High-Resolution PM2.5 Modeling and Respiratory Disease Correlation in Chiang Mai Using Himawari-9 and Ground Observations

Katanchalee Taweepornwattanakul[1,2*] and Wataru Takeuchi[2]

[1] Department of Civil Engineering, The University of Tokyo, Tokyo, Japan

[2] Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

*katanchalee-katan@g.ecc.u-tokyo.ac.jp

Abstract *Haze pollution from fine particulate matter ($PM_{2.5}$) is a major environmental and health concern in Southeast Asia, with northern Thailand heavily impacted by seasonal biomass burning. Chiang Mai experiences recurrent haze episodes each February–April, with concentrations far above health guidelines. Ground monitoring offers direct $PM_{2.5}$ observations but remains spatially limited. Geostationary satellites such as Himawari-9 provide high-frequency coverage, enabling near-real-time monitoring when calibrated properly. This study, focused on 2023, developed a framework linking Himawari-9 top-of-atmosphere reflectance, AOD retrievals, meteorological factors, and respiratory health outcomes. AOD was retrieved using a 6S-based look-up table and integrated with ERA5 reanalysis in Random Forest models. Seasonal evaluation showed strong predictive skill (R^2 = 0.82 during burning, 0.76 in non-haze months), while year-round validation highlighted strong non-stationarity. Monthly exposure estimates correlated with hospital admissions, with asthma and COPD showing significant lagged responses. The framework demonstrates the potential of geostationary satellites for health-relevant air quality monitoring while identifying the need for multi-year and multi-sensor studies.*

Keywords: Himawari9, AOD, $PM_{2.5}$ Modeling, Seasonal Haze Episodes, Respiratory health

1. Introduction

Air pollution from fine particulate matter ($PM_{2.5}$) is a critical environmental and health issue worldwide, with Southeast Asia among the most vulnerable regions. Northern Thailand, particularly Chiang Mai, experiences recurrent haze episodes during the dry season (February–April), primarily driven by local biomass burning and transboundary smoke transport. These events routinely exceed the World Health Organization (WHO) guideline of 5 µg/m³, disrupting daily life, affecting tourism, and straining healthcare systems.

Ground-based air quality monitoring in Thailand remains spatially limited, restricting comprehensive regional assessment. Satellite remote sensing offers a powerful alternative, enabling large-scale, high-frequency aerosol observations. Instruments such as MODIS and VIIRS provide valuable data but with limited temporal coverage, while geostationary sensors like Himawari-9's Advanced Himawari Imager (AHI) deliver 10-minute observations, facilitating near-real-time monitoring of haze dynamics. However, translating aerosol optical depth (AOD) into reliable surface-level $PM_{2.5}$ estimates is challenging due to boundary-layer dynamics, meteorology, and emission variability.

Previous studies across Asia have emphasized the necessity of integrating meteorological parameters into AOD–$PM_{2.5}$ models to account for aerosol vertical distribution and hygroscopic growth (Tan et al., 2022). In Thailand, sensitivity to aerosol chemical schemes has been demonstrated (Bran et al., 2022), while source apportionment studies highlight contributions from biomass burning, traffic, and transboundary transport (Chansuebsri et al., 2024). Health-focused research also links haze exposure to asthma, COPD, pneumonia, and premature mortality. These findings underscore the need for spatially explicit, epidemiologically relevant modeling frameworks.

Building on this foundation, the present study develops a Himawari-9–based pipeline that integrates Top-of-Atmosphere (TOA) reflectance, radiative transfer modeling (6S LUT), and machine learning to estimate $PM_{2.5}$ in Chiang Mai. By applying strict quality control and season-specific calibration, the framework addresses the strong non-stationarity of AOD–$PM_{2.5}$ relationships. Beyond exposure modeling, it links monthly $PM_{2.5}$ estimates to hospital admissions for major respiratory diseases, using time-lag regression to assess delayed health impacts. This integrated approach contributes a replicable framework for high-frequency, health-relevant air quality monitoring in Southeast Asia.

2. Methodology

2.1 Study Area and Data Sources

This study focuses on Chiang Mai Province in northern Thailand, a mountainous valley region frequently affected by severe haze episodes during February–April. Top-of-atmosphere (TOA) reflectances were obtained from the Advanced Himawari Imager (AHI) onboard Himawari-9 at 10-minute temporal resolution. Three channels were selected for analysis: the blue band (0.47 μm), the red band (0.64 μm), and the shortwave infrared (SWIR) band (2.3 μm). Ground-based $PM_{2.5}$ concentrations were collected from the Pollution Control Department station at Chang Phueak. Meteorological variables, including air temperature, relative humidity, surface pressure, wind speed, and precipitation, were retrieved from ERA5 reanalysis. Monthly hospital admission records for asthma, chronic obstructive pulmonary disease (COPD), pneumonia, and influenza were provided by the Chiang Mai Provincial Health Office.

2.2 Data Preprocessing and AOD Retrieval

All datasets were harmonized to Coordinated Universal Time (UTC) and resampled to hourly resolution for consistency. Quality control (QC) included limiting observations to daytime scenes with solar zenith angle (SZA) $<70°$, discarding reflectances outside the physical range (0–1), and excluding rows with missing meteorological data. Retrieved AOD values were constrained to

0.05–2.5, yielding 1,179 valid records from 3,795 initial observations. Seasonal stratification defined February–April as the burning season and the remaining months as non-haze conditions. Aerosol optical depth (AOD) at 550 nm was retrieved using a look-up table (LUT) generated from the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) radiative transfer model through the Py6S interface. The inversion relied on the blue (0.47 μm) and red (0.64 μm) channels, with surface reflectance parameterized using the SWIR band (2.3 μm) under a dark-target approach. Biomass burning aerosol profiles were assigned during February–April, while continental profiles were applied for non-haze months. A fixed view zenith angle (48.3°) and relative azimuth angle (0°) were assumed, recognizing potential biases from these simplifications.

2.3 $PM_{2.5}$ Modeling Framework

The retrieved AOD was merged with ERA5 meteorology and ground-based $PM_{2.5}$ observations to construct predictive models. Random Forest regression was employed for its robustness against nonlinearity and multicollinearity. Predictor variables included AOD and five meteorological parameters, with $PM_{2.5}$ concentrations as the dependent variable. The model was configured with 600 trees, a minimum leaf size of three, and a random seed of 42. Two evaluation strategies were used: (i) seasonal hold-in validation to assess performance during haze and non-haze periods separately, and (ii) rolling time-series cross-validation to examine generalizability across the full year. Performance was measured using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). Feature importance was analyzed to quantify the relative contribution of predictors.

2.4 Health Impact Assessment

Monthly averages of modeled $PM_{2.5}$ concentrations were linked to hospital admission records for asthma, COPD, pneumonia, and influenza. Admissions were normalized per 100,000 population to ensure temporal comparability. Exposure–response relationships were evaluated using correlation analysis and time-lag regression with one- and two-month delays to capture delayed health effects. The analysis emphasized chronic respiratory diseases (asthma, COPD), which are more strongly influenced by cumulative pollution exposure, while acknowledging the limitations of monthly health data in capturing acute outcomes.

3. Results and Discussion

3.1 TOA Reflectance and Quality Control

The Himawari-9 TOA reflectance dataset initially contained 3,795 hourly records across 2023. After applying strict quality control—including restriction to daytime observations (SZA < 70°), reflectance validity checks, and filtering out incomplete meteorological data—1,179 high-quality records remained. Seasonal stratification yielded 237 records for February–April (burning season) and 942 for the rest of the year. Figure 1 shows TOA reflectance variability after filtering, where anomalous spikes from cloud contamination were effectively removed while retaining seasonal patterns.

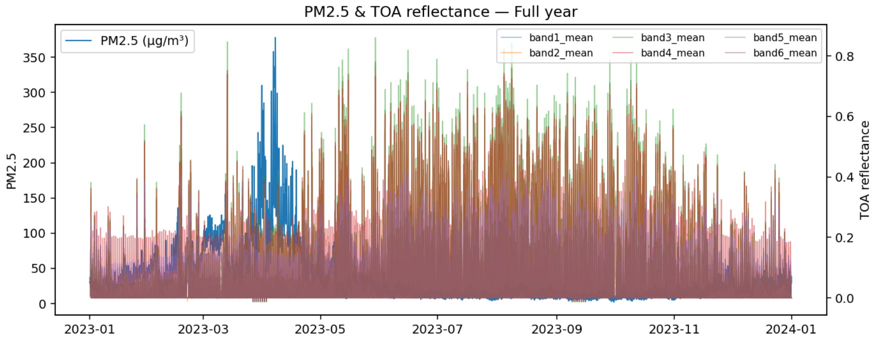

Figure 1. TOA reflectance after quality control, showing effective removal of anomalous cloud-contaminated observations while preserving seasonal variability.

3.2 AOD Retrieval and Seasonal Variability

AOD at 550 nm was retrieved via 6S LUT inversion using blue (0.47 µm) and red (0.64 µm) channels, with SWIR (2.3 µm) reflectance parameterized by a dark-target scheme. Seasonal aerosol models were applied: biomass burning profiles for February–April and continental profiles for other months. The retrieved AOD distribution was physically consistent (median = 0.41, range 0.05–2.5). Seasonal contrasts were clear, with mean AOD > 0.80 in haze months and < 0.30 in non-haze months. Figure 2 illustrates monthly AOD variability, showing peak values in March 2023 consistent with intensified open burning.

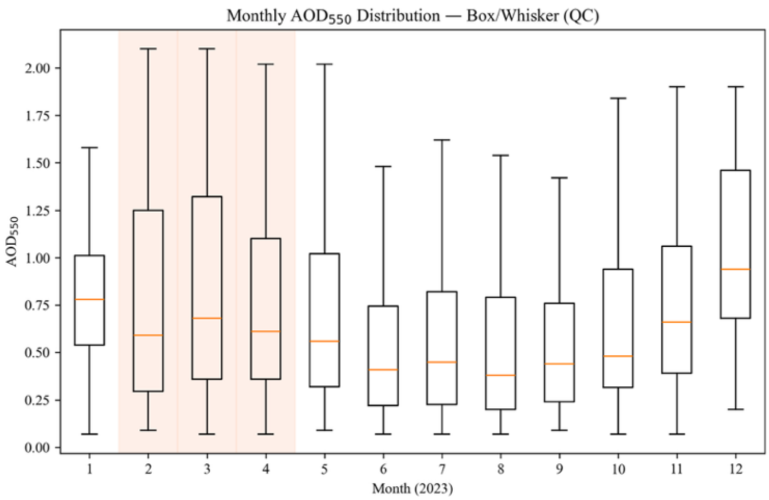

Figure 2. Temporal distribution of retrieved AOD at 550 nm, showing sharp seasonal peaks during biomass burning episodes.

3.3 $PM_{2.5}$ Modeling with Random Forest

Random Forest regression models combining AOD and meteorological covariates achieved strong performance when stratified by season. During the burning season, R^2 reached 0.82, though with larger errors (MAE ≈ 21.6 µg/m³, RMSE ≈ 31.0 µg/m³), reflecting the extreme variability of haze pollution. Non-haze months achieved $R^2 = 0.76$ with substantially lower errors (MAE ≈ 2.9 µg/m³, RMSE ≈ 4.3 µg/m³), consistent with the narrower $PM_{2.5}$ range. Feature importance analysis showed relative humidity and temperature as dominant predictors, while AOD was particularly important during haze episodes. In the other hands, rolling time-series cross-validation revealed poor year-round generalization, with R^2 near zero or negative. This indicates strong non-stationarity: predictor–response relationships valid during haze periods break down in non-haze conditions. Such non-stationarity emphasizes the necessity of season-specific calibration and cautions against using a single annual model without adaptive or hybrid strategies.

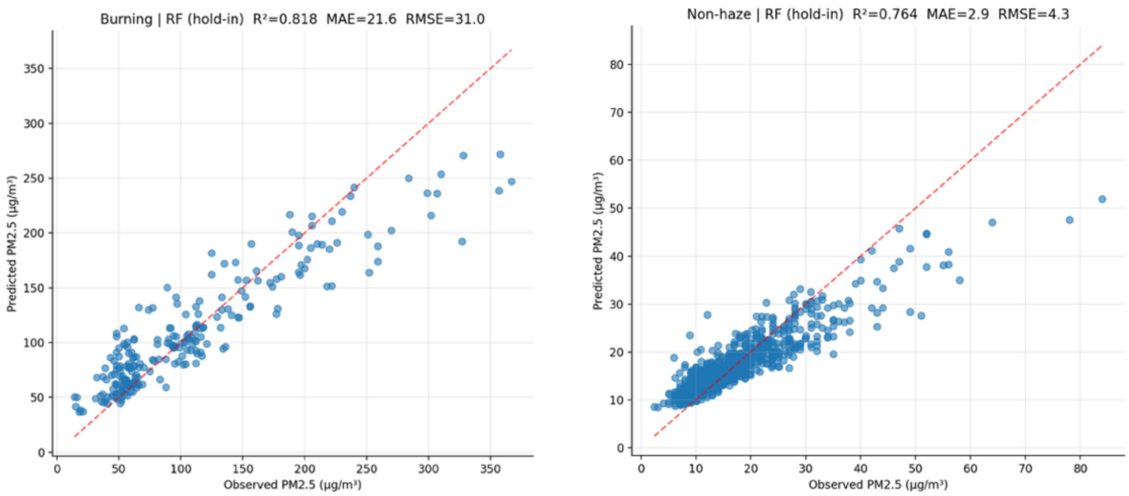

Figure 3. Observed versus predicted $PM_{2.5}$ for haze and non-haze Random Forest models, highlighting stronger AOD influence during burning season conditions.

3.4 Monthly Exposure and Health Associations

Monthly AOD and $PM_{2.5}$ averages peaked during March 2023 (AOD > 1.0, $PM_{2.5}$ > 70 µg/m³), far exceeding WHO guidelines. Outside haze months, averages dropped below 25 µg/m³. Linking these to hospital admissions revealed strong associations for asthma and COPD, particularly with a two-month lag (adjusted $R^2 \approx 0.97$). Pneumonia and influenza showed weaker associations, reflecting additional drivers beyond air pollution.

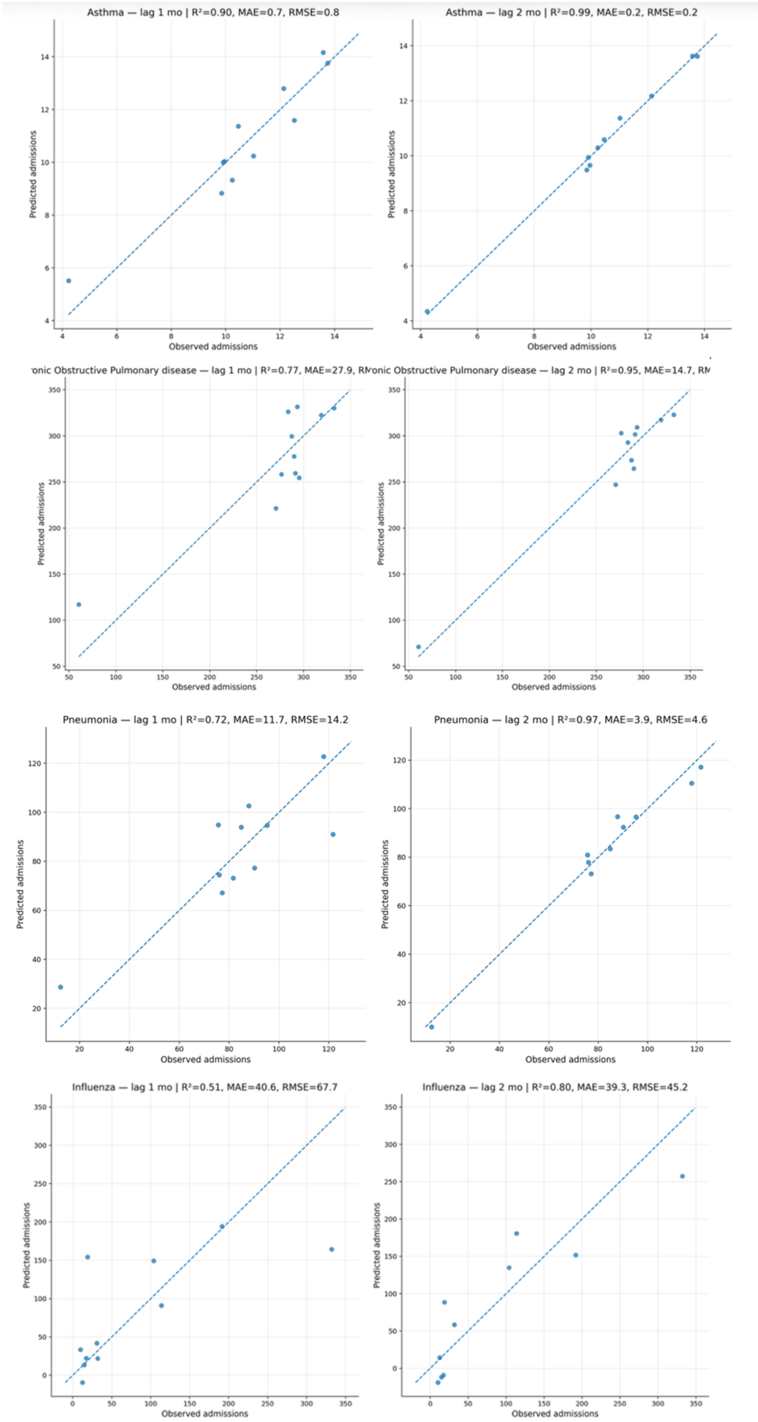

Figure 4. Monthly AOD, $PM_{2.5}$, and hospital admissions for respiratory diseases, showing strong lagged responses for asthma and COPD.

These findings collectively demonstrate that geostationary Himawari-9 observations can be effectively transformed into epidemiologically relevant surface $PM_{2.5}$ estimates when combined with strict quality control, radiative transfer modeling, and meteorological covariates. The season-specific Random Forest models provided reliable predictions during both haze and non-haze regimes, while the sharp performance decline in year-round cross-validation underscores the non-stationary nature of the $PM_{2.5}$–AOD relationship. Importantly, the preliminary health analysis suggested strong lagged associations between monthly exposure and respiratory hospital admissions, particularly for asthma and COPD. However, the exceptionally high adjusted R^2 values (≈0.97) are likely inflated due to the small sample size and coarse monthly aggregation, and thus should be interpreted with caution. These results highlight both the promise of geostationary AOD–$PM_{2.5}$ modeling for health applications in northern Thailand and the need for further validation with multi-year datasets and higher-resolution health information.

4. Conclusion

This study developed a preliminary framework to estimate surface-level $PM_{2.5}$ in Chiang Mai using Himawari-9 TOA reflectance, radiative transfer-based AOD retrieval, meteorological drivers, and machine learning. The results showed that season-specific Random Forest models achieved strong predictive skill (R^2 up to 0.82 during haze and 0.76 during non-haze periods), confirming the utility of geostationary satellites for capturing rapid aerosol dynamics. Moreover, the integration of monthly exposure estimates with hospital admissions provided initial evidence of lagged health impacts, particularly for chronic respiratory diseases. These outcomes emphasize the potential of satellite-driven monitoring for public health applications in Southeast Asia. Nonetheless, the findings remain preliminary due to key limitations, including reliance on a single-year dataset, the use of fixed viewing geometry (VZA, RAA) in AOD retrieval, and dependence on monthly aggregated health data. Future work should extend this framework to multi-year and multi-sensor observations, incorporate actual geometric parameters to refine retrieval accuracy, and integrate higher-frequency health datasets to strengthen epidemiological evidence and policy relevance.

References

Abidin, M., et al. (2025). Environmental health risks and impacts of $PM_{2.5}$ exposure on human health in residential areas, Bantul, Indonesia. Toxicology Reports, https://doi.org/10.1016/j.toxrep.2025.101949

Bran, M., et al. (2022). Surface $PM_{2.5}$ concentrations during the dry season over northern Thailand: Sensitivity to model aerosol chemical schemes and the effects on regional meteorology. https://doi.org/10.1016/j.atmosres.2022.106303

Chansuebsri, K., et al. (2024). Chemical composition and origins of $PM_{2.5}$ in Chiang Mai (Thailand) by integrated source apportionment and potential source areas. https://doi.org/10.1016/j.atmosenv.2024.120517.

Chen, X., Li, K., Yang, T., et al. (2024). Trends and drivers of aerosol vertical distribution over China from 2013 to 2020: Insights from integrated observations and modeling. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2024.170485.

Shi, B., et al. (2020). Urbanization and regional air pollution across South Asian developing countries: A nationwide land use regression for ambient $PM_{2.5}$ in Pakistan. Environmental Pollution. https://doi.org/10.1016/j.envpol.2020.115145

Tan, W., et al. (2022). Assessing spatiotemporal variations of AOD in Japan based on Himawari-8 L3 data. Atmospheric Pollution Research. https://doi.org/10.1016/j.apr.2022.101439

Thao, P., et al. (2022). Public health and economic impact assessment of $PM_{2.5}$ from open biomass burning over Southeast Asia during haze episodes. Atmospheric Pollution Research. https://doi.org/10.1016/j.apr.2022.101418

Yang, Z., Zdanski, C., Farkas, D., Bang, J., & Williams, H. (2020). Evaluation of Aerosol Optical Depth (AOD) and $PM_{2.5}$ associations for air quality assessment. Remote Sensing Applications: Society and Environment, 20, 100396. https://doi.org/10.1016/j.rsase.2020.100396