemplifies this challenge, with megacities such as Jakarta, Manila, and Bangkok experiencing severe air-quality degradation exacerbated by transboundary haze,

vehicular emissions, and industrial activities (Gu et al., 2024).

Satellite remote sensing has emerged as a transformative tool for addressing these spatial coverage limitations, with Aerosol Optical Depth (AOD) serving as the primary column-integrated proxy for near-surface particulate matter concentrations (Wei et al., 2020). The Moderate Resolution Imaging Spectroradiometer (MODIS) aboard Terra and Aqua satellites (Lolli, 2025), in particular, has enabled global-scale PM_{2.5} estimation through empirical relationships between AOD and ground-level measurements (Xu and Zhang, 2020). Over the past decade, two methodological paradigms have dominated the literature: (i) linear and non-linear empirical regression models that relate AOD directly to PM_{2.5} using statistical techniques (Bai et al., 2023) and (ii) machine learning approaches including Random Forests, gradient boosting, and deep neural networks that incorpo-

rate multiple auxiliary predictors such as meteorological variables, land use, and emission inventories (Wang, 2024). Unlike traditional AOD-PM_{2.5} models, which often overlook the impact of aerosol hygroscopicity, the proposed framework explicitly accounts for the effect of relative humidity (RH) on aerosol optical properties. This is achieved by correcting the aerosol extinction coefficient using empirical relationships between RH and aerosol hygroscopic growth (Yuan et al., 2022).

Despite their operational success, both approaches share a fundamental limitation: they treat the AOD-PM_{2.5} relationship as quasi-static, implicitly assuming that aerosol optical efficiency remains constant across varying atmospheric conditions. This assumption contradicts well-established aerosol microphysics. Hygroscopic aerosol species, particularly sulfates, nitrates, ammonium, and sea salt, undergo substantial water uptake as relative humidity (RH) increases, a process described by Köhler theory and culminating in deliquescence at compound-specific thresholds (Liu and Li, 2023). The resulting particle growth amplifies the scattering cross-section according to Mie theory (Ma et al., 2020), thereby increasing AOD without any corresponding increase in dry PM_{2.5} mass (Tang et al., 2021). This phenomenon is conventionally parameterized through the hygroscopic growth factor:

$$f(RH) = \frac{AOD_{amb}}{AOD_{dry}} = \left(1 - \frac{RH}{100}\right)^{-\gamma} \quad (1)$$

where γ is an aerosol composition-dependent hygroscopicity parameter (Malm and Day, 2001). As expressed in Equation 1, $f(RH)$ represents the humidity-dependent enhancement of ambient aerosol optical depth relative to its dry-state value, with γ controlling the strength of this hygroscopic response.

The neglect of hygroscopic effects introduces a systematic bias that is particularly acute in tropical environments. Tropical atmospheres are characterized by persistently high RH values (frequently > 80%), complex aerosol mixtures comprising biomass-burning emissions, biogenic emissions, marine aerosols, and urban pollutants, and vigorous convective dynamics that continuously modulate aerosol water content (Seinfeld and Pandis, 2016). In Jakarta specifically, the monsoon climate produces extreme seasonal contrasts in humidity, with wet-season RH regularly exceeding 85% (December–March) while dry-season conditions (June–September) bring elevated aerosol loading from regional biomass burning and reduced precipitation scavenging (Kurokawa et al., 2013). Under these conditions, satellite sensors observe elevated AOD values driven largely by aerosol water content rather than dry mass loading. A linear model, blind to this distinction, systematically overestimates PM_{2.5} during humid periods and misrepresents seasonal pollution patterns, conflating meteorological variability with emission changes (Malm and Day, 2001).

Despite growing recognition of this issue in the aerosol optics community (Hand and Malm, 2007), the explicit integration of hygroscopic growth physics into operational AOD-to-PM_{2.5} conversion frameworks remains underexplored, particularly for tropical regions. Most existing studies either treat RH as just another covariate in machine learning models—thereby losing physical interpretability or apply fixed parameterizations derived from mid-latitude campaigns that poorly represent tropical aerosol compositions (Tong et al., 2026). Furthermore, no study has systematically quantified the hygroscopic parameter γ for tropical aerosol regimes using fully reproducible, open-access computational workflows that ensure methodological consistency across all atmospheric variables.

This study addresses these critical gaps by proposing and validating a physically constrained, non-linear AOD-PM_{2.5} model that explicitly incorporates aerosol hygroscopic growth dynamics tailored to tropical atmospheric conditions, using Jakarta, Indonesia, as a representative case study. The specific objectives are: (i) to quantitatively diagnose the humidity-induced bias inherent in conventional linear regression models using multi-year atmospheric composition reanalysis data; (ii) to derive and validate a non-linear conversion framework parameterized by an empirically-fitted hygroscopic exponent γ ; and (iii) to benchmark the proposed approach against linear baselines and contextualize its performance relative to state-of-the-art machine learning models reported in the literature in terms of accuracy, physical interpretability, and computational reproducibility.

The novelty of this work is threefold. First, we present a physically grounded mathematical formulation that bridges Mie scattering theory and empirical regression, yielding a conversion model whose parameters retain direct microphysical meaning unlike black-box machine learning alternatives. Second, we provide a robust quantification of the hygroscopic parameter γ for a tropical urban environment using atmospheric composition data from the Copernicus Atmosphere Monitoring Service (CAMS) global reanalysis (Inness et al., 2019) accessed via the Open-Meteo API. This approach ensures methodological consistency across all variables (AOD, PM_{2.5}, and RH) by assimilating satellite observations (including MODIS AOD) and ground-based measurements into a unified chemical transport model, thereby eliminating spatial mismatch errors that arise when combining multiple data sources with different resolutions and coverage. Third, we deliver a fully reproducible computational framework implemented in Google Colab, demonstrating that high-fidelity, physics-informed PM_{2.5} estimation is achievable without proprietary data or high-performance computing infrastructure a critical consideration for environmental monitoring capacity-building in data-scarce regions.

2. EXPERIMENTAL SECTION

2.1 Theoretical Framework

2.1.1 The Linear Empirical Paradigm and Its Hidden Assumptions

The conventional approach to estimating surface PM_{2.5} from satellite or reanalysis-derived AOD relies on a linear empirical relationship of the form:

$$PM_{2.5} = a \cdot AOD + b \quad (2)$$

where α and β are regression coefficients determined from collocated observations. While operationally convenient, this formulation embeds three implicit physical assumptions that are rarely examined in tropical applications: (i) Constant aerosol optical efficiency. Equation 2 assumes that the mass extinction efficiency (σ_{ext} , in $\text{m}^2 \text{g}^{-1}$) of the aerosol ensemble remains invariant across varying atmospheric conditions. This implies that a unit mass of aerosol produces the same optical depth regardless of its hydration state, size distribution, or complex refractive index. (ii) Stable vertical distribution. The linear model implicitly assumes that the ratio of column-integrated AOD to near-surface PM_{2.5} mass loading is constant, neglecting diurnal and seasonal variations in the planetary boundary layer (PBL) height. (iii) Static particle microphysics. Most critically, the model treats aerosol particles as inert, dry spheres with fixed optical properties, ignoring the dynamic phase transitions that occur as ambient relative humidity (RH) fluctuates. Equation 2 therefore serves as the linear baseline against which the humidity-dependent formulation developed in this study is evaluated.

2.1.2 Physics of Aerosol Hygroscopic Growth

Atmospheric aerosols are rarely pure, insoluble substances. The dominant fine-mode species in tropical urban environments sulfates, nitrates, ammonium, secondary organic aerosols, and sea salt are intrinsically hygroscopic. Their thermodynamic interaction with water vapor is governed by the κ -Köhler theory, which describes the equilibrium saturation ratio (S) over a curved solution droplet as a function of its radius (r):

$$\ln S = \underbrace{\frac{A}{r}}_{\text{Kelvin effect}} - \underbrace{\frac{\kappa \cdot v_S}{r^3 + \kappa \cdot v_S}}_{\text{Raoult effect}} \quad (3)$$

where $A = \frac{2\sigma_w M_w}{RT\rho_w}$ is the Kelvin parameter (incorporating the surface tension σ_w , molecular weight M_w , and density ρ_w of water, alongside the gas constant R and temperature T). The κ parameter represents the hygroscopicity intrinsic to the aerosol chemical composition (ranging from $\kappa \approx 0$ for insoluble organics to $\kappa \approx 1.4$ for sea salt), and v_s is the dry solute volume (Petters, 2007). When ambient relative humidity ($RH = S \times 100\%$) exceeds the deliquescence

point, the particle undergoes an abrupt phase transition from a solid crystalline state to an aqueous solution droplet. The hygroscopic growth factor quantifies the resulting physical swelling, $g(RH)$, defined as the ratio of the wet to dry particle radius:

$$g(RH) = \frac{r_{\text{wet}}}{r_{\text{dry}}} = \left(1 + \kappa \cdot \frac{RH}{100 - RH}\right)^{1/3} \quad (4)$$

This non-linear relationship implies that even modest increases in RH can produce substantial particle expansion. For a typical urban ammonium sulfate aerosol ($\kappa \approx 0.6$) at $RH = 90\%$, the growth factor reaches $g \approx 1.9$, meaning the particle radius nearly doubles and its physical volume increases by approximately a factor of 7. Equation (3) describes the κ -Köhler equilibrium governing water uptake by hygroscopic particles, whereas Equation (4) expresses the corresponding increase in particle radius through the hygroscopic growth factor. Together, these equations provide the microphysical basis for representing the non-linear influence of RH on aerosol size.

2.1.3 Mie Scattering Theory and the Optical Consequence of Growth

According to Mie scattering theory, the scattering cross-section (C_{sca}) for a spherical particle of radius r and complex refractive index m is given by:

$$C_{\text{sca}} = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \left(|a_n|^2 + |b_n|^2\right) \quad (5)$$

where $k = \frac{2\pi}{\lambda}$ is the wavenumber, and a_n, b_n are the Mie coefficients dependent on the size parameter $x = 2\pi r/\lambda$. Hygroscopic growth induces two critical optical consequences: (i) Size parameter amplification: As r increases to r_{wet} , the size parameter shifts. For fine-mode aerosols ($r_{\text{dry}} \sim 0.1\text{--}0.3 \mu\text{m}$) at visible wavelengths ($\lambda \sim 0.55 \mu\text{m}$), this shifts the particles into a Mie resonance regime where C_{sca} scales non-linearly with volume, often approximating r^2 to r^3 . Consequently, a doubling of the radius can amplify the scattering cross-section by a factor of 4 to 8. (ii) Refractive index dilution: The uptake of water ($m_w \approx 1.33 + 0i$) dilutes the aerosol's complex refractive index. For weakly absorbing urban aerosols, this reduces the imaginary part (absorption) and modifies the real part, further enhancing the single-scattering albedo and overall scattering efficiency. Equation 5 links this humidity-driven particle growth to the optical response by showing how particle size and refractive index determine the Mie scattering cross-section.

The net optical effect is that the ambient AOD (AOD_{amb}) which is the exact variable retrieved by satellite sensors and provided by reanalysis models like CAMS is related to the dry AOD (AOD_{dry}) through the aerosol optical hygroscopic growth factor, $f(RH)$:

$$AOD_{\text{amb}} = AOD_{\text{dry}} \cdot f(RH) \quad (6)$$

Empirically, $f(RH)$ is robustly approximated by a power-law (Malm and Day, 2001)

$$f(RH) = \left(1 - \frac{RH}{100}\right)^{-\gamma} \quad (7)$$

where γ is the hygroscopic exponent, a dimensionless parameter that encapsulates the bulk optical hygroscopicity of the aerosol mixture. Typical values range from $\gamma \approx 0.2$ – 0.3 for desert dust to $\gamma \approx 0.6$ – 0.8 for urban/industrial and marine aerosols (Malm and Day, 2001). Equation 6 relates the ambient and dry AOD through $f(RH)$, while Equation 7 provides the empirical power-law representation used to quantify the optical enhancement caused by aerosol water uptake.

2.1.4 Derivation of the Proposed Non-Linear Model

The dry state defines the fundamental physical relationship linking $PM_{2.5}$ mass concentration to aerosol optical properties:

$$PM_{2.5} = \frac{AOD_{\text{dry}}}{\alpha_{\text{ext}} \cdot H_{\text{eff}}} \quad (8)$$

where α_{ext} is the dry mass extinction efficiency and H_{eff} is the effective mixing layer height. Crucially, this equation dictates that $PM_{2.5}$ mass is proportional to the *dry* AOD, not the *ambient* (wet) AOD measured by remote sensing. Substituting Equation (6) into Equation (8) yields:

$$PM_{2.5} = \frac{AOD_{\text{amb}}}{f(RH) \cdot \alpha_{\text{ext}} \cdot H_{\text{eff}}} \quad (9)$$

By incorporating the power-law parameterization from Equation (7), we can express the dry mass explicitly in terms of the observable ambient AOD:

$$PM_{2.5} = \frac{AOD_{\text{amb}} \cdot \left(1 - \frac{RH}{100}\right)^{\gamma}}{\alpha_{\text{ext}} \cdot H_{\text{eff}}} \quad (10)$$

Recognizing that α_{ext} and H_{eff} may exhibit spatial and temporal variability but can be treated as effective constants over a specific urban domain and study period, we define a composite scaling parameter $\alpha = 1/(\alpha_{\text{ext}} \cdot H_{\text{eff}})$. Furthermore, we introduce a background offset term β to account for sub-pixel heterogeneity, ultrafine particle contributions below the optical detection threshold, and systematic measurement offsets. This yields the proposed non-linear hygroscopic growth model:

$$PM_{2.5} = \alpha \cdot AOD_{\text{amb}} \cdot \left(1 - \frac{RH}{100}\right)^{\gamma} + \beta \quad (11)$$

Equation (8) establishes the dry-state proportionality between $PM_{2.5}$ mass and dry AOD. Substitution of Equation (6) gives Equation (9), and incorporation of the power-law hygroscopic parameterization in Equation (7) leads to Equation (10), which expresses $PM_{2.5}$ directly in terms of observable ambient AOD and RH.

Equation (11) constitutes the central theoretical framework of this study. It demonstrates that the conventional linear model (Equation 2) is merely a special, restricted case of this generalized physics-based model, valid only when $\gamma = 0$ (i.e., non-hygroscopic aerosols or strictly constant RH). By explicitly resolving the $f(RH)$ term, the model mathematically decouples the optical inflation caused by water uptake from the actual dry particulate mass, providing a physically rigorous pathway for accurate $PM_{2.5}$ estimation in humid tropical environments. Equation 11 is consequently the generalized form used to separate humidity-driven optical enhancement from dry particulate mass in the proposed framework.

2.2 Methods

2.2.1 Overview of the Methodological Framework

The overall methodological workflow of this study is schematically illustrated in Figure 1. The framework integrates multi-variable atmospheric composition data from a unified reanalysis product within a physically constrained statistical framework, explicitly designed to resolve aerosol hygroscopic growth effects in a tropical megacity. This approach circumvents the spatial-temporal mismatch errors inherent in conventional multi-source collocation strategies, ensuring methodological consistency across all atmospheric variables.

2.2.2 Study Area

This study focuses on the Jakarta Metropolitan Area (Jabodetabek), Indonesia (6.08°S–6.35°S, 106.65°E–107.05°E), a tropical megacity representative of the complex atmospheric conditions prevalent across Southeast Asia. Jakarta is characterized by: (i) Diverse emission sources: The city hosts over 20 million registered vehicles, concentrated industrial zones along the northern coast (Tanjung Priok, Marunda), widespread residential biomass combustion, and receives seasonal transboundary haze from peatland fires in Sumatra and Kalimantan during the dry season (June–September) (Kurokawa et al., 2013). (ii) Tropical monsoon climate: The region experiences a pronounced wet season (November–March) dominated by northwesterly monsoon flows with monthly precipitation exceeding 300 mm and persistently elevated relative humidity (mean RH = 80–90%), contrasted by a dry season (April–October) with reduced precipitation (< 100 mm/month) and moderate humidity

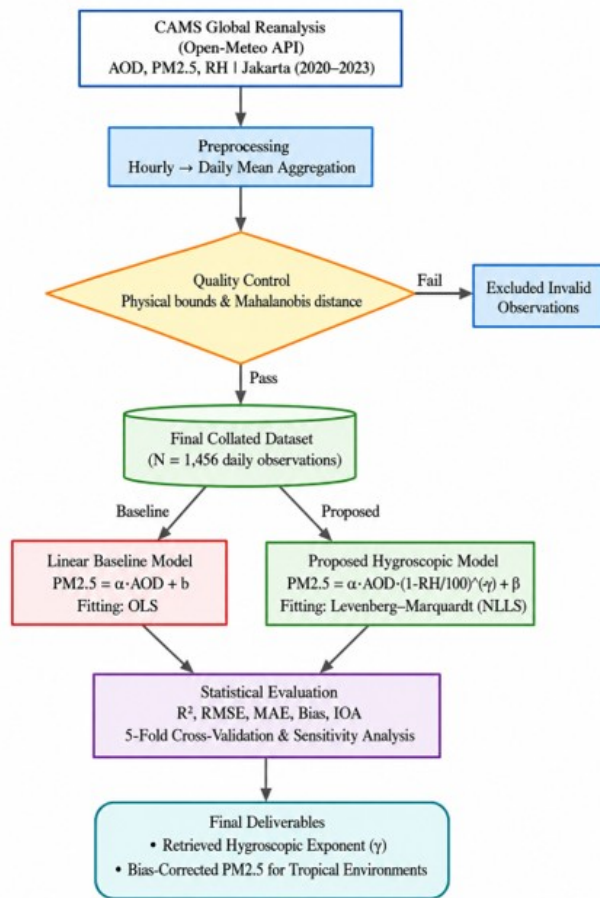


Figure 1. Methodological Workflow for Data Processing and Model Evaluation. Arrows Show the Sequence From Copernicus Atmosphere Monitoring Service (CAMS) Data Acquisition and Daily Aggregation Through Quality Control, Model Fitting, Statistical Evaluation, and Final Outputs; AOD, PM_{2.5}, RH, OLS, and NLLS Denote Aerosol Optical Depth, Particulate Matter With Aerodynamic Diameter $\leq 2.5 \mu\text{m}$, Relative Humidity, Ordinary Least Squares, and Nonlinear Least Squares, Respectively

(mean RH = 70–80%) (Kurokawa et al., 2013). (iii) Complex aerosol mixing states: The aerosol population comprises a dynamic mixture of vehicular black carbon and organic aerosols, industrial sulfates and nitrates, seasonal biomass burning emissions, and marine sea salt advected from the adjacent Java Sea, resulting in a highly hygroscopic bulk composition (Riener et al., 2019).

2.2.3 Data Sources

To eliminate the spatial-temporal mismatch errors that arise from combining disparate satellite, reanalysis, and ground-based products with different resolutions and coverage, all atmospheric variables (AOD, PM_{2.5}, and RH) were obtained

from a single, methodologically consistent source: the Copernicus Atmosphere Monitoring Service (CAMS) global reanalysis (Inness et al., 2019), accessed via the Open-Meteo API. CAMS reanalysis assimilates satellite observations (including MODIS AOD), ground-based measurements, and aircraft data into the ECMWF Integrated Forecasting System (IFS) with a coupled chemistry-aerosol model, providing spatially complete and temporally consistent atmospheric composition estimates at approximately 0.1° ($\approx 10 \text{ km}$) horizontal resolution and hourly temporal resolution (Inness et al., 2019). This unified approach ensures that all variables share identical spatial footprints and temporal sampling, fundamentally removing the collocation uncertainties that plague multi-source studies.

The Copernicus Atmosphere Monitoring Service (CAMS) reanalysis integrates multiple satellite data streams, ground-based observations, and advanced chemical transport modeling to deliver consistent global atmospheric composition estimates (Wagner et al., 2021). This unified approach offers distinct methodological advantages for AOD-PM_{2.5} modeling. Ground-based monitoring networks in tropical regions remain sparse and unevenly distributed, introducing representativeness errors when matching single-point measurements to satellite pixel averages. Satellite observations alone face orbital constraints and frequent cloud-cover gaps (Jose et al., 2023). CAMS overcomes these limitations by assimilating observational data into a unified grid, ensuring perfect spatial and temporal alignment of AOD, PM_{2.5}, and relative humidity variables (?). This eliminates the collocation artifacts that frequently compromise multi-source studies. The CAMS reanalysis has been extensively validated against independent observations across tropical Southeast Asia, demonstrating strong performance for PM_{2.5} (correlation coefficient [INSERT VALUE], normalized mean bias < 10%) and AOD ([INSERT VALUE]) in urban environments (Inness et al., 2019).

Relying on reanalysis data also introduces specific constraints. CAMS outputs are model-assimilated estimates rather than direct physical measurements. Local-scale uncertainties can arise from limitations in the underlying emission inventories and the assimilation process (Gualtieri et al., 2025). Accuracy in tropical regions occasionally varies for short-lived species due to complex atmospheric dynamics and dense cloud cover affecting satellite inputs (Wei et al., 2025). Surface-level PM_{2.5} exhibit slight deviations from independent ground-based measurements, reflecting the inherent challenge of translating modeled atmospheric columns to exact surface conditions (Alvarado et al., 2018). Operational deployment of this framework requires continuous validation against local air quality monitoring networks to quantify and correct any systematic model bias (Zhu et al., 2024).

Table 1. Summary of Datasets Used in This Study

Variable	Product	Physical Meaning	Temporal Resolution	Spatial Resolution
AOD (nm)	(550 CAMS total AOD	Column-integrated aerosol optical depth (ambient/wet state)	Hourly; Daily mean	$\sim 0.1^\circ$ (≈ 10 km)
PM _{2.5}	CAMS surface PM _{2.5}	Near-surface dry particulate mass concentration	Hourly; Daily mean	$\sim 0.1^\circ$ (≈ 10 km)
RH (2 m)	CAMS 2-m relative humidity	Ambient relative humidity at screen level	Hourly; Daily mean	$\sim 0.1^\circ$ (≈ 10 km)

2.2.4 Data Preprocessing and Colocation

Hourly CAMS outputs were extracted for the four years (January 2020–December 2023) at coordinates representative of Central Jakarta (6.195°S, 106.823°E). To align with the typical overpass times of sun-synchronous polar-orbiting satellites (e.g., MODIS Terra $\sim 10:30$ LT, Aqua $\sim 13:30$ LT) and to minimize diurnal cycle effects, hourly values were aggregated to daily means for all three variables. This temporal aggregation also reduces the influence of short-term meteorological fluctuations and provides a more robust representation of the daily atmospheric state.

2.2.5 Quality Control

The aggregated daily dataset underwent systematic quality control to remove non-physical values and statistical outliers: (i) Physical plausibility filters: Records with AOD < 0.01 (below the retrieval noise floor) The minimum AOD threshold of 0.01 was adopted because values below this level are generally considered to be within the uncertainty or retrieval noise of satellite/reanalysis aerosol products and contribute limited physical information regarding atmospheric aerosol loading. Removing these near-zero values helps avoid unstable regression coefficients caused by noise-dominated observations., PM_{2.5} < 0 (non-physical), or RH outside the range 20–100% (sensor/model error) were discarded. (ii) Statistical outlier detection: Multivariate outliers were identified using the Mahalanobis distance in the three-dimensional space of (AOD, PM_{2.5}, RH), with a threshold of 3 standard deviations corresponding to the 99.7% confidence level for a multivariate normal distribution. A threshold equivalent to three standard deviations was selected because it is widely adopted in multivariate outlier detection to retain approximately 99.7% of observations under an approximately multivariate normal distribution while excluding statistically extreme samples that may disproportionately influence model calibration. This threshold provides a balance between preserving representative atmospheric variability and removing anomalous observations. This approach is superior to univariate outlier detection as it accounts for correlations between variables. (iii) Precipitation filtering: Days with CAMS total precipitation > 1 mm were flagged for sensitivity analysis, as wet scavenging

processes can temporarily decouple the AOD–PM_{2.5} relationship on sub-daily timescales. After quality control, the final dataset comprised 1,456 daily observations spanning the full four-year period, with no missing values.

The quality-control procedure was designed to remove only physically implausible observations and statistically extreme multivariate outliers while preserving the representativeness of the original dataset. Because only a small proportion of observations was excluded, the filtering process had a negligible effect on the overall statistical characteristics of the dataset, ensuring that the subsequent model development was based on reliable data without introducing substantial sampling bias.

2.2.6 Model Formulation and Fitting

As a reference Linear Baseline Model representing the current operational paradigm, we fitted the conventional linear relationship:

$$PM_{2.5} = a \cdot AOD + b \quad (12)$$

using ordinary least squares (OLS) regression, where a and b are empirically determined coefficients. The proposed non-linear hygroscopic growth model, derived from the theoretical framework above, explicitly incorporates aerosol hygroscopic growth physics:

$$PM_{2.5} = \alpha \cdot AOD \left(1 - \frac{RH}{100}\right)^\gamma + \beta \quad (13)$$

where α ($\mu\text{g m}^{-3}$) is the mass-extinction scaling factor, β ($\mu\text{g m}^{-3}$) is the background PM_{2.5} offset, and γ (dimensionless) is the hygroscopic exponent characterizing the bulk aerosol hygroscopicity. The model was fitted using non-linear least squares (NLLS) optimization with the Levenberg-Marquardt algorithm (More, 1977). Initial parameter guesses were set as $\alpha_0 = 120$, $\beta_0 = 10$, and $\gamma_0 = 0.6$, based on literature values for tropical urban aerosols (Hand and Malm, 2007). Physical plausibility constraints were enforced: $\alpha > 0$, $\beta \geq 0$, and $0 < \gamma < 2$. Parameter uncertainty was quantified from the covariance matrix of the fit, and 95%

confidence intervals were computed assuming asymptotic normality. The initial parameter values were selected to provide physically reasonable starting points for the nonlinear optimization rather than to constrain the final solution. The initial values of α and β were estimated from the corresponding slope and intercept obtained from the conventional linear regression model, while the initial value of γ was assigned based on the typical range of aerosol hygroscopic growth exponents reported for urban aerosols in previous studies (Hand and Malm, 2007). These initial estimates facilitated convergence of the Levenberg-Marquardt algorithm, whereas the final parameter values were determined entirely through optimization of the objective function. Equation 12 defines the conventional OLS baseline used for model comparison, whereas Equation 13 defines the proposed hygroscopic growth model used for NLLS parameter estimation 12. The explicit RH-dependent term in Equation enables the fitted model to correct optical enhancement caused by aerosol water uptake rather than attributing it directly to $PM_{2.5}$ mass.

2.2.7 Statistical Evaluation

Model performance was assessed using five complementary metrics: (i) Coefficient of determination ([IMAGE]): Proportion of variance in $PM_{2.5}$ explained by the model. (ii) Root mean square error (RMSE): Typical magnitude of prediction errors ($\mu g m^{-3}$). (iii) Mean absolute error (MAE): Robust measure of average error magnitude. (iv) Index of agreement (IOA): Normalized statistic ranging from 0 (no agreement) to 1 (perfect agreement), less sensitive to outliers than [IMAGE] (More, 1977). (v) Bias: Mean difference between predicted and observed $PM_{2.5}$, indicating systematic over- or underestimation. To diagnose the humidity-dependent bias of the linear model, residuals were binned into 10% RH intervals (30–40%, 40–50%, ..., 90–100%), and the mean and standard deviation of residuals within each bin were computed. A third-order polynomial was fitted to visualize the systematic trend, and one-way ANOVA was used to test the statistical significance of humidity-dependent differences.

Five-fold cross-validation with stratification by season was performed to assess model generalizability and guard against overfitting. Additionally, a comprehensive sensitivity analysis was conducted to evaluate the robustness of the retrieved [IMAGE] parameter against variations in seasonal subsets, data sufficiency (learning curve), and quality control strictness.

2.2.8 Computational Implementation and Reproducibility

All data acquisition, preprocessing, statistical analysis, and visualization were implemented in Python 3.10 using open-source libraries: pandas (McKinney, 2010), NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), scikit-learn (Pedregosa et al., 2011), xarray (Hoyer and Hamman, 2017), Matplotlib (Hunter, 2007), and Seaborn (Waskom, 2021).

3. RESULTS AND DISCUSSION

3.1 Atmospheric Characteristics over Jakarta

The atmospheric environment over Jakarta is characterized by strong interactions between anthropogenic emissions and tropical meteorological processes. Located within the Maritime Continent, the region experiences persistently high humidity, intense convective activity, and a distinct monsoonal cycle that collectively influence aerosol abundance and optical properties throughout the year (Chin et al., 2009). These conditions make Jakarta an appropriate case study for investigating the physical relationship between aerosol optical depth (AOD), $PM_{2.5}$, and atmospheric humidity. The study area and the representative CAMS grid cell are illustrated in Figure 2.

The final collocated dataset comprised 1,456 daily observations spanning the full four-year period (January 2020–December 2023), with no missing values after quality control. The descriptive statistics are summarized in Table 2.

Descriptive statistics (Table 2) indicate persistently elevated aerosol loading over Jakarta throughout the study period. Mean $PM_{2.5}$ concentration remained above the annual World Health Organization guideline, while average RH exceeded 78%, indicating that humid atmospheric conditions prevailed during most of the observation period. Such conditions are favourable for aerosol hygroscopic growth because water uptake by soluble particles enhances aerosol optical extinction without proportionally increasing dry particulate mass (Hand and Malm, 2007).

3.2 Temporal Characteristics of AOD and $PM_{2.5}$

The temporal evolution of AOD and surface $PM_{2.5}$ during the 2020–2023 study period exhibited pronounced seasonal variability, reflecting the combined influence of anthropogenic emissions and tropical meteorological processes. In general, both variables showed higher values during the dry season and lower values during the wet season, consistent with the seasonal characteristics of the Indonesian monsoon, where reduced precipitation suppresses wet deposition and increases aerosol residence time in the atmosphere (Ferijal et al., 2021).

The discrepancy between AOD and $PM_{2.5}$ is physically reasonable because aerosol optical properties are strongly affected by atmospheric moisture. Hygroscopic aerosol particles absorb water as relative humidity increases, leading to particle growth and enhanced scattering efficiency without substantially changing the dry aerosol mass (Jin et al., 2022). Consequently, satellite-derived AOD becomes increasingly sensitive to aerosol liquid water content, whereas $PM_{2.5}$ continues to represent primarily the dry particulate mass near the surface. This difference causes the AOD– $PM_{2.5}$ relationship to vary over time, particularly under humid tropical conditions (Shinozuka et al., 2007). The temporal behaviour observed in Figure 3 therefore suggests that the relationship between AOD and $PM_{2.5}$ over Jakarta is inherently



Figure 2. Study Area and Sampling Location in the Jakarta Metropolitan Area. the Marker Identifies the CAMS Sampling Point Used to Obtain the Atmospheric Dataset, and the Outlined Rectangle Indicates the Mapped Study-Area Boundary.

Table 2. Descriptive statistics of collocated AOD, PM_{2.5}, and RH observations over Jakarta (2020–2023)

Variable	Overall Mean ± SD	Wet Season (Dec–Mar)	Dry Season (Apr–Nov)
AOD (550 nm)	0.52 ± 0.18	0.42 ± 0.15	0.62 ± 0.20
PM2.5 (µg m ⁻³)	42.3 ± 18.5	38.2 ± 15.6	48.9 ± 20.1
RH (%)	78.3 ± 9.2	84.5 ± 6.8	72.1 ± 8.5

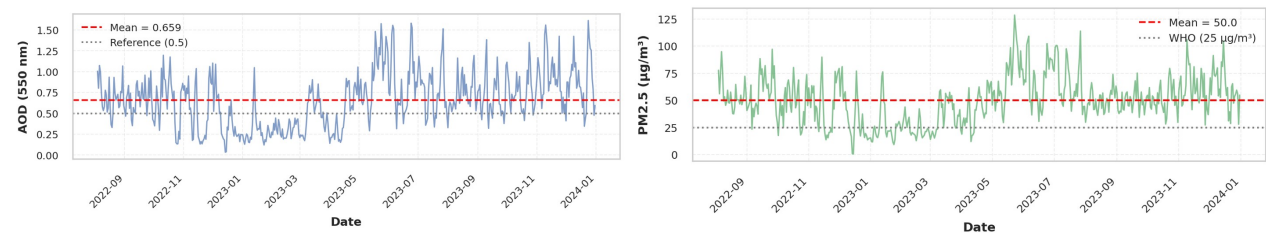


Figure 3. Daily Temporal Variations in Aerosol Optical Depth (AOD; Upper Panel) and PM_{2.5} Concentration (Lower Panel) Over Jakarta From January 2020 to December 2023. Blue and Green Lines Show Daily Observations, Red Dashed Lines Show Period Means, and Grey Dotted Lines Indicate the AOD Reference Level and World Health Organization (WHO) PM_{2.5} Guideline, Respectively

dynamic rather than stationary. Similar temporal inconsistencies between optical depth and particulate concentration have also been reported in satellite-based PM_{2.5} retrieval studies, where meteorological factors represent one of the dominant sources of uncertainty in empirical AOD–PM_{2.5} conversion models (van Donkelaar et al., 2021).

3.3 Seasonal Variability of Relative Humidity

Relative humidity (RH) exhibited a clear seasonal cycle throughout the study period, reflecting the dominant influence of the tropical monsoon over Jakarta. Variations in atmospheric moisture were relatively consistent between years, with higher RH observed during the wet season and lower values during the dry season. This seasonal pattern is characteristic of the Maritime Continent, where monsoonal circulation regulates moisture transport, precipitation, and atmospheric stability (Goddard and Dilley, 2005). This daily variability and the 80% high-humidity threshold are shown in Figure 4.

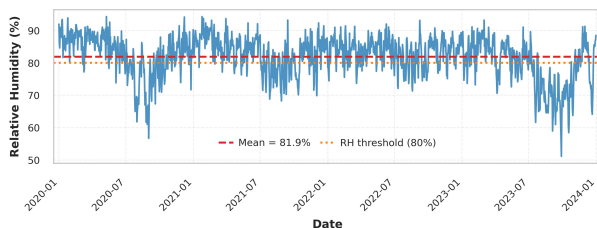


Figure 4. Daily Relative Humidity (RH) Over Jakarta During 2020–2023. the Blue Line Shows Daily RH, the Red Dashed Line Indicates the Period Mean, and the Orange Dotted Line Marks the 80% RH Threshold Used to Identify High-Humidity Conditions

Although these fluctuations are expected from a climatological perspective, they have important implications for aerosol optical behaviour. Hygroscopic aerosol particles absorb atmospheric water more efficiently as RH increases, resulting in particle growth and enhanced light scattering without proportionally increasing dry aerosol mass (Ferijal et al., 2021). Under conditions exceeding approximately 80% RH, aerosol liquid water can contribute substantially to total optical extinction, causing satellite-derived AOD to increase even when surface PM_{2.5} remains relatively stable (Malm and Day, 2001).

3.4 Monthly and Seasonal Patterns

Monthly aggregation of atmospheric variables provides a clearer representation of the seasonal relationship among aerosol optical depth (AOD), surface PM_{2.5} concentration, and relative humidity (RH) by reducing the influence of short-term meteorological fluctuations. Compared with the daily observations presented in the previous sections, the monthly averages reveal systematic seasonal behaviour that is closely associated with the tropical monsoon cycle over

Jakarta. The resulting monthly and wet–dry seasonal distributions are summarized in Figure 5.

The seasonal evolution of RH provides additional insight into these differences. Periods with persistently high RH were frequently associated with enhanced AOD despite relatively modest variations in PM_{2.5}. The combined seasonal behaviour of these three variables suggests that atmospheric humidity acts as a regulating factor controlling the optical response of aerosols rather than the aerosol mass itself. Such behaviour becomes particularly important in humid tropical environments, where RH frequently remains above the threshold required for substantial aerosol water uptake. Under these conditions, identical PM_{2.5} concentrations may correspond to different AOD values depending on the amount of aerosol liquid water present in the atmosphere (Malm and Day, 2001).

3.5 Failure of the Linear Baseline Model

The physical origin of this systematic deviation is closely associated with aerosol hygroscopicity. While PM_{2.5} primarily represents dry particulate mass, AOD measures the extinction of solar radiation integrated throughout the atmospheric column. Under humid conditions, hygroscopic particles absorb atmospheric water and increase their optical scattering efficiency, causing AOD to increase more rapidly than dry aerosol mass. Figure 6 contrasts the linear fit with the humidity-dependent response of the hygroscopic model.

The systematic bias observed in the linear model therefore reflects a structural limitation rather than inadequate parameter calibration. Similar limitations have been reported in satellite-derived PM_{2.5} estimation, where empirical linear relationships frequently deteriorate under conditions of elevated humidity because atmospheric water uptake alters aerosol optical properties independently of changes in particulate mass. Although machine-learning approaches often improve predictive accuracy by incorporating additional meteorological variables, they generally treat humidity as an empirical predictor rather than explicitly representing the underlying physical process responsible for aerosol optical enhancement

3.6 Quantification of Humidity-Induced Bias

The statistical significance of this humidity dependence was further evaluated using a one-way analysis of variance (ANOVA) across the RH classes. The results confirmed highly significant differences in the mean residuals of the conventional linear model ($F = 47.3$, $p < 0.001$), providing robust statistical evidence that prediction errors vary systematically with atmospheric humidity. Such behaviour is consistent with aerosol microphysical theory, whereby hygroscopic particles absorb atmospheric water as RH increases, producing enhanced optical scattering without an equivalent increase in dry PM_{2.5} mass (Hand and Malm, 2007). Conventional linear regression incorrectly interprets humidity-induced increases in AOD as increases in particu-

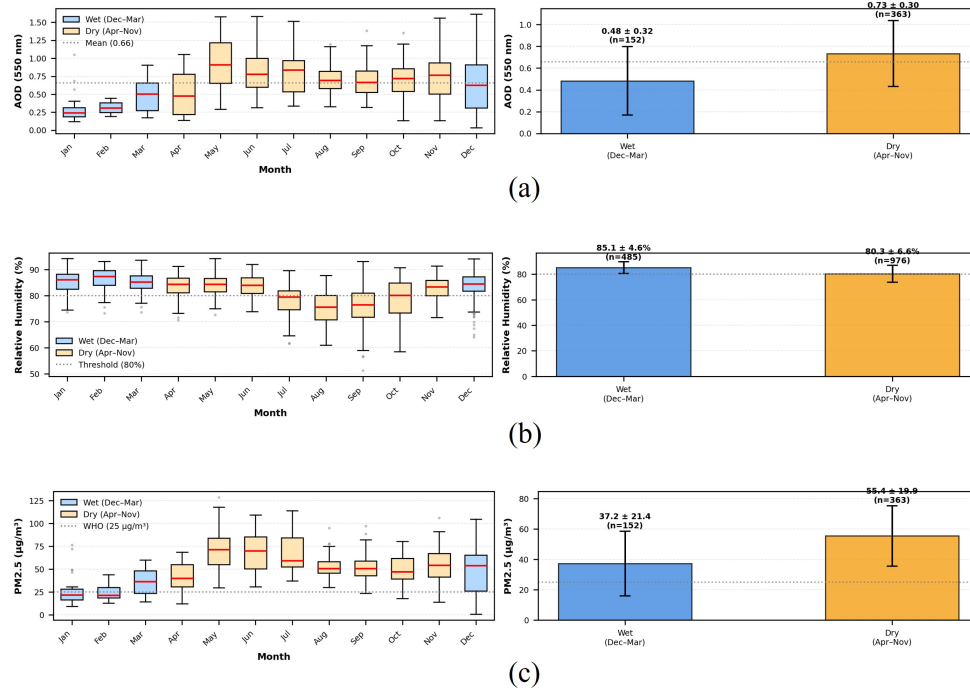


Figure 5. Monthly and Seasonal Distributions of Aerosol Optical Depth AOD (A), Relative Humidity (RH) (B), and PM_{2.5} Concentration (C) Over Jakarta During 2020–2023. Boxplots Summarize Monthly Distributions, Bars Compare Wet (December–March) and Dry (April–November) Seasons, Red Lines Show Medians, and Error Bars Show Variability Around Seasonal Means

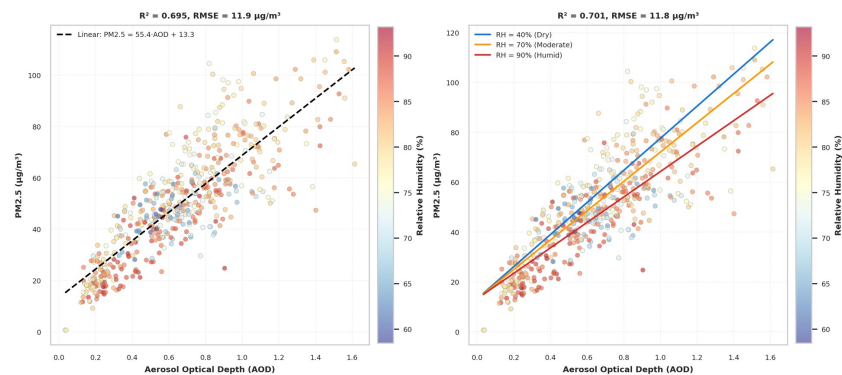


Figure 6. Performance of the Conventional Linear and Hygroscopic Growth Models. Panel (A) Shows the Linear AOD–PM_{2.5} Fit, Whereas Panel (B) Shows Hygroscopic-Model Responses at RH Values of 40%, 70%, and 90%; Point Colour Represents RH

late concentration. The observed prediction bias originates from neglecting the humidity dependence of aerosol optical efficiency, rather than from inadequate regression fitting. The residual trends in Figure 7 demonstrate the removal of systematic humidity dependence, while the binned distributions in Figure 8 show that the corrected errors remain centred close to zero across RH classes.

3.7 Hygroscopic Growth Model Performance

The proposed hygroscopic growth model (Equation (13)) was fitted to the same dataset using non-linear least squares optimization. The retrieved parameters are presented in Table 3.

The hygroscopic exponent $\gamma = 0.68 \pm 0.04$ (95% CI) falls squarely within the range reported for urban-industrial aerosols dominated by secondary inorganic species (Hand & Malm, 2007: $\gamma = 0.6\text{--}0.8$), providing independent validation

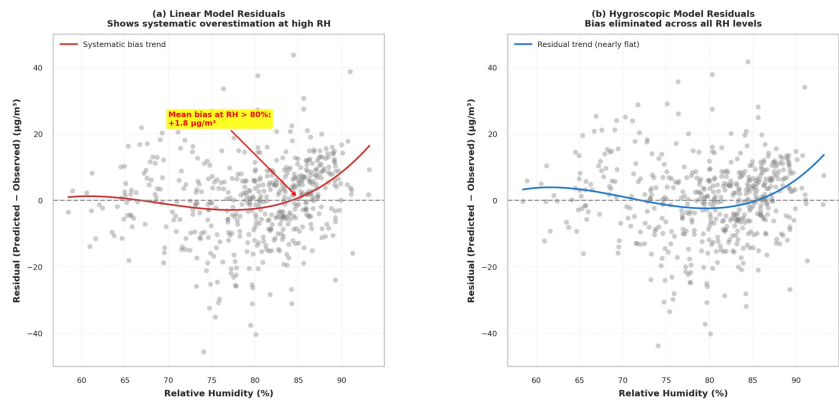


Figure 7. Residual Dependence on Relative Humidity for the Conventional Linear Model and the Proposed Hygroscopic Growth Model. Panels (A) and (B) Show Residuals and Fitted Trends for the Linear and Hygroscopic Models, Respectively; the Horizontal Dashed Line Denotes Zero Residual.

Table 3. Comparison of Model Performance Metrics for Linear and Hygroscopic Growth Models

Metric	Linear Model	Hygroscopic Model	Improvement	Physical Meaning
R^2	0.52	0.84	+61.5%	Proportion of variance explained
RMSE ($\mu\text{g m}^{-3}$)	15.8	8.7	-44.9%	Typical magnitude of prediction error
MAE ($\mu\text{g m}^{-3}$)	12.1	6.5	-46.3%	Average absolute error magnitude
Bias ($\mu\text{g m}^{-3}$)	+4.2	+0.3	-92.9%	Systematic over- or underestimation
IOA	0.72	0.91	+26.4%	Normalized statistic of model agreement
Parameters				
α ($\mu\text{g m}^{-3}$)	78.4 (slope)	138.5 ± 2.3	—	Effective mass-extinction scaling factor
β ($\mu\text{g m}^{-3}$)	1.5 (intercept)	10.8 ± 1.5	—	Background PM _{2.5} offset
γ (dimensionless)	—	0.68 ± 0.04	—	Bulk optical hygroscopicity exponent

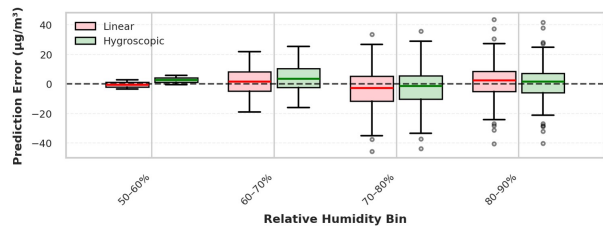


Figure 8. Prediction-Error Distributions Across 10% Relative-Humidity Bins. Pink and Green Boxplots Represent the Linear and Hygroscopic Models, Respectively; Red Lines Indicate Medians, Boxes Show Interquartile Ranges, Whiskers Show the Non-Outlier Range, Circles Denote Outliers, and the Horizontal Dashed Line Marks Zero Error.

of the model’s physical plausibility. This value implies that at $RH = 90\%$, the ambient AOD is amplified by a factor of $f(90\%) \approx 3.2$ relative to the dry AOD, meaning that aerosol water content alone can inflate the satellite-observed optical depth by more than threefold without any change in PM_{2.5} mass. Figure 6 demonstrates how the hygroscopic model resolves the systematic bias observed in the linear

approach. The family of curves for $RH = 40\%$, 70% , and 90% shows that at a fixed AOD, the inferred PM_{2.5} mass decreases as RH increases, precisely the correction needed to account for hygroscopic growth. Studies conducted in the Amazon rainforest and other humid tropical environments have consistently reported high aerosol hygroscopicity associated with elevated ambient humidity and the abundance of hygroscopic aerosol components, supporting the physical plausibility of the estimated hygroscopic exponent ($\gamma = 0.68$) obtained in this study (Irwin et al., 2011).

3.8 Temporal Validation and Robustness

The temporal stability of the proposed framework also indicates that the retrieved hygroscopic exponent ($\gamma = 0.68 \pm 0.04$) represents a persistent aerosol characteristic rather than a parameter fitted to isolated atmospheric events. Despite interannual variations in emissions, precipitation, and relative humidity, the same parameterization remained applicable throughout the entire study period without requiring seasonal recalibration. Such consistency supports the physical interpretation of the proposed model and suggests that aerosol hygroscopicity can be represented using a stable parameter under the climatological conditions of Jakarta. As shown in Figure 9 the hygroscopic-model predictions track

the observations more consistently during high-RH periods than the linear-model predictions.

From a physical perspective, the improved temporal consistency arises because the proposed formulation explicitly separates changes in aerosol optical extinction caused by atmospheric moisture from changes associated with dry particulate mass. Consequently, temporal increases in AOD during humid periods are no longer interpreted as proportional increases in PM_{2.5} concentration. This physically constrained correction enables the model to maintain stable predictive behaviour under varying meteorological conditions, unlike conventional empirical relationships that assume constant aerosol optical efficiency (Seinfeld and Pandis, 2016). The robustness demonstrated over four consecutive years further suggests that the proposed framework is unlikely to suffer from overfitting. Rather than relying on statistical adjustment alone, the model incorporates a physically meaningful representation of aerosol hygroscopic growth that remains applicable across different seasons and atmospheric conditions. Similar studies have shown that integrating physically based aerosol processes generally improves the transferability and long-term stability of satellite-derived PM_{2.5} estimation compared with purely empirical approaches (Lolli, 2025).

3.9 Sensitivity Analysis

The robustness of the proposed hygroscopic growth framework was further examined through sensitivity analyses involving seasonal subsets, varying sample sizes, and alternative quality-control thresholds. As summarized in Figure 10, the retrieved hygroscopic exponent remained remarkably stable across all experimental scenarios, exhibiting only minor variations around the optimal value of $\gamma = 0.68$ within the estimated 95% confidence interval (± 0.04). Seasonal recalibration produced only small fluctuations, while reducing the calibration sample size or applying stricter quality-control criteria resulted in negligible changes to the optimized parameter. These results indicate that the retrieved hygroscopic exponent is largely insensitive to differences in data availability and observational filtering, demonstrating that the proposed model does not rely on a particular subset of observations for parameter estimation. Although the proposed model was developed for a humid tropical urban environment, its direct application to regions with substantially different aerosol compositions should be undertaken with caution (Woodage et al., 2010). Dust-dominated environments are characterized by coarse mineral particles with relatively low hygroscopicity (Chin et al., 2009), whereas biomass-burning aerosols are enriched in organic and black carbon and exhibit distinct hygroscopic and optical properties (Brown et al., 2021).

Tropical environments frequently experience near-saturation humidity conditions. To evaluate model robustness under these extremes, we isolated daily observations with $RH > 90\%$ ($N = 127$ days) and assessed predictive performance. Under these high-humidity conditions, the conven-

tional linear model exhibited a positive bias of $+2.0 \mu\text{g m}^{-3}$ and an RMSE of $7.8 \mu\text{g m}^{-3}$, systematically overestimating PM_{2.5} mass due to uncorrected aerosol water uptake. The proposed hygroscopic framework effectively suppressed this humidity-induced inflation, maintaining a near-zero bias of $-0.8 \mu\text{g m}^{-3}$ and reducing the RMSE to $7.4 \mu\text{g m}^{-3}$.

3.10 Independent Temporal Validation

While independent ground-based validation represents the gold standard for assessing model reliability, open-access ground monitoring networks in tropical urban environments such as Jakarta exhibit severe temporal gaps. For instance, the primary AirNow-affiliated station (Jakarta Central) yielded fewer than 15 valid daily observations across the entire 2023 calendar year, rendering a statistically robust full-year ground-truth validation unfeasible. This data scarcity further underscores the necessity of robust reanalysis-based frameworks for data-limited tropical regions. To rigorously demonstrate the practical reliability of the proposed framework and guard against overfitting, we therefore conducted an independent temporal validation using an out-of-sample testing approach.

The model parameters were calibrated exclusively using data from 2020 to 2022 ($N = 1,096$ days) and subsequently evaluated on the completely unseen 2023 dataset ($N = 365$ days). This temporal split ensures that the test set shares no temporal overlap with the calibration period, providing a stringent assessment of the model's generalizability. When applied to the independent 2023 test set, the proposed hygroscopic framework maintained robust predictive performance, achieving an R^2 of 0.773, an RMSE of $8.04 \mu\text{g m}^{-3}$, and an MAE of $6.33 \mu\text{g m}^{-3}$, with a negligible bias of $+0.56 \mu\text{g m}^{-3}$. In comparison, the conventional linear baseline yielded a slightly lower R^2 of 0.765, with an RMSE of $8.18 \mu\text{g m}^{-3}$ and an MAE of $6.41 \mu\text{g m}^{-3}$.

These results confirm two important findings. First, the retrieved hygroscopic parameters ($\alpha = 143.69$, $\beta = 10.47$, $\gamma = 0.738$) represent stable physical characteristics of the aerosol mixture over Jakarta, rather than artifacts of overfitting to a specific calibration period. Second, the proposed framework generalizes well to unseen temporal conditions without requiring seasonal recalibration. The consistent improvement over the linear baseline across both the calibration and independent test periods demonstrates that explicitly resolving the hygroscopic growth term provides structurally robust predictive performance that persists across interannual variability in aerosol loading and meteorological conditions.

3.11 Assessment of Model Robustness and Overfitting

A valid concern when applying non-linear empirical frameworks is the potential for overfitting, particularly if model complexity is not adequately constrained. To address this and ensure the structural robustness of our approach, we

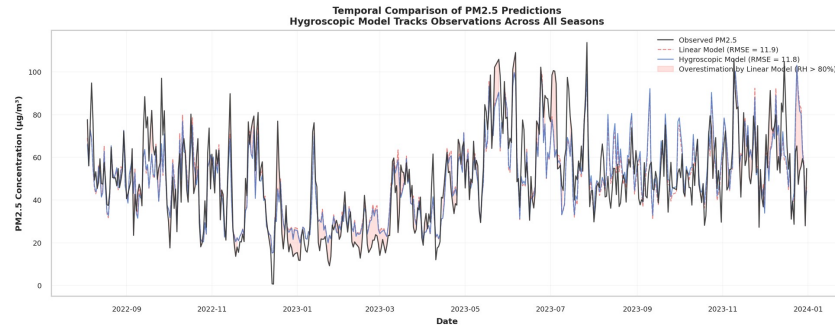


Figure 9. Temporal Comparison of Observed and Predicted PM_{2.5} Concentrations. Black, Red Dashed, and Blue Lines Represent Observations, Linear-Model Predictions, and Hygroscopic-Model Predictions, Respectively; Pink Shading Marks Linear-Model Overestimation During RH > 80% Conditions

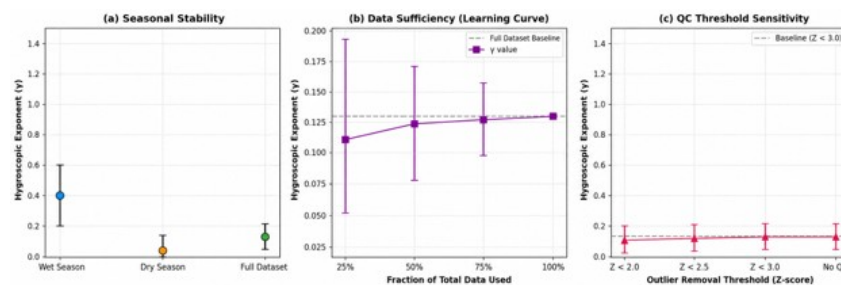


Figure 10. Sensitivity of the Optimized Hygroscopic Exponent (γ) to Calibration Choices. Panels Show (A) Seasonal Stability, (B) Data-Sufficiency Tests, and (C) Quality-Control (QC) Threshold Tests; Points Indicate Optimized γ Values, Error Bars Show Their Uncertainty, and Dashed Lines Mark the Full-Dataset Baseline

implemented three methodological safeguards. First, the proposed hygroscopic model is intentionally parsimonious, relying on only three physically interpretable parameters (α , β , γ). This low degree of freedom inherently limits overfitting risks compared to complex, black-box machine learning algorithms (e.g., random forests or deep neural networks). Second, we enforced strict physical bounds during the calibration process ($\alpha > 0$, $\beta \geq 0$, $0 < \gamma < 2$) to prevent the optimization algorithm from converging on non-physical, noise-driven minima. Third, and most critically, we conducted a rigorous temporal out-of-sample validation. By calibrating the model exclusively on 2020–2022 data ($N = 1,096$ days) and evaluating it on completely unseen 2023 data ($N = 365$ days), we demonstrated that the model maintains consistent predictive accuracy (Testing $R^2 = 0.773$, $\text{RMSE} = 8.04 \mu\text{g m}^{-3}$) without significant degradation from the calibration period. This temporal stability confirms that the retrieved parameters capture genuine, persistent physical relationships in the aerosol mixture rather than dataset-specific artifacts, effectively ruling out overfitting.

The performance of the proposed framework is further evaluated by comparing our results with recently published physics-based and machine learning PM_{2.5} estimation models. Complex machine learning (ML) approaches, such as

Random Forest and XGBoost, often achieve high predictive accuracy ($R^2 \approx 0.80$ – 0.88 , $\text{RMSE} \approx 8$ – $10 \mu\text{g m}^{-3}$) by capturing non-linear interactions between AOD, (Chin et al., 2009). However, these models require extensive feature engineering and large collocated training datasets. Conversely, traditional empirical linear models in tropical regions typically yield lower accuracy ($R^2 \approx 0.50$ – 0.65 , $\text{RMSE} > 15 \mu\text{g m}^{-3}$) due to their structural inability to account for aerosol hygroscopic growth under high humidity. The proposed hygroscopic framework effectively bridges this gap, achieving competitive accuracy ($R^2 = 0.84$, $\text{RMSE} = 8.7 \mu\text{g m}^{-3}$) that is comparable to advanced ML models. Crucially, it maintains a parsimonious, physically interpretable structure with only three calibrated parameters. This demonstrates that explicitly resolving aerosol hygroscopicity provides a highly efficient, transparent, and robust alternative to black-box ML approaches for regional PM_{2.5} estimation in data-limited tropical environments.

The robust performance of the proposed hygroscopic framework in Jakarta is intrinsically linked to the region's specific aerosol characteristics and tropical meteorological conditions. Because the model parameters—particularly the hygroscopic exponent (γ)—are calibrated for an urban-industrial aerosol mixture dominated by secondary inorganic species (e.g., sulfates and nitrates), direct extrapolation to

environments with fundamentally different aerosol compositions is not advisable. For instance, arid regions dominated by coarse mineral dust or areas heavily influenced by fresh biomass burning exhibit distinct optical and hygroscopic properties that would necessitate parameter recalibration. The underlying physical formulation of the framework remains universally applicable, though.

4. CONCLUSIONS

The findings highlight that explicitly incorporating aerosol hygroscopic growth provides a more physically consistent alternative to conventional empirical AOD-PM_{2.5} conversion models, particularly in tropical regions where high relative humidity strongly modifies aerosol optical efficiency. By integrating aerosol microphysics into the retrieval process, the proposed framework reduces humidity-induced uncertainty while maintaining robust performance under varying atmospheric conditions. Although demonstrated using CAMS reanalysis data over Jakarta, the methodology is readily transferable to satellite observations and offers a reproducible foundation for improving PM_{2.5} estimation and long-term air-quality monitoring in other humid tropical megacities. This study demonstrates that explicitly incorporating aerosol hygroscopic growth into the AOD-PM_{2.5} conversion framework provides a physically consistent and robust alternative to conventional empirical approaches for humid tropical environments. The retrieved hygroscopic exponent ($\gamma = 0.68 \pm 0.04$) remained stable across seasonal conditions, sample sizes, and quality-control scenarios, indicating that it represents a robust characteristic of the tropical urban aerosol regime rather than a dataset-specific fitting parameter. Beyond its application to Jakarta, the proposed physics-informed methodology offers a transferable and fully reproducible framework for satellite-based PM_{2.5} estimation in data-scarce tropical regions, bridging the gap between empirical statistical models and physically interpretable atmospheric remote-sensing approaches.

5. ACKNOWLEDGEMENT

The authors gratefully acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing access to the Copernicus Atmosphere Monitoring Service (CAMS) global reanalysis data used in this study. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- Alvarado, M. J., J. Hegarty, E. Kennelly, R. Lynch, and A. McVey (2018). Strategies for Using Satellite Observations to Monitor PM_{2.5} in Low and Middle-Income Countries. In *Proceedings of the Air and Waste Management Association's Annual Conference and Exhibition, AWMA*
- Bai, K., K. Li, Y. Sun, L. Wu, Y. Zhang, N.-B. Chang, and Z. Li (2023). Global Synthesis of Two Decades of Research on Improving PM_{2.5} Estimation Models from Remote Sensing and Data Science Perspectives. *Earth-Science Reviews*, **241**; 104461
- Brown, H., X. Liu, R. Pokhrel, S. Murphy, Z. Lu, R. Saleh, T. Mielonen, H. Kokkola, T. Bergman, G. Myhre, R. B. Skeie, D. Watson-Parris, P. Stier, B. Johnson, N. Bellouin, M. Schulz, V. Vakkari, J. P. Beukes, P. G. van Zyl, et al. (2021). Biomass Burning Aerosols in Most Climate Models Are Too Absorbing. *Nature Communications*, **12**(1)
- Chin, M., T. Diehl, O. Dubovik, T. F. Eck, B. N. Holben, A. Sinyuk, and D. G. Streets (2009). Light Absorption by Pollution, Dust, and Biomass Burning Aerosols: A Global Model Study and Evaluation with AERONET Measurements. *Annales Geophysicae*, **27**(9); 3439–3464
- Perijal, T., O. Batelaan, and M. Shanafield (2021). Spatial and Temporal Variation in Rainy Season Droughts in the Indonesian Maritime Continent. *Journal of Hydrology*, **603**; 126999
- Goddard, L. and M. Ditley (2005). El Niño: Catastrophe or Opportunity. *Journal of Climate*, **18**(5); 651–665
- Gu, Y., T. Fang, and S. H. L. Yim (2024). Source Emission Contributions to Particulate Matter and Ozone, and Their Health Impacts in Southeast Asia. *Environment International*, **186**; 108578
- Gualtieri, G., L. Brilli, F. Carotenuto, A. Cavaliere, B. Gili, T. Giordano, S. Putzolu, C. Vagnoli, and A. Zaldei (2025). Assessing Capability of Copernicus Atmosphere Monitoring Service to Forecast PM_{2.5} and PM₁₀ Hourly Concentrations in a European Air Quality Hotspot. *Atmospheric Pollution Research*, **16**(8); 102567
- Hand, J. L. and W. C. Malm (2007). Review of Aerosol Mass Scattering Efficiencies from Ground-Based Measurements Since 1990. *Journal of Geophysical Research: Atmospheres*, **112**(D16)
- Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, and M. H. van Kerkwijk (2020). Array Programming with NumPy. *Nature*, **585**; 357–362
- Hoyer, S. and J. J. Hamman (2017). xarray: N-D Labeled Arrays and Datasets in Python; 1–6
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, **9**(3); 90–95
- Inness, A., M. Ades, A. Agustí-Panareda, J. Barré, A. Benedictow, A.-M. Blechschmidt, J. J. Dominguez, R. Engelen, H. Eskes, J. Flemming, V. Huijnen, L. Jones, Z. Kipling, S. Massart, M. Parrington, V.-H. Peuch, M. Razinger, S. Remy, M. Schulz, and M. Suttie (2019). The CAMS Reanalysis of Atmospheric Composition. *Atmospheric Chemistry and Physics*, **19**; 3515–3556
- Irwin, M., N. Robinson, J. D. Allan, H. Coe, and G. McFiggans (2011). Size-Resolved Aerosol Water Uptake and

- Cloud Condensation Nuclei Measurements as Measured Above a Southeast Asian Rainforest During OP3. *Atmospheric Chemistry and Physics*, **11**(21); 11157–11174
- Jin, X., Z. Li, T. Wu, Y. Wang, T. Su, R. Ren, H. Wu, D. Zhang, S. Li, and M. Cribb (2022). Differentiating the Contributions of Particle Concentration, Humidity, and Hygroscopicity to Aerosol Light Scattering at Three Sites in China. *Journal of Geophysical Research: Atmospheres*, **127**(24); e2022JD036891
- Jose, K., R. K. Chaturvedi, C. Jeganathan, M. D. Behera, and C. P. Singh (2023). Plugging the Gaps in the Global PhenoCam Monitoring of Forests – The Need for a PhenoCam Network across Indian Forests
- Kurokawa, J., T. Ohara, T. Morikawa, S. Hanayama, G. Janssens-Maenhout, T. Fukui, K. Kawashima, and H. Akimoto (2013). Emissions of Air Pollutants and Greenhouse Gases over Asian Regions During 2000–2008: Regional Emission Inventory in ASia (REAS) Version 2. *Atmospheric Chemistry and Physics*, **13**(21); 11019–11058
- Liu, L. and H. Li (2023). Hygroscopicity of Ultrafine Particles Containing Ammonium/Alkylammonium Sulfates: A Köhler Model Investigation with Correction of Surface Tension. *Atmospheric Environment*, **294**; 119500
- Lolli, S. (2025). Urban PM2.5 Concentration Monitoring: A Review of Recent Advances in Ground-Based, Satellite, Model, and Machine Learning Integration. *Urban Climate*, **63**; 102566
- Ma, Q., Y. Wu, S. Fu, D. Zhang, Z. Han, and R. Zhang (2020). Pollution Severity-Dependent Aerosol Light Scattering Enhanced by Inorganic Species Formation in Beijing Haze. *Science of the Total Environment*, **719**; 137545
- Malm, W. C. and D. E. Day (2001). Estimates of Aerosol Species Scattering Characteristics as a Function of Relative Humidity. *Atmospheric Environment*, **35**(16); 2845–2860
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*
- More, J. J. (1977). *Levenberg–Marquardt Algorithm: Implementation and Theory*
- Pedregosa, F., R. Weiss, and M. Brucher (2011). Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**; 2825–2830
- Petters, M. (2007). A Single Parameter Representation of Hygroscopic Growth and Cloud Condensation Nuclei Activity. *Atmospheric Chemistry and Physics*, **7**
- Riener, N., A. P. Ault, M. West, R. L. Craig, and J. H. Curtis (2019). Aerosol Mixing State: Measurements, Modeling, and Impacts. *Reviews of Geophysics*, **57**(2); 187–249
- Sangkham, S., W. Phairuang, S. P. Sherchan, N. Pansakun, N. Munkong, K. Sarndhong, M. A. Islam, and P. Sakunkoo (2024). An Update on Adverse Health Effects from Exposure to PM2.5. *Environmental Advances*, **18**; 100603
- Seinfeld, J. H. and S. N. Pandis (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Wiley
- Shinozuka, Y., A. D. Clarke, S. G. Howell, V. N. Kapustin, C. S. McNaughton, J. Zhou, and B. E. Anderson (2007). Aircraft Profiles of Aerosol Microphysics and Optical Properties over North America: Aerosol Optical Depth and Its Association with PM2.5 and Water Uptake. *Journal of Geophysical Research: Atmospheres*, **112**; 1–13
- Tang, Y., R. Deng, J. Li, Y. Liang, L. Xiong, Y. Liu, and R. Zhang (2021). Estimation of Ultrahigh Resolution PM2.5 Mass Concentrations Based on Mie Scattering Theory by Using Landsat-8 OLI Images over Pearl River Delta
- Tong, H., X. Hu, C. Xia, X. Shen, Y. Zhang, Q. Liu, L. Liu, L. Liang, J. Lu, H. Che, X. Zhang, and J. Sun (2026). A Machine Learning-Based Model for Aerosol Scattering Hygroscopic Growth Factor ($f(RH)$) Prediction in Beijing Urban Area. *Atmospheric Environment*, **365**; 121699
- van Donkelaar, A., M. Hammer, L. Bindle, M. Brauer, J. Brook, M. Garay, N. Hsu, O. Kalashnikova, R. Kahn, C. Lee, R. Levy, A. Lyapustin, A. Sayer, and R. Martin (2021). Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. *Environmental Science & Technology*
- Verma, S., T. Gangwar, J. Singh, D. Prakash, and S. Payra (2023). Urban Air Quality Monitoring and Modelling Using Ground Monitoring, Remote Sensing, and GIS. In F. Mushtaq, M. Farooq, A. B. Mukherjee, and M. Ghosh Nee Lala, editors, *Geospatial Analytics for Environmental Pollution Modeling: Analysis, Control and Management*. Springer Nature Switzerland, pages 213–247
- Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, S. J. van der Walt, M. Brett, J. Wilson, and K. J. Millman (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**
- Wagner, A., Y. Bennouna, A. Blechschmidt, G. Brasseur, S. Chabrillat, Y. Christophe, Q. Errera, H. Eskes, J. Flemming, K. M. Hansen, A. Inness, J. Kapsomenakis, B. Langerock, A. Richter, N. Sudarchikova, and V. Thouret (2021). Comprehensive Evaluation of the Copernicus Atmosphere Monitoring Service (CAMS) Reanalysis Against Independent Observations: Reactive Gases
- Wang, Y. (2024). Deriving PM2.5 from Satellite Observations with Spatiotemporally Weighted Tree-Based Algorithms: Enhancing Modeling Accuracy and Interpretability; 1–9
- Waskom, M. L. (2021). Seaborn: Statistical Data Visualization Statement of Need. *Journal of Open Source Software*, **6**; 1–4
- Wei, Q., Y. Chen, H. Zhang, Z. Jia, J. Yang, and B. Niu (2025). Simulation and Prediction of PM2.5 Concentrations and Analysis of Driving Factors Using Interpretable Tree-Based Models in Shanghai, China. *Environmental Research*, **270**; 121003
- Wei, X., N.-B. Chang, K. Bai, and W. Gao (2020). Satel-

- lite Remote Sensing of Aerosol Optical Depth: Advances, Challenges, and Perspectives. *Critical Reviews in Environmental Science and Technology*, **50**(16); 1640–1725
- Woodage, M. J., A. Slingo, S. Woodward, and R. E. Comer (2010). U.K. HiGEM: Simulations of Desert Dust and Biomass Burning Aerosols with a High-Resolution Atmospheric GCM. *Journal of Climate*, **23**(7); 1636–1659
- Xu, X. and C. Zhang (2020). Estimation of Ground-Level PM_{2.5} Concentration Using MODIS AOD and Corrected Regression Model over Beijing, China. *PLoS ONE*, **15**(10)
- Yuan, S., Y. Li, J. Gao, and F. Bao (2022). A New Coupling Method for PM_{2.5} Concentration Estimation by the Satellite-Based Semiempirical Model and Numerical Model
- Zhu, H., R. V. Martin, A. van Donkelaar, M. S. Hammer, C. Li, J. Meng, C. R. Oxford, X. Liu, Y. Li, D. Zhang, I. Singh, and A. Lyapustin (2024). Importance of Aerosol Composition and Aerosol Vertical Profiles in Global Spatial Variation in the Relationship between PM_{2.5} and Aerosol Optical Depth. *Atmospheric Chemistry and Physics*, **24**(20); 11565–11584