Data in Brief 66 (2026) 112752

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Air quality dataset of particulate matter PM2.5 & PM10 from the monitoring network in Asuncion metropolitan area, Paraguay

Carolina Recalde*, Luis Bernal, Angel Rincón, Diego Stalder

Universidad Nacional de Asunción, Facultad de Ingeniería, Laboratorio de Mecánica y Energía, San Lorenzo, Paraguay

A R T I C L E I N F O

Article history:
Received 11 February 2026
Revised 27 March 2026
Accepted 2 April 2026
Available online 7 April 2026

Dataset link: Metropolitan Asunción Air Quality Monitoring: Processed PM2.5 and PM10 Data with Meteorological Corrections (2019–2021) (Original data)

Keywords:
Urban pollution
Meteorology
Open data
South America
Low-cost sensors
Temporal distribution
Environmental health
Atmospheric aerosols

A B S T R A C T

The present dataset contains three years of continuous, high-frequency air-quality monitoring of particulate matter (PM2.5 and PM10) in the metropolitan area of Asunción and the Western Region of Paraguay. Data were collected from eleven monitoring stations: nine in urban Asunción, one at the National University of Asunción campus in San Lorenzo, and one in the semi-rural Cerrito, Paraguayan Chaco. The monitoring period covers from April 2019 to December 2021. The monitoring network was established to record spatio-temporal variations in particulate matter concentrations throughout urban and semi-rural environments. Key meteorological variables, including temperature, relative humidity, atmospheric pressure, and wind direction, were continuously recorded to enable analysis of dispersion trends. Particulate matter concentrations were measured at five-minute intervals using Optical Particle Counters (OPC–N2), with data automatically transmitted to a central server every thirty minutes for quality control, storage, and processing. This dataset allows direct comparison with national air-quality standards established by the National Resolution No 259/2015 of the Ministry of Environment and Sustainable Development. Its full coverage supports policy evaluation, cross-sector assessments, and studies on environmental and public health impacts. Full technical documentation is included to secure transparency, traceability, and reproducibility, pro-

* Corresponding author.
E-mail address: nrecalde@ing.una.py (C. Recalde).

https://doi.org/10.1016/j.dib.2026.112752

viding the basis for management and scientific research in Paraguay.

Specifications Table

Subject	Earth & Environmental Sciences
Specific subject area	Urban Air Quality Monitoring; Atmospheric Pollution Assessment; Low-cost Sensor Networks.
Type of data	• Raw data: Original, unprocessed CSV files from air quality monitoring stations. • Intermediate data: Cleaned, resampled, and quality-controlled data in CSV format. • Analyzed data: Fully processed, corrected, and aggregated data in CSV format, including meteorological supplementation, AQI calculations, and statistical summaries.
Data collection	Continuous three years measurements were obtained from 2019 to 2021. Data was collected from a network of eleven low-cost air quality monitoring stations (Libelium Smart Environment Pro) equipped with Alphasense OPC–N2 optical particle sensors for PM1, PM2.5, and PM10 detection, and 9370-P sensors for temperature, relative humidity, and atmospheric pressure. Each station was powered by a solar panel and battery system and included GPS localization, SD backup, and 4 G cellular communication. Data were sampled every 5 min and transmitted every 30 min to a central server running a MySQL database.
Data source location	Asunción Metropolitan Area, Paraguay Station grouping by region: • Metropolitan Region: Estacion1, Estacion2, Estacion3, Estacion4, Estacion9, Estacion10 • Central Region – Asunción: Estacion5, Estacion6, Estacion7, Estacion8 • Rural Region – Chaco: Estacion11
Data accessibility	Repository name: Metropolitan Asunción Air Quality Monitoring: Processed PM2.5 Data with Meteorological Corrections (2019–2021). Data identification number: 10.17632/7zxs93dv42.1 Direct URL to data: https://data.mendeley.com/datasets/7zxs93dv42/1
Related research article	• J. C. Pulfer, C. Recalde, L. F. Bernal, F. Barrios, D. F. Palacios, R. Duré, and L. Gavilán, "Implementación de una red de monitoreo de material particulado mp2.5 y mp10 en la ciudad de asunción," 2015. [Online]. Available: http://hdl.handle.net/20.500.14066/2473. • C. Recalde and L. Bernal, "Determinación y análisis de las concentraciones de material particulado medido por una red de monitoreo contínuo para la ciudad de Asunción." Asunción, Paraguay: Asociación Interamericana de Ingeniería Sanitaria y Ambiental, Apr. 2021. • F. Carlés, C. Recalde, C. Sauer, L. Bernal and D. Stalder, "Air Quality Time Series Forecasting Using Machine Learning Algorithms," 2023 XLIX Latin American Computer Conference (CLEI), La Paz, Bolivia, 2023, pp. 1–9, 10.1109/CLEI60451.2023.10346175.

1. Value of the Data

- Regional Data Gaps and Spatial Coverage: This dataset provides the first three-year record, spatially distributed record of continuous PM2.5 and PM10 measurements in Paraguay. By

monitoring three districts' environments – the Metropolitan Area, Central Asunción, and the Rural Chaco. This study allows for a direct comparative analysis between urban and rural settings. This specifically addresses the lack of regional data previously highlighted in global assessment likeGEO-6 [1].
- Modelling and Emission Inventories: The empirical baseline provided in this work is important for calibrating atmospheric dispersion models and refining local emission inventories. In regions where vehicle fleets and fuel types are diesel, these ground-truth observations address the persistent lack of local observations.
- Regulatory Impact and Public Policy: Beyond its technical value, the measurements offer an evidence base for evaluating compliance with the National Resolution N° 259/2015 [2]. These data support the design of data-driven urban policies and emission control strategies that reflect local realities. Health and Epidemiological Insights: When integrated with clinical records, this high-frequency dataset enables researchers to quantify population exposure. Given the sparsity of the monitoring network in Paraguay, these findings are essential for epidemiological studies focused on respiratory health problems within subtropical urban populations. Sensor Validation: Finally, the open-access nature of this dataset facilitates the performance testing of low-cost sensors, such as the Alphasense OPC–N2. Comparing these sensors using reference methods is particularly relevant for methodological research in resource-limited areas characterized by high humidity.

2. Background

Air quality in cities has declined. As a matter of fact, in most urban centers where data are available, concentrations of particulate matter and ozone frequently exceed World Health Organization (WHO) guidelines [3]. This trend increases the vulnerability of urban populations to respiratory diseases, with $>$100 million people in the region living in areas susceptible to air pollution [1]. In Paraguay, the availability of continuous, spatially distributed measurements of fine (PM2.5) and coarse (PM10) particulate matter has historically been limited, constraining the ability to evaluate long-term trends and enforce regulatory standards such as National Resolution of MADES No 259/15 [2].

To address this problem, this work established the first high-density air quality monitoring network dedicated to particulate matter in the Asunción metropolitan area. This network comprises eleven fixed monitoring stations equipped with optical light-scattering sensors (Alphasense OPC–N2) capable for real-time data collection and automatic transmission to a centralized database. The current infrastructure enables continuous, high-frequency monitoring of particulate concentrations across different urban microenvironments.

The data collection and dataset organization were conducted as part of the project "Implementation of a Monitoring Network for PM2.5 and PM10 in the City of Asunción" (PINV15–658), co-funded by the National Council of Science and Technology (CONACYT).[1] The resulting dataset complements existing scientific publications focused on urban air quality characterization in Paraguay, and provides a strong empirical groundwork for future research on the dynamics of atmospheric pollution, emission sources, and public health policies.

3. Data Description

This dataset is organized into three main folders: data, notebooks and figures. Within the data folder, three subdirectories correspond to the stages of the processing workflow: the raw

[1] J. C. Pulfer et al., "Implementación de una red de monitoreo de material particulado MP2.5 y MP10 en la ciudad de Asunción - [Poster]," Asunción, Paraguay: Consejo Nacional de Ciencia y Tecnología (CONACYT), 2020. Accessed: Jan. 26, 2026. [Online]. Available: http://hdl.handle.net/20.500.14066/2473.

one, contains the original, unprocessed CSV files organized by monitoring year; the intermediate, holds cleaned and resampled data along with quality–control reports such as missing data summaries; and the final, stores the fully corrected, analysis–ready datasets. The notebooks directory provides four Jupyter notebooks that document the complete processing sequence—from data ingestion and integrity checks, through meteorological correction and interpolation, to visualization and spatial mapping. Finally, the figures folder includes the publication–ready graphical outputs generated by the workflow.

4. Spatial Organization and Regional Classification

The monitoring network covers three distinct regions of Paraguay (Fig. 1), enabling comparative analysis across Asunción and its area of influence, with the background rural environments:

- **Metropolitan Region** (6 stations): High-density urban areas and major traffic corridors in the Metropolitan Area of Asunción.

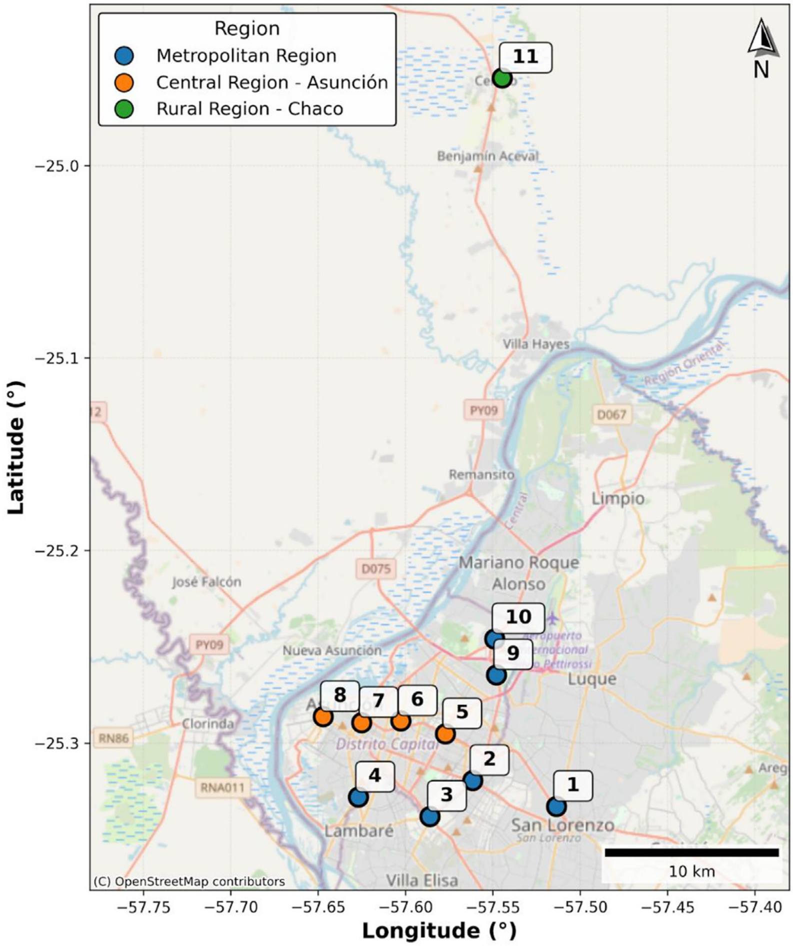

Fig. 1. Spatial distribution of the 11 air quality monitoring stations in the Asunción Metropolitan Area and Chaco region, Paraguay. Stations are color/symbol-coded by regional groups: Metropolitan (blue), Central Asunción (orange), and Rural Chaco (green). Background map: OpenStreetMap. Coordinates in WGS84.

- **Central Asunción Region** (4 stations): Mixed residential-commercial zones within the city center.
- **Rural Chaco Region** (1 station): Background site in the Paraguayan Chaco, representative of regional baseline conditions.

The file **Station_Metadata.csv** includes:

- Geographic coordinates (WGS84 and UTM 21S).
- Station type (urban traffic, urban background, semi-urban, rural background).
- Regional classification (Metropolitan, Central Asunción, Rural Chaco).
- Land use characteristics and data completeness percentage.

5. Raw Data (/DATA/RAW/)

This folder contains 11 CSV files (Estation1.csv to Estation11.csv) with unprocessed 5-minute measurements from April 2019 to December 2021. Each file retains the original structure exported from the acquisition system (Table 1).

Table 1
Structure of raw data files.

Field	Description	Unit of measurement	Data type
ID	Unique identifier used to register each data frame.	N°	int
DATE	Specifies the day, month, and year on which the measurement was recorded.	DD/MM/AA	char(10)
HOUR	Indicates the hour and minute of the observation.	hh:mm	char(6)
MP1	Concentration of ultrafine airborne particles with an aerodynamic diameter $\leq$ 1 µm.	µg/m³	decimal(10,2)
MP2.5	Concentration of fine particulate matter with an aerodynamic diameter $\leq$ 2.5 µm.	µg/m³	decimal(10,2)
MP10	Concentration of coarse particulate matter with an aerodynamic diameter $\leq$ 10 µm.	µg/m³	decimal(10,2)
TEMPERATURE	External ambient air temperature.	°C	decimal(10,2)
RELATIVE HUMIDITY	External ambient relative humidity.	%	decimal(10,2)
ATMOSPHERIC PRESSURE	External ambient atmospheric pressure.	Pa	decimal(10,2)
BATTERY LEVEL	Indicates the charge level of the station's power system.	%	int

6. Main Output File

The complete outputs of the data processing pipeline are organized in the (/data/final/) directory, with accompanying code in the /notebooks/ directory. This structure ensures reproducibility and facilitates direct reuse of the processed data.

Table 2 summarizes the structure and content of the main processed output file generated by the data processing pipeline. This table documents the complete 5-minute time series used for analysis, integrating station identifiers, raw particulate matter measurements, and meteorological variables with interpolated, humidity-corrected, and derived values. In addition to regulatory metrics, such as 24-hour moving averages and daily limit exceedance flags, the dataset includes Air Quality Index (AQI) values and corresponding categorical classifications.

Table 2
Structure of the main processed output file (/data/final/).

Field	Description	Unit of measurement	Data type
DATETIME_UTCS	Coordinated Universal Time (UTC) timestamp of the 5-minute measurement interval.	YYYY-MM-DD hh:mm	datetime
STATION_CODE	Unique identifier of the monitoring station.	–	char
REGION	Geographic region to which the monitoring station belongs.	–	char
MP1	Raw concentration of particulate matter with aerodynamic diameter $\leq$ 1 µm	µg/m³	decimal(10,2)
MP2.5	Raw concentration of particulate matter with aerodynamic diameter $\leq$ 2.5 µm.	µg/m³	decimal(10,2)
MP10	Raw concentration of particulate matter with aerodynamic diameter $\leq$ 10 µm.	µg/m³	decimal(10,2)
MP2.5_INTERPOLATED	MP2.5 values linearly interpolated for gaps $\leq$ 3 h	µg/m³	decimal(10,2)
MP10_INTERPOLATED	MP10 values linearly interpolated for gaps $\leq$ 3 h	µg/m³	decimal(10,2)
INTERPOLATION_FLAG	Indicates whether particulate matter values were interpolated	Boolean	bool
MP2.5_CORRECTED [4]	Humidity-corrected MP2.5 concentration for RH > 80%	µg/m³	decimal(10,2)
MP10_CORRECTED	Humidity-corrected MP10 concentration for RH > 80%	µg/m³	decimal(10,2)
HUMIDITY_CORRECTION_FLAG	Indicates application of humidity correction	Boolean	bool
TEMPERATURE	Ambient air temperature	°C	decimal(10,2)
RELATIVE_HUMIDITY	Ambient relative humidity	%	decimal(10,2)
ATMOSPHERIC_PRESSURE	Ambient atmospheric pressure	Pa	decimal(10,2)
TEMPERATURE_INTERPOLATED	Interpolated temperature values for short gaps	°C	decimal(10,2)
RELATIVE_HUMIDITY_INTERPOLATED	Interpolated relative humidity values for short gaps	%	decimal(10,2)
MP2.5_CORRECTED_24H_AVG	24-hour moving average of corrected MP2.5 concentrations	µg/m³	decimal(10,2)
MP10_CORRECTED_24H_AVG	24-hour moving average of corrected MP10 concentrations	µg/m³	decimal(10,2)
EXCEEDS_PM2.5_DAILY_LIMIT	Indicates exceedance of the daily regulatory limit for PM2.5	Boolean	bool
EXCEEDS_PM10_DAILY_LIMIT	Indicates exceedance of the daily regulatory limit for PM10	Boolean	bool
AQI_PM2.5	Air Quality Index computed from corrected MP2.5 concentrations	AQI	int
AQI_PM10	Air Quality Index computed from corrected MP10 concentrations	AQI	int
CATEGORIA_AQI_PM2.5	Air quality category derived from AQI_PM2.5	—	char

7. Experimental Design, Materials and Methods

This section describes the methodology used to develop the dataset, beginning with the design and deployment of the monitoring network. Then, it covers the instrumentation and measurement principles, and the automated data acquisition and transmission system. Finally, a multi-stage computational pipeline for data processing, quality control, and enhancement is presented.

7.1. Network design and strategic site selection

The monitoring network was established to address a critical scarcity of high-resolution, location-specific air quality data for Paraguay. A strategic site-selection process was used to represent diverse urban typologies a regional background. This process accounts for special variability of particulate matter (PM) concentrations. These eleven (11) fixed monitoring stations were deployed across three defined regions (Fig. 1):

1. **Asuncion Metropolitan Area (6 stations)**: These sensors were positioned along major thoroughfares, such as *Av*. Mariscal López and *Av*. Santísimo Sacramento avenues, as well as within commercial hubs and dense residential zones. This placement enables the characterization of pollution profiles specially associated with cores and high-volume transit corridors.
2. **Central Asunción (4 stations)**: These units are distributed within the administrative boundaries of Asunción, specifically in zones characterized by mixed land use, high population density, and moderate traffic flow. This configuration allows these sites to act as indicators for general urban background conditions, capturing the baseline pollution levels to which the majority of the city residents are exposed.
3. **Rural Chaco (1 station)**: Established in Cerrito, Benjamín Aceval (Chaco), this station is located within a sparsely populated, semi-arid environment. Serving as a regional background site, it provides a baseline minimally affected by local anthropogenic activity. This strategic positioning is essential for distinguishing local pollution from long-range transport events affecting the region.

The site-selection process was governed by four primary criteria to ensure data integrity and representativeness. First, priority was given to locations near identified or potential emission sources, such as major roadways, while ensuring the network reflected dominant land-use categories across the study area. Second, logistical feasibility was essential to guarantee operational continuity, especially regarding reliable electrical power access and physical security. Finally, all sites were evaluated for adherence to air quality monitoring guidelines, maintaining prescribed distances from localized obstructions and immediate interference sources to ensure unobstructed airflow.

7.2. Instrumentation: specifications and measurement principles

Each monitoring node was designed as an autonomous, solar-powered unit centered on the Libelium Smart Environment Pro platform. Its core instrumentation integrates a suite of specialized sensors and communication modules, detailed as follows:

- **Particulate Matter Monitoring**: Aerosol concentrations are measured using an Alphasense OPC–N2 optical particle counter. This instrument uses a 660 nm laser diode and a multi-element photodetector to detect light scattered by individual particles within a 0.38–17 µm size range. The resulting raw particle count is categorized into 24 programmable size bins, which are subsequently converted into mass concentrations (µg/m^3) for PM1, PM2.5, and PM10. These mass calculations are derived via onboard algorithms assuming a constant particle density of 1.65 g/cm^3 and a nominal flow rate of 1.2 L/min.
- **Meteorological Monitoring**: Complementary meteorological parameters are monitored via a Libelium 9370-P sensor. Integrating probes for temperature (±0.5 °C accuracy), relative humidity (±3% accuracy), and atmospheric pressure (±1 hPa accuracy).
- **Communication, Localization and Power Management Hardware**: Operational reliability is supported by a 4 G cellular modem (GSM/GPRS) for data transmission. In addition to this, geography integrity is maintained via a GPS module. To prevent data loss during network outages, each node uses a Secure Digital (SD) card for local backup logging. Finally, the system is engineered for complete energy autonomy, featuring a 20 W solar panel coupled with a 12V/7Ah sealed lead-acid battery.

Fig. 2. Solar-powered air quality monitoring station based on the Libelium Smart Environment Pro platform, integrating an Alphasense OPC–N2 sensor for PM_1, $PM_{2.5}$, and PM_{1n} measurements, a meteorological sensor, 4 G communication, and an IP65 weatherproof enclosure with autonomous power supply.

The sensors and electronics were shielded within a weatherproof (IP65) enclosure to ensure operational stability under diverse conditions. This protection features a proprietary passive inlet head, specifically designed to reduce particle leakage while preventing interference from rain and insects. The autonomous air quality monitoring stations deployed in this study is illustrated in Fig. 2.

7.3. Data acquisition, transmission, and centralized storage

The data lifecycle, from its initial measurement to centralized repository, was managed through a fully automated pipeline, including:

1. **Acquisition**: At each station, environmental and diagnostic parameters were sampled at a fixed 5-minute interval. The microcontroller digitized all incoming analog signals from the sensors using a 10-bit Analog-to-Digital Converter ADC.
2. **Local Buffering**: To ensure resilience against cellular network outages, each station implemented a First-In-First-Out (FIFO) buffer in non-volatile memory. The implemented mechanism locally stores every 5-minute record (timestamp, PM and meteorological values, battery status, GPS coordinates) until successful transmission is confirmed.
3. **Secure Transmission**: Every 30 min, the station initiates a Transmission Control Protocol (TCP) session over the 4 G network with a central server. A custom, lightweight application-layer protocol managed packetization and flow control. This protocol used an acknowledgment-based (ACK) handshake signal to guarantee reliable data delivery over the network.
4. **Central Storage**: Incoming data packets from all 11 stations were parsed and ingested in real-time into a MySQL relational database hosted on a dedicated server. This database served as the primary and versioned repository for all raw measurements.

7.4. Computational data processing pipeline

All data processing steps are documented in reproducible Jupyter notebooks available in the /notebooks/ folder. The workflow is carried out as follows:

Step 1. Data Integrity and Cleaning (DataIntegrity.ipynb)

- **Flexible parsing**: Automatic detection of CSV format (delimiter, quoting).
- **Duplicate removal**: Duplicate records (same station, date, time) are removed.
- **Missing packet identification**: Gaps in the 5-minute timestamp sequence are logged.
- **Resampling**: All data are uniformly resampled to a consistent 5-minute frequency; missing intervals are flagged.

Step 2. Meteorological Supplementation and Gap Filling (CorrectionCalculation.ipynb)

- **Meteostat API integration**: Missing temperature and humidity data for Station 11 (Chaco) are supplemented using the nearest reliable weather station (Cerrito Benjamín Aceval).
- **Short-gap interpolation**: Gaps $\leq$3 h in PM and meteorological series are filled via linear interpolation.
- **Long-gap handling**: Gaps $>$3 h are preserved as NA to avoid over-imputation.

Step 3. Humidity Correction

PM2.5 and PM10 concentrations are corrected for high relative humidity ($>$80%) using the Alphasense OPC–N2 correction formula [5]:

$$\boldsymbol{PM_{corrected}} = \frac{\boldsymbol{PM}}{1 + \boldsymbol{\alpha}\ \boldsymbol{x}\ \left(\boldsymbol{R_H} - \boldsymbol{R_{ref}}\right)},$$

where $\alpha \approx 0.014$ and $R_{ref} = 40\%$ (Primary physical reference site) [6].

Step 4. Temporal Aggregation and AQI Calculation

- **Hourly averages**: Computed from valid 5-minute measurements.
- **24-hour moving averages**: Centered rolling averages.
- **Air Quality Index (AQI)**: Hourly AQI values are computed separately for PM2.5 and PM10 using the U.S. Environmental Protection Agency (EPA) AQI calculation method [7], which employs piecewise linear interpolation between concentration breakpoints.

$$AQI = \frac{I_{high} - I_{low}}{C - C_{low}} \times (C - C_{low}) + I_{low}$$

where:

$$C = measured\ pollutant\ concentration\ \left(\mu g/m^3\right)$$

$$C_{low},\ C_{high} = concentration\ breakpoints\ bracketing\ C$$

$$I_{low},\ I_{high} = AQI\ breakpoints\ to\ C_{low}\ and\ C_{high}$$

Step 5. Regional Statistical Summaries

a. PM2.5 and PM10 distributions are summarized by region (Metropolitan, Central Asunción, Rural Chaco), year, and month.
b. Metrics include the mean, median, 95th percentile, and exceedance-day counts relative to national standards.

8. Visualization Outputs

The processing pipeline generates a comprehensive set of publication-quality visualizations in PNG (300 DPI) and SVG formats through Visualize.ipynb and mapa.ipynb.

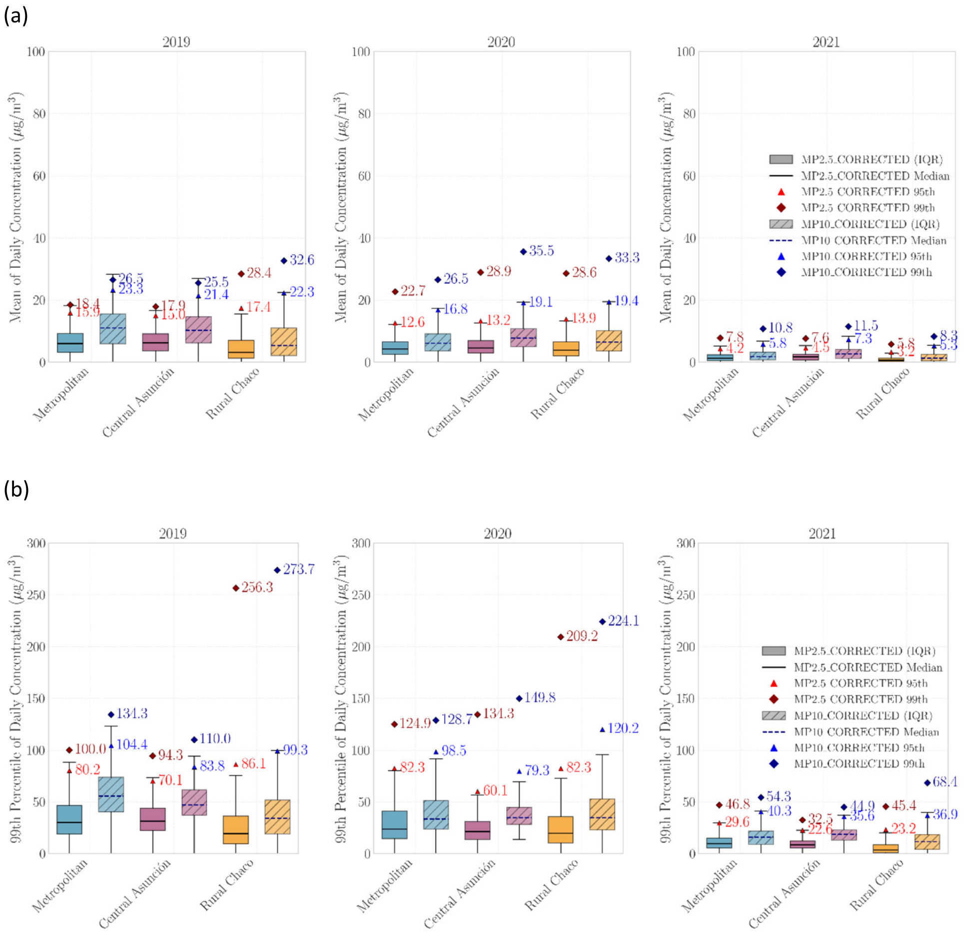

Fig. 3. Comparative boxplots showing PM2.5 and PM10 distributions across urban, semi-urban, and rural environments by period 2019–2021. (a) Mean values, (b) 90th Percentile.

8.1. Particulate matter distribution analysis

Fig. 3 illustrates the statistical distribution of PM2.5 and PM10 concentrations throughout urban (Metropolitan), semi-urban (Central Asunción), and rural (Chaco) environments for all three annual periods analyzed. The boxplots reveal a pronounced spatial gradient, with systematically higher median concentrations and wider interquartile ranges at urban stations, intermediate levels at semi-urban sites, and substantially lower values at the rural background station.

Urban stations showed pronounced right-skewed distributions, particularly for PM10, reflecting episodic high-concentration events associated with traffic density, road dust resuspension, and other localized anthropogenic sources. Semi-urban stations showed similar but attenuated patterns, indicating reduced emission intensity and greater dispersion capacity. In contrast, the rural background station displayed narrow distributions with low medians and limited upper tails, consistent with background conditions minimally influenced by direct anthropogenic sources as road transport.

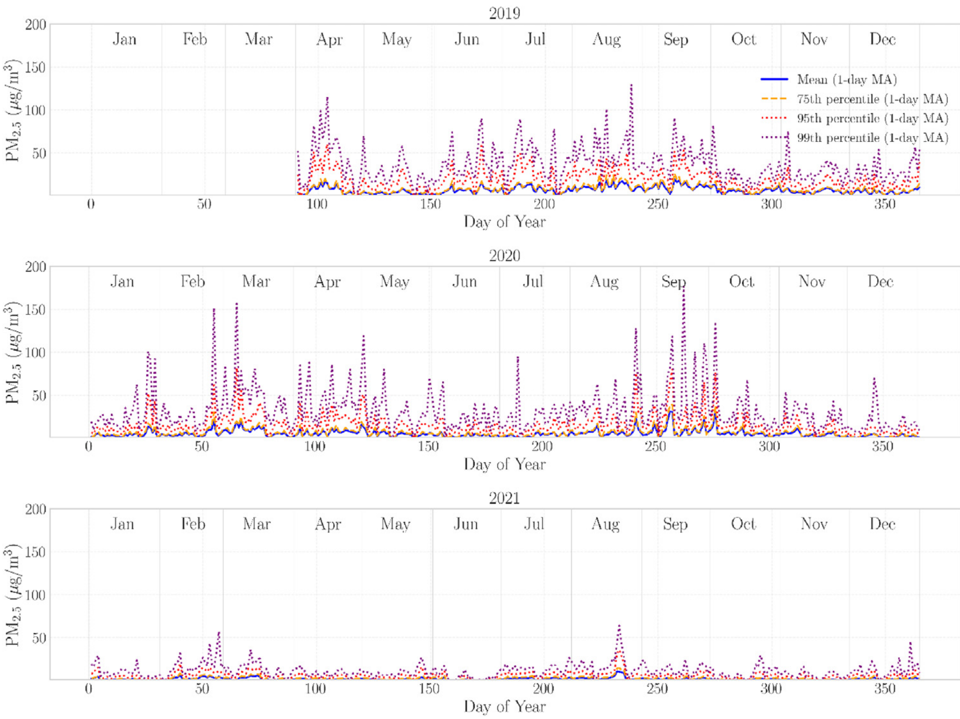

Fig. 4. Temporal evolution of regional PM2.5 concentrations and variability during 2019–2021.

Across all regions, both PM2.5 and PM10 distributions showed a consistent downward shift from 2019 to 2021, accompanied by a contraction of the 95th and 99th percentiles. This temporal evolution suggests a reduction in extreme pollution episodes over the study period, most notably in urban areas.

8.2. Temporal trend analysis

Figs. 4 and 5 illustrate the temporal evolution of mean PM2.5 and PM10 concentrations, respectively. Together, with the 75th and 95th percentile bands, they allow the assessment of both central tendency and variability across regions. For both pollutants, urban and semi-urban regions consistently exhibited higher mean concentrations and greater temporal variability compared to the rural site throughout the three-year period.

PM2.5 concentrations showed a gradual decline across all three annual periods in all regions, with the steepest reduction observed between 2020 and 2021. Urban stations showed the greatest temporal variability, reflected in wider percentile bands and recurrent short-term concentration peaks, indicative of fluctuating combustion-related emissions. Semi-urban stations followed similar trends with reduced amplitude, suggesting transitional pollution dynamics between dense urban cores and background conditions.

PM10 exhibited analogous long-term behavior with higher dispersion and more pronounced peaks, particularly in urban areas. This shows a stronger influence from coarse particle sources, such as mechanical abrasion and resuspension processes. On the other hand, the rural Chaco station maintained low and relatively stable PM2.5 and PM10 concentrations over time, with minimal separation between mean and percentile curves. This supports its function as a regional background reference.

Fig. 5. Temporal evolution of regional PM10 concentrations and variability during 2019–2021.

8.3. Assessment of air quality health categories and temporal AQI patterns

Figs. 6 and 7 combine stacked bar charts of Air Quality Index (AQI) category frequencies with the time series of AQI values derived from PM2.5 and PM10 concentrations, respectively. These visualizations highlight the strong contrasts in air quality health risk across urban, semi-urban, and rural environments.

Urban and semi-urban regions exhibit higher frequencies of “Moderate” and “Unhealthy for Sensitive Groups” AQI categories [7], particularly during 2019 and 2020, reflecting elevated particulate matter concentrations and greater exposure potential for the population. In contrast, the rural Chaco site is dominated by the “Good” AQI category throughout all three periods, consistent with its low particulate burden.

Temporal AQI trends indicate a progressive reduction in both the frequency and magnitude of adverse air quality episodes across all regions. This improvement is most evident in the urban environment, where high-AQI events become less frequent and less intense by 2021. These patterns corroborate the observed decline in PM2.5 and PM10 concentrations and emphasize the combined spatial and temporal heterogeneity of particulate-related health risks in the Asunción Metropolitan Area.

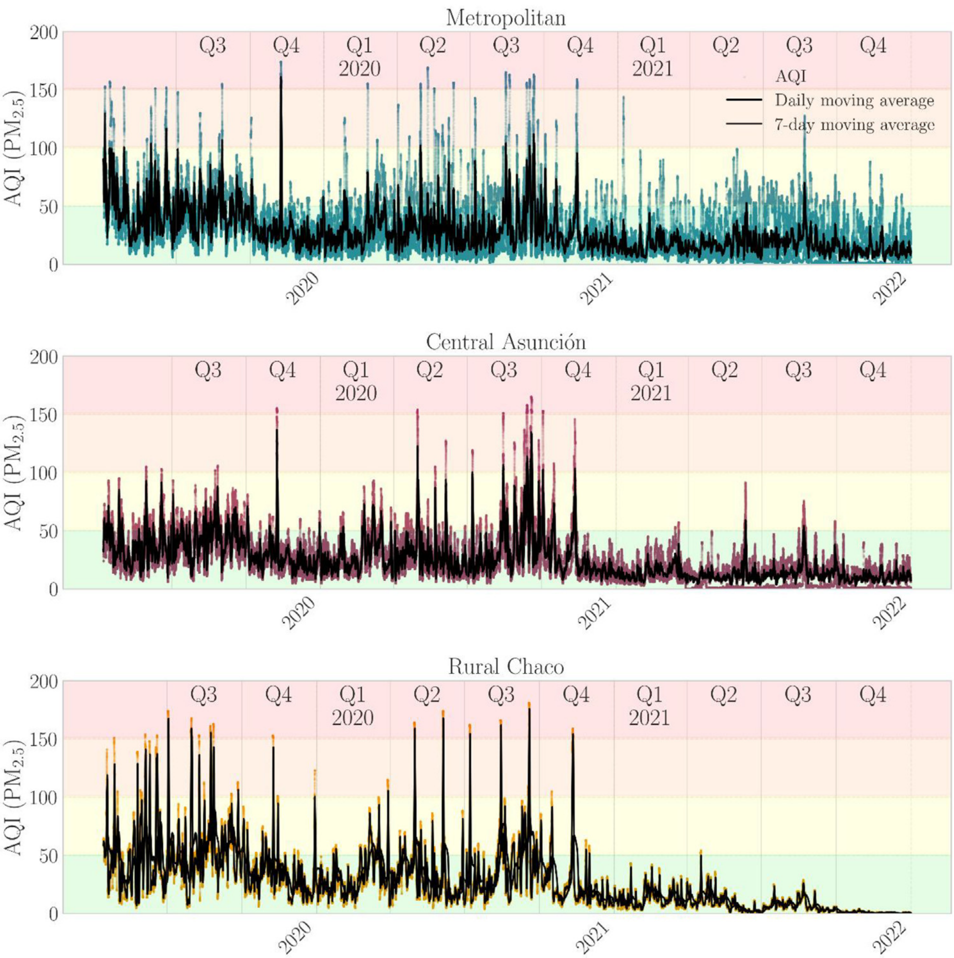

Fig. 6. Regional distribution and temporal evolution of PM2.5-based Air Quality Index categories across 2019–2021.

8.4. Statistical results

The statistical results, summarized in Table 3, highlight clear spatial and temporal differences in PM2.5 and PM10 concentrations across Metropolitan (urban), Central Asunción (semi-urban), and Rural Chaco environments during the 2019–2021 period. In 2019, the Metropolitan region exhibited elevated mean concentrations for both PM2.5 (11.4 µg/m^3) and PM10 (17.9 µg/m^3), reflecting the influence of dense traffic and urban activities. Central Asunción showed slightly lower mean values, consistent with its semi-urban character. Meanwhile, Rural Chaco unexpectedly recorded the highest mean concentrations in 2019, particularly for PM10 (23.3 µg/m^3), accompanied by extremely high standard deviation (102.2 µg/m^3). This indicates sporadic but intense coarse-particle events, likely linked to natural sources and resuspension under dry conditions rather than continuous anthropogenic emissions. Across all regions and years, PM10 presented substantially higher standard deviations than PM2.5, indicating greater temporal variability and episodic behavior of coarse particles compared to fine particulate matter.

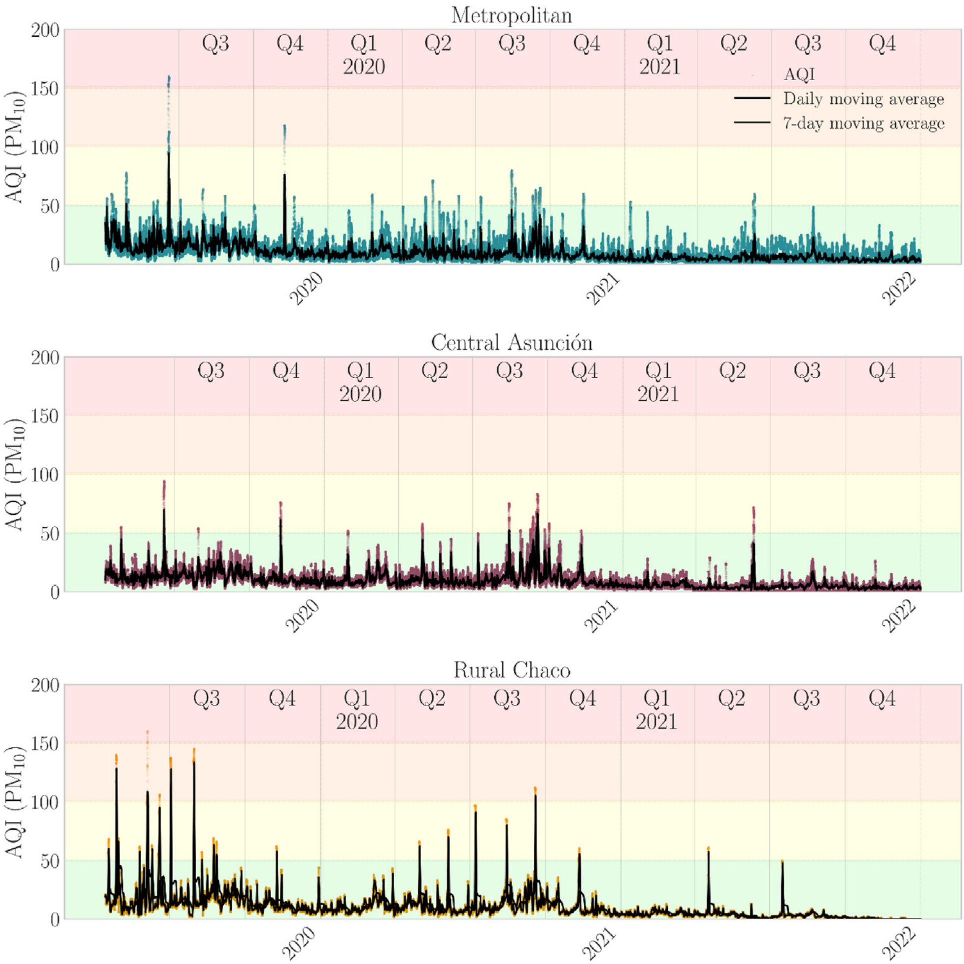

Fig. 7. Regional distribution and temporal evolution of PM10-based Air Quality Index categories across 2019–2021.

The observed evolution of the decline in particulate matter concentrations, from 2019 to 2021, is consistent for all regions. In the Metropolitan area, mean PM2.5 decreased from 11.4 µg/m³ in 2019 to 4.5 µg/m³ in 2021, while PM10 dropped from 17.9 to 5.9 µg/m³. Central Asunción followed a similar trajectory, with PM2.5 means declining from 9.5 to 3.2 µg/m³ and PM10 from 14.9 to 5.2 µg/m³. The most pronounced reduction occurred in the Rural Chaco, where mean PM2.5 fell from 14.7 µg/m³ in 2019 to 1.9 µg/m³ in 2021, and PM10 from 23.3 to 3.7 µg/m³. This sharp decrease is accompanied by a substantial reduction in standard deviation, indicating not only lower average concentrations but also a disappearance of extreme episodic events by 2021.

Table 3
Statistical results. Air quality index by region, by year and PM studied.

Region	Year	Pollutant	N	Mean (μg/m³)	Median (μg/m³)	Standard Deviation (μg/m³)
Metropolitan	2019	PM2.5	389.219	11.4	6.3	18.8
		PM10	389.219	17.9	10.5	50.7
	2020	PM2.5	619.544	8.4	4.4	14.9
		PM10	619.544	11.1	6.3	19.7
	2021	PM2.5	630.720	4.5	2.3	8.3
		PM10	630.720	5.9	3.0	12.7
Central Asunción	2019	PM2.5	309.194	9.5	5.9	12.7
		PM10	309.194	14.9	9.5	31.5
	2020	PM2.5	410.029	8.3	4.5	13.5
		PM10	410.029	12.3	7.2	23.6
	2021	PM2.5	420.480	3.2	2.0	5.0
		PM10	420.480	5.2	2.9	15.2
Rural Chaco	2019	PM2.5	75.557	14.7	6.5	29.9
		PM10	75.557	23.3	10.2	102.2
	2020	PM2.5	105.408	9.9	4.2	22.6
		PM10	105.408	14.3	7.0	34.5
	2021	PM2.5	105.120	1.9	0.8	4.7
		PM10	105.120	3.7	1.3	17.8

Limitations

- The distribution of the monitoring sensors within the Asunción Metropolitan Area provides good spatiotemporal resolution of particulate matter concentrations for the study area. Therefore, it cannot be generalized nationwide.
- The low-cost sensors for data collection are factory-calibrated and were validated before service. Their precision may vary over the time interval of the study. At the time of submission of the present study, the National Normalized-Metric Institute of Paraguay (Instituto Nacional de Tecnología y Normalización - INTN) does not provide calibration certification for these instruments.
- For humidity-correction procedures, the humidity value in the central area of Asunción has been used as a reference. This limits the precision of the estimated corrected concentration values at each monitoring station within the Asunción Metropolitan Area.

Ethics Statement

The authors confirm that they have read and comply with the ethical requirements for publication in Data in Brief. This work does not involve human subjects, animal experiments, or any data collected from social media platforms. During the preparation of this work the authors used chatGPT in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

Credit Author Statement

Carolina Recalde: Conceptualization, methodology, data validation, writing and original draft preparation. Luis Bernal: Methodology, software and data validation. Angel Rincón: Supervision, translation, analysis results and writing-reviewing. Diego Stalder: Supervision, data analysis, visualization and writing-reviewing.

Data Availability

Metropolitan Asunción Air Quality Monitoring: Processed PM2.5 and PM10 Data with Meteorological Corrections (2019–2021) (Original data) (Mendeley Data).

Acknowledgements

This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT) of Paraguay through Project PINV15-658, PINV01-40. The authors would like to thank Jean Claude Pulfer for his collaboration in the development of this Project, and to Cristhian Zárate for reviewing the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] United Nations Publications, Global Environment Outlook 6 (Geo-6): Regional Assessment For Latin America and the Caribbean, vol. 1. United Nations Environment Programme, 2017
[2] Ministerio del Ambiente y Desarrollo Sostenible (MADES), Resolución N° 259/15 “POR LA CUAL SE ESTABLECE PARÁMETROS PERMISIBLES DE CALIDAD DEL AIRE.” Paraguay, 2015.
[3] World Health OrganizationWHO Global Air Quality guidelines: Particulate Matter (PM2.5 and PM10), ozone, Nitrogen dioxide, Sulfur Dioxide and Carbon Monoxide, WHO European Centre for Environment and Health, Geneva, 2021.
[4] L.R. Crilley, et al., Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring, Atmos. Meas. Tech. 11 (2) (Feb. 2018) 709–720, doi:10.5194/amt-11-709-2018.
[5] H. Hojaiji, H. Kalantarian, A.A.T. Bui, C.E. King, M. Sarrafzadeh, Temperature and humidity calibration of a low-cost wireless dust sensor for real-time monitoring, IEEE Sens Appl Symp (SAS), IEEE, Mar. 2017, doi:10.1109/SAS.2017.7894056.
[6] Alphasense Ltd., “Application Note AAN 003 ‘NON-LINEARITY OF MASS FLOW CONTROLLED OXYGEN SENSORS,’” Braintree, Essex, UK.
[7] U.S. Environmental Protection Agency, “Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI),” Research Triangle Park, NC, Sep. 2018.