



UDC 528.8:004.8:551.51

TEMPORAL AND SPATIAL PATTERNS OF PM_{2.5} IN MAINZ, GERMANY: INSIGHTS FROM SATELLITE OBSERVATIONS AND MACHINE LEARNING

 Felien TEE¹, Hakan LANE²
¹*Applied Geo-Information Science, HAS Green Academy, 's-Hertogenbosch, Netherlands*
²*Johannes Gutenberg University, Mainz, Germany*

Article History:

- received 9 August 2025
- accepted 25 April 2026

Abstract. Fine particulate matter (PM_{2.5}) remains one of the most harmful air pollutants affecting human health and urban environments. Although ground-based monitoring stations provide accurate measurements, their limited spatial coverage constrains detailed assessments of local air quality. This study investigates the temporal and spatial distribution of satellite-derived PM_{2.5} concentrations in Mainz, Germany, between 2017 and 2022 using statistical and machine learning approaches. Monthly and annual PM_{2.5} data from the Atmospheric Composition Analysis Group were combined with meteorological variables obtained from the WorldClim database. Temporal trends were evaluated using linear regression analyses of monthly and seasonal averages, while spatial patterns were modelled using Thin Plate Spline (TPS) interpolation and Random Forest (RF) regression. The results demonstrate a consistent decline in annual PM_{2.5} concentrations throughout the study period, with the strongest reductions observed during winter. Summer and autumn exhibited relatively stable concentration levels. Spatial modelling revealed persistent pollution hotspots in the southern part of Mainz, whereas lower concentrations occurred in the northern and peripheral areas. TPS produced smooth and interpretable spatial surfaces suitable for visualization, while RF captured additional spatial variability by incorporating meteorological predictors. However, the RF model was limited by the availability of climate data and the exclusion of anthropogenic variables such as traffic intensity and industrial emissions. The findings demonstrate that integrating satellite observations with geospatial modelling provides valuable insights into urban air pollution dynamics and supports environmental monitoring where ground-based measurements are sparse. The proposed workflow offers a practical framework for regional air quality assessment and future spatial prediction studies.

Keywords: PM_{2.5}, air pollution, satellite data, urban air quality, Thin Plate Spline interpolation, Random Forest machine learning.

✉Corresponding author. E-mail: f.d.tee@outlook.com

1. Introduction

Air pollution remains one of the leading environmental threats to human health. Fine particulate matter (PM_{2.5}) is particularly harmful because it penetrates deep into the respiratory system and is associated with cardiovascular and respiratory diseases (Pope & Dockery, 2006). Although PM_{2.5} concentrations have declined across Europe due to stricter emission regulations, concentrations in many urban areas still exceed the WHO guideline (European Environment Agency [EEA], 2022; World Health Organization, 2021).

Because only two official PM_{2.5} monitoring stations operate in Mainz, satellite-derived PM_{2.5} products provide valuable complementary information with continuous spatial coverage. The Atmospheric Composition Analysis

Group dataset combines satellite observations, atmospheric modelling and ground measurements to estimate ground-level PM_{2.5} concentrations.

Previous studies have applied both spatial interpolation and machine learning techniques to analyse PM_{2.5} distributions. However, direct comparisons of Thin Plate Splines and Random Forest using the same satellite-derived dataset in a medium-sized European city remain limited.

This study analyses temporal and spatial PM_{2.5} patterns in Mainz (2017–2022) using satellite-derived observations (Hammer et al., 2022), linear regression, Thin Plate Spline interpolation and Random Forest modelling. The objective is to evaluate temporal trends, seasonal variation and the applicability of both spatial modelling approaches for urban air quality assessment.

2. Materials and methods

2.1. Research area

The study was conducted in the city of Mainz, located in Rhineland-Palatinate, Germany. Mainz is a medium-sized urban area characterized by residential neighbourhoods, industrial zones, major transport corridors, and surrounding agricultural land. These contrasting land uses make the city suitable for analysing spatial variation in air pollution.

A rectangular bounding box surrounding Mainz was defined to ensure that all spatial datasets shared an identical spatial extent during preprocessing and analysis. The bounding box also included neighbouring grid cells required for interpolation, thereby reducing edge effects near the administrative boundary. Figure 1 illustrates both the study area and the satellite sampling grid. Because only two official PM_{2.5} monitoring stations are located within Mainz, satellite-derived observations were used to provide complete spatial coverage.



Figure 1. Study area and satellite data-grid

2.2. Data sources

Annual and monthly PM_{2.5} concentrations for the period 2017–2022 were obtained from the Atmospheric Composition Analysis Group at Washington University in St. Louis (Hammer et al., 2020, 2022; van Donkelaar et al., n.d.). The dataset combines satellite-derived Aerosol Optical Depth (AOD), atmospheric chemical transport models and ground-based observations to estimate ground-level PM_{2.5} concentrations on a regular spatial grid. The study used GWRPM_{2.5} dataset (Shen et al., 2024). Meteorological predictors were obtained from the WorldClim version 2.1 database and consisted of temperature, wind speed and precipitation. These variables were included because they strongly influence atmospheric dispersion and PM_{2.5} transport (Fick & Hijmans, 2017).

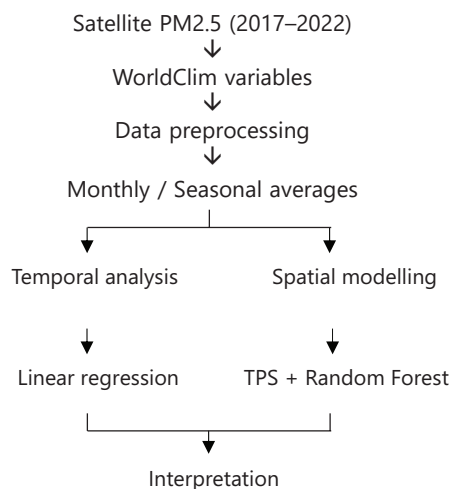
2.3. Data processing

All datasets were processed in RStudio. NetCDF files containing annual and monthly PM_{2.5} concentrations were imported and clipped to the predefined study area. Raster datasets were transformed to a common WGS84 coordinate reference system to ensure spatial consistency between datasets (Van Donkelaar et al., 2021).

Monthly PM_{2.5} observations were aggregated to seasonal and annual averages for temporal analyses. Meteorological variables from WorldClim were spatially matched to the PM_{2.5} raster cells using bilinear extraction. Mean

annual temperature was calculated from minimum and maximum temperature layers, while cumulative annual precipitation and average wind speed were derived for each raster cell before inclusion in the Random Forest model. Quality control included the removal of missing values, verification of raster alignment, and consistency checks across all datasets.

2.4. Workflow



2.5. Statistical analyses

Temporal variation in PM_{2.5} concentrations was analysed using simple linear regression models. Monthly mean concentrations were extracted from the satellite dataset and regressed against time to quantify long-term trends between 2017 and 2022. Seasonal analyses were performed by grouping monthly observations into winter, spring, summer and autumn before fitting separate regression models for each season.

The statistic regression analyses were conducted in R. Model diagnostics consisted of visual inspection of residuals and regression fits. Because the primary objective was descriptive trend analysis rather than prediction, emphasis was placed on the direction and consistency of temporal change rather than statistical significance alone.

2.6. Spatial modelling

Two complementary spatial modelling approaches were applied.

2.6.1. Thin Plate Spline (TPS)

Thin Plate Spline interpolation was used to generate continuous PM_{2.5} concentration surfaces from the satellite observations. TPS was selected because preliminary analyses showed that it produced smooth spatial gradients while avoiding the local artefacts generated by Inverse Distance Weighting. Ordinary Kriging was also evaluated but resulted in fragmented prediction surfaces that poorly represented the observed spatial patterns. Consequently, only TPS was retained for the final comparison.

2.6.2. Random Forest

Random Forest regression was used to investigate whether meteorological variables could improve spatial prediction of PM_{2.5} concentrations. Predictor variables included:

1. mean annual temperature;
2. annual precipitation;
3. mean wind speed.

The model was developed using the 2017 dataset because complete meteorological predictor layers were available for that period. Later years were reserved for independent comparison of spatial patterns rather than model training. Random Forest was implemented using the Random Forest package in R.

Model performance was evaluated by comparing predicted and observed PM_{2.5} values and by examining variable importance scores.

3. Results

3.1. Temporal trends

3.1.1. Monthly trends

Monthly PM_{2.5} concentrations showed a consistent downward trend between 2017 and 2022 (Figure 2). Annual mean concentrations gradually declined throughout the study period, while monthly values displayed the expected seasonal variability. Concentrations were generally highest

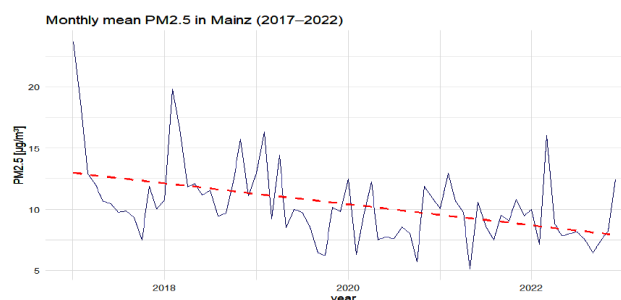


Figure 2. Monthly and annual PM_{2.5} concentrations in Mainz (2017–2022)

during the winter months and lowest during summer. A recurring increase was observed in late autumn, particularly in November.

Linear regression of the monthly means indicated a negative overall trend throughout the study period, suggesting a gradual improvement in air quality.

3.1.2. Seasonal trends

Seasonal aggregation revealed clear differences in temporal behaviour (Figure 3). Winter exhibited the strongest decline in PM_{2.5} concentrations over the study period, whereas spring showed a more moderate decrease. Summer and autumn remained comparatively stable, with only slight reductions in mean concentrations.

Separate linear regression models confirmed negative slopes for all seasons, although the magnitude differed between seasons.

3.2 Spatial modelling

3.2.1. Thin Plate Spline

The Thin Plate Spline interpolation generated a continuous representation of PM_{2.5} concentrations across the study area (Figure 4). Higher concentrations were consistently observed in the southern part of Mainz, whereas lower concentrations occurred towards the northern and peripheral areas. The interpolated surface showed gradual spatial transitions without pronounced local artefacts.

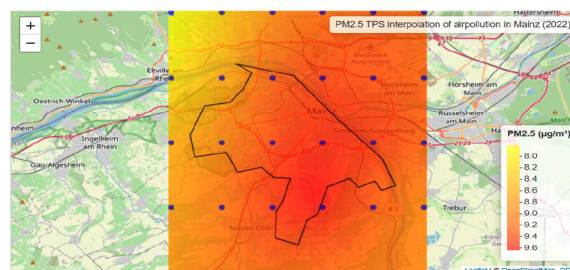


Figure 4. Thin Plate Spline interpolation of PM_{2.5} concentrations

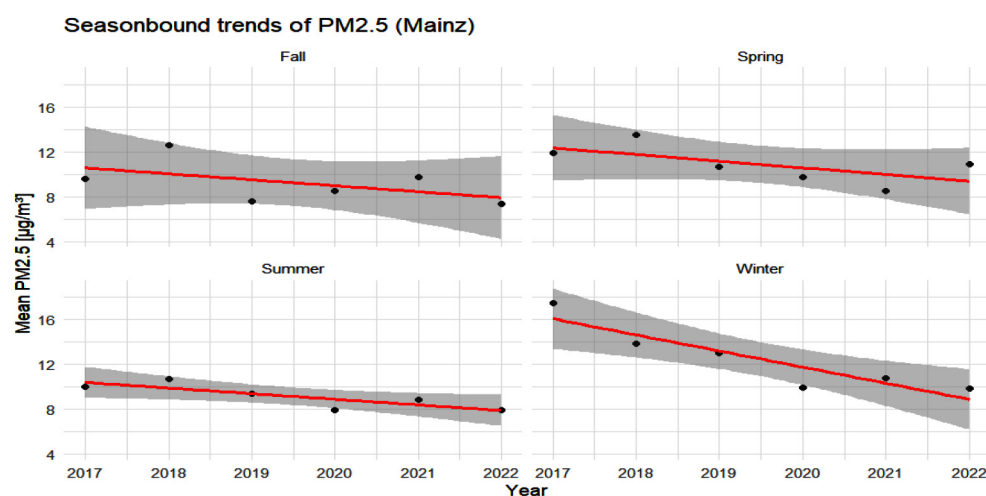


Figure 3. Seasonal PM_{2.5} trends between 2017 and 2022

3.2.2 Random Forest

The Random Forest model produced a spatial prediction based on annual meteorological variables, including temperature, wind speed and precipitation (Figure 5). Similar to the TPS interpolation, the highest predicted PM_{2.5} concentrations occurred in the southern and southwestern parts of the study area.

The Random Forest predictions displayed greater local spatial variation than the TPS interpolation, reflecting the influence of the environmental predictor variables incorporated into the model.

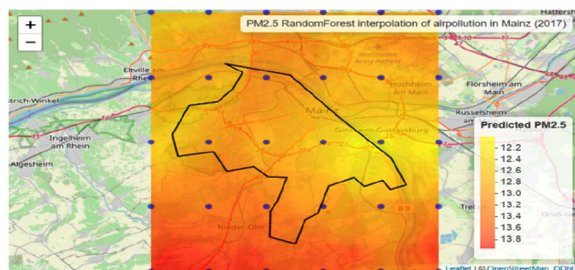


Figure 5. Random Forest map 2017

4. Discussion

This study investigated the temporal and spatial variability of satellite-derived PM_{2.5} concentrations in Mainz, Germany, between 2017 and 2022 by combining statistical trend analysis with two spatial modelling approaches. Overall, the results indicate a gradual improvement in air quality throughout the study period, while persistent spatial differences remained visible across the study area.

The observed decline in PM_{2.5} concentrations is consistent with findings reported elsewhere in Europe. Previous studies have demonstrated that stricter emission legislation, cleaner vehicle technologies, reduced industrial emissions and improvements in residential heating have contributed to declining particulate matter concentrations over the past decade (EEA, 2022; Hammer et al., 2020). The downward trends identified in this study therefore appear to reflect broader regional developments rather than changes unique to Mainz.

Seasonal analyses demonstrated that the strongest reductions occurred during winter, whereas summer and autumn exhibited relatively small changes. Winter PM_{2.5} concentrations are generally influenced by residential heating, increased atmospheric stability and frequent temperature inversions that inhibit vertical mixing of pollutants. Consequently, emission reductions are often most apparent during winter months, while summer concentrations are more strongly controlled by meteorological dispersion processes. Although meteorological variables were not explicitly incorporated into the temporal regression analyses, the seasonal patterns observed in this study are consistent with previous investigations of European urban air quality (e.g., Cameletti et al., 2013; Liang et al., 2020).

Both spatial modelling approaches identified relatively elevated PM_{2.5} concentrations in the southern part of the study area. Although this study did not explicitly analyse emission inventories, these locations coincide with major transport infrastructure and mixed industrial land use surrounding Mainz. The observed spatial gradients therefore appear plausible and correspond with the expected distribution of anthropogenic emission sources. Nevertheless, this interpretation should be regarded as indicative because traffic density, land-use characteristics and industrial emissions were not included as explanatory variables.

The comparison between Thin Plate Spline and Random Forest demonstrates that both methods provide complementary information rather than competing alternatives. Thin Plate Spline generated a smooth and easily interpretable representation of regional PM_{2.5} gradients that is particularly suitable for cartographic visualization. Random Forest incorporated meteorological predictors and therefore captured greater local spatial variability. However, the predictive performance of the Random Forest model remained constrained because only meteorological variables were available as predictors. Important explanatory variables such as traffic intensity, population density, industrial emissions and land use were unavailable and therefore not incorporated into the modelling framework.

Several methodological limitations should be considered when interpreting the results. First, the temporal analyses were based on monthly and annual averages, which inevitably smooth short-term pollution episodes. Second, the study relied on satellite-derived PM_{2.5} estimates rather than dense networks of ground-based monitoring stations. Although the Atmospheric Composition Analysis Group dataset has been extensively validated, satellite products remain indirect estimates of ground-level concentrations. Third, the Random Forest analysis was developed using a single reference year because consistent meteorological predictor data were not available for the complete study period. Consequently, the machine learning results should primarily be interpreted as a demonstration of methodological applicability rather than a fully validated predictive model.

Despite these limitations, the integration of satellite observations, statistical trend analysis and spatial modelling provides a reproducible framework for analysing urban air quality in regions with limited monitoring infrastructure. The workflow presented in this study can readily be transferred to other medium-sized European cities and may support future environmental monitoring, exposure assessment and spatial decision-making.

5. Conclusions

The present study analysed temporal and spatial patterns of satellite-derived PM_{2.5} concentrations in Mainz, Germany, between 2017 and 2022 using statistical trend analysis, Thin Plate Spline interpolation and Random Forest modelling.

The analyses demonstrated an overall decline in PM_{2.5} concentrations during the study period, indicating continued improvements in regional air quality. Seasonal analyses showed that the strongest reductions occurred during winter, while summer and autumn exhibited comparatively stable concentration levels. Spatial modelling consistently identified relatively elevated PM_{2.5} concentrations in the southern part of the study area and lower concentrations towards the north.

Comparison of the two spatial modelling approaches showed that Thin Plate Splines provide smooth and easily interpretable spatial representations suitable for environmental mapping, whereas Random Forest incorporates meteorological information and captures additional spatial heterogeneity. Both approaches therefore offer complementary perspectives for analysing urban air pollution.

Several limitations should be acknowledged. The analyses relied on satellite-derived annual and monthly averages, which do not capture short-term pollution episodes. Furthermore, the Random Forest model included only meteorological predictors and excluded important anthropogenic variables such as traffic intensity, industrial emissions and population density. Future research should therefore combine satellite observations with ground-based monitoring networks and additional explanatory variables to improve predictive performance and strengthen model validation.

Overall, this study demonstrates that satellite-derived PM_{2.5} products, combined with statistical and machine learning techniques, provide an effective framework for assessing spatial and temporal air quality patterns in urban environments where ground-based monitoring is limited.

Acknowledgements

I would like to sincerely thank my supervisor for his valuable guidance throughout this work. His critical review and insightful suggestions greatly improved the quality of this article.

References

- Cameletti, M., Lindgren, F., Simpson, D., & Rue, H. (2013). Spatio-temporal modelling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97, 109–131.
- European Environment Agency. (2022). *Air quality in Europe – 2022 report*.
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315.
- Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R. C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., Brauer, M., Apte, J. S., Henze, D. K., Zhang, L., Zhang, Q., Ford, B., Pierce, J. R., & Martin, R. V. (2020). Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environmental Science & Technology*, 54(13), 7879–7890.
- Hammer, M. S., van Donkelaar, A., Li, A., Lyapustin, A. M., Sayer, N. C., Hsu, R. C., Levy, M. J., Garay, O. V., Kalashnikova, R. A., Kahn, R. A., Brauer, M., Apte, J. S., Henze, D. K., Zhang, L., Zhang, Q., Ford, B., Pierce, J. R., & Martin, R. V. (2022). *Global annual PM_{2.5} grids from MODIS, MISR and SeaWiFS aerosol optical depth (AOD), 1998–2019, V4.GL.03*. [Data set]. NASA Socioeconomic Data and Applications Center (SEDAC).
- Liang, F., Xiao, Q., Huang, K., Yang, X., Liu, F., Li, J., Liu, X., & Gu, D. (2020). The 17-y spatiotemporal trend of PM_{2.5} and its mortality burden in China. *Proceedings of the National Academy of Sciences*, 117(41), 25601–25608. <https://doi.org/10.1073/pnas.1919641117>
- Pope, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association*, 56(6), 709–742. <https://doi.org/10.1080/10473289.2006.10464485>
- Shen, S., van Donkelaar, A., Jacobs, N., Wang, C., & Martin, R. V. (2024). Enhancing global estimation of fine particulate matter concentrations by including geophysical *a priori* information in deep learning. *ACS ES&T Air*, 1(5), 332–345. <https://doi.org/10.1021/acsestair.3c00054>
- van Donkelaar, A., Martin, R. V., & Atmospheric Composition Analysis Group. (n.d.). *V5GL0502 GWR PM_{2.5} (0.05°) annual and monthly data* [Data set]. Washington University in St. Louis. <https://www.satpm.org/v5-gl-05-02>
- van Donkelaar, A., Martin, R. V., Li, C., & Burnett, R. T. (2021). *Global estimates of fine particulate matter (PM_{2.5}), Version 5.GL.05.02, 0.05° × 0.05°* [Data set]. Atmospheric Composition Analysis Group. <https://www.satpm.org/v5-gl-05-02>
- World Health Organization. (2021). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. <https://www.who.int/publications/i/item/9789240034228>