

IIDEPROQ Scientific Journal
Vol. 3 • 2026

DEVELOPMENT OF A PLATFORM FOR VISUALIZATION AND ANALYSIS OF AIR QUALITY TIME SERIES

Mishel Cortez[1], Iris Jabneel Calle[2,*], Marcelo Antonio Gorritty[3]

ABSTRACT

This study aimed to develop an interactive platform for visualizing and analyzing air-quality time series in Tableau®, using PM10 records from three automatic monitoring stations of the MoniCA network in Cochabamba. The work was intended to strengthen environmental management through a tool integrating data cleaning, statistical analysis, and dynamic visualization.

The methodology comprised the systematization of hourly records for 2019-2023, the removal of negative and outlying values, the assessment of data sufficiency, the treatment of missing values, and time-series decomposition. As an operational acceptance criterion, the proportion of valid data in each series was evaluated, adopting 75% valid data as a reference threshold to support comparable analyzes across periods and stations.

PM10 concentrations from Fuerza Aerea, Parque Kanata, and SEMAPA were obtained through continuous automatic monitoring. The integration of Tableau® with Python/ TabPy enabled imputation routines, descriptive metrics, and graphical outputs aimed at identifying trends, seasonality, and time windows with elevated concentrations relative to the applicable standards.

Results showed marked differences in data completeness among stations and a consistent seasonal pattern, with higher PM10 concentrations during the dry and winter period. Moving averages proved useful for descriptive analysis of short gaps, whereas linear interpolation was used for time-series decomposition because a complete sequence was required. Overall, the platform improved analytical exploration of the data and proved useful as a technical support tool for interpreting air-quality information.

Keywords: air quality, PM10, time series, data visualization, Tableau®.

1. INTRODUCTION

Globally, air pollution is one of the main environmental risks to human health due to its association with respiratory and cardiovascular diseases (WHO, 2024; Röbbel, n.d.). In Bolivia, this problem is particularly relevant in urban centers where vehicle emissions, industrial activities, and weather conditions that limit the dispersion of pollutants converge.

[1] Environmental Engineering Program, Faculty of Engineering. UMSA
[2] IIDEPROQ. Environmental Engineering. Faculty of Engineering. UMSA
[3] IIDEPROQ. Environmental Engineering. Faculty of Engineering. UMSA
***Corresponding author:** jabneeliris@gmail.com

https://doi.org/10.53287/2026.3p7k0w4f

Air pollution represents one of the main environmental problems in urban centers, particularly in Bolivia, where emissions from the motor vehicle fleet, industrial and commercial activity are the predominant sources (MMAyA, 2021).

In cities like Santa Cruz, La Paz, and Cochabamba, air pollution has become a major public health problem. In Cochabamba, the valley topography and high traffic density contribute to the accumulation of particulate matter, making systematic monitoring of PM10 essential for interpreting changes in air quality over time. Motor vehicles are responsible for approximately 70% of the country's air pollution, exacerbated by forest fires and indiscriminate burning practices (El País, 2021).

In the city of Cochabamba, one in five people owns a vehicle. Motor vehicles are responsible for 92% of particulate matter smaller than 10 microns (PM10) and nitrogen oxides (NOx), making them the primary source of pollutants. Environmental audit report K2/AP06/Y13 indicates that the Cochabamba Valley is particularly susceptible to pollution due to its topographical and climatic characteristics (Acosta, 2022).

Since 2000, thanks to the Clean Air project financed by COSUDE, the installation of air quality monitoring networks has been promoted, consolidating into the National Network MoniCA Bolivia (MMAYA, 2021).

The MoniCA network is present in eight departments of the country, and employs passive, active, and automatic methodologies for measuring criteria pollutants such as PM10, NO_2, CO and O_3. The objectives of the monitoring include describing the temporal behavior of concentrations, verifying compliance with air quality regulations, generating information for environmental management, and providing input for risk communication and the activation of response measures when appropriate. However, the lack of continuity in the records and the heterogeneity in the cleaning and analysis procedures have limited the technical use of these databases. Sufficient information and high-quality records are essential for time series to be interpretable and comparable across stations and periods.

Although data from the MoniCA network is available in the National Environmental Information System (SNIA), there is no platform that allows for dynamic and interactive analysis of time series. This prevents in-depth analysis of seasonal patterns, long-term trends, and comparisons between periods and seasons. The lack of such a tool limits comprehensible access to information for both decision-makers and the general public.

In this context, tools like Tableau® offer a viable alternative for integrating data cleaning routines, descriptive metrics, time decomposition, and interactive visualization. Rather than simply generating 'intuitive visualizations,' the focus of this study is on structuring a reproducible analytical environment that allows for the examination of trends, seasonality, concentration distribution, and comparison with normative limits.

Finally, the development of a digital air quality analysis platform addresses not only a technical need, but also a social, environmental, and academic one. It facilitates timely decision-making in public health, promotes the use of modern visualization and analysis methodologies, and contributes to building analytical capacity in the field of environmental management. Furthermore, it lays the groundwork for future research in predictive modeling and environmental monitoring.

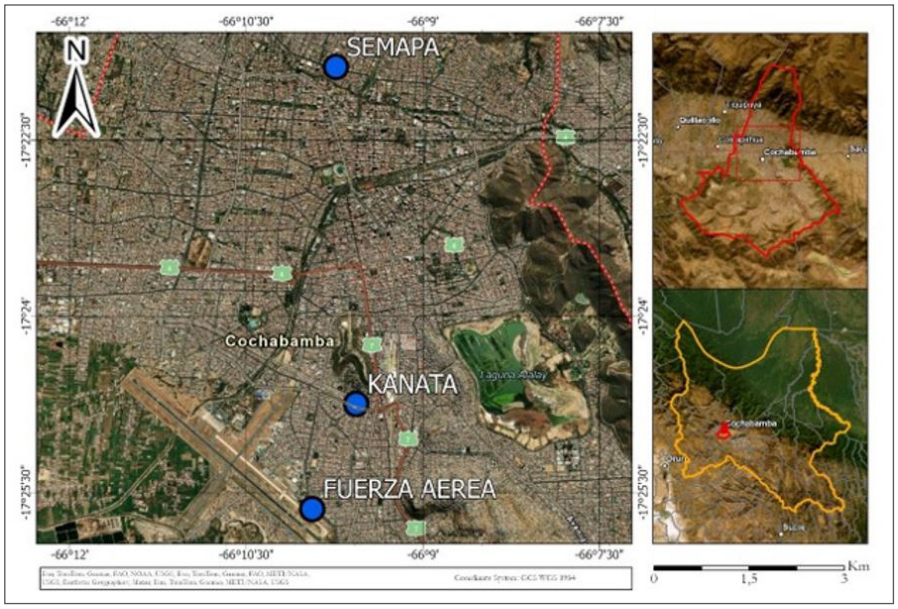
2. MATERIALS AND METHODS

The data analyzed comes from the MoniCA Network of Cochabamba and corresponds to the PM10 pollutant measured in three active automatic stations: Fuerza Aérea (E1), Parque Kanata (E2) and SEMAPA (E3).

The study period spanned five years (2019–2023) and included hourly records and, for some periods, 30-minute resolution series that were subsequently standardized to hourly resolution. Measurements were taken using continuous automatic monitoring with Thermo equipment based on beta particle attenuation (Bascopé, 2001).

PM10 was selected because this variable showed the greatest relative continuity within the available data set, allowing for a consistent temporal analysis. The locations of the three active automatic monitoring stations are shown below.

Figure 1. Visualization of the location of the monitoring stations



Source: Cortez, (2025)

2.1. Time series processing with Tableau® tools

All negative values were filtered out because negative concentrations of air pollutants have no physical meaning. The removal was applied automatically at the data source to prevent erroneous records from affecting descriptive metrics and subsequent analysis routines.

Outlier removal was performed using 4 methodologies:

i) Interquartile range (IQR, Index Quart),
ii) Z-score method,
iii) clustering method using k-means and
iv) method by calculating percentiles using window functions, the latter being the one that best maintains the behavior of the series (Aggarwal, 2016)

The handling of missing values was performed using Python scripts integrated into Tableau® via TabPy. At this stage, an explicit criterion for data sufficiency was incorporated: for comparative purposes, a series was considered to have at least 75% valid data to support robust interpretations. Furthermore, the interpretation of results took into account the inherent uncertainty of the measurement system and data processing, which depends on instrumental performance, calibration, preventive and corrective maintenance, and the magnitude of the gaps present in each series.

Three methods for imputing missing data were evaluated: linear interpolation, polynomial interpolation, and moving average, implemented using the NumPy and Pandas packages. The comparison between methods considered the length of the gaps, the risk of series distortion, and the preservation of the observed temporal structure.

2.2. Preliminary analysis

The analysis of measures of central tendency and stationarity of the time series was performed using box plots by season and year. This graphical representation allowed for the evaluation of the median as a robust measure of central tendency, the dispersion (interquartile range), and the relative variability between seasons and periods. Additionally, by comparing the monthly averages grouped by season and year, it was possible to identify seasonal patterns and evaluate stationarity in the mean, considering as stationary those series whose monthly averages remained relatively constant over time.

To analyze the relationship between dispersion and the central level of the series, the difference between the 80th and 20th percentiles was plotted against the natural logarithm of the median daily concentrations. This scatter plot allows the determination of whether the components of the series combine additively or multiplicatively, depending on the relationship between the level (median) and variability.

2.3. Statistical analysis and decomposition of time series

To statistically compare PM10 concentrations, a summary table was created in Tableau applying various statistical measures to the [concentration] variable, including minimum, maximum, mean, median, standard deviation, 10th percentile, and 90th percentile. Additionally, the 25th (Q1), 50th (Q2), and 75th (Q3) quartiles were calculated using custom calculated fields (ADS, 2024). All calculations were performed disaggregated by monitoring station over a five-year period, allowing for a detailed characterization of the data distribution.

To analyze the behavior of hourly PM10 concentrations, an additive time series decomposition was applied (Esparza, 2004) using the seasonal_decompose function from the statsmodels package, integrating Python code into Tableau via TabPy. This allowed for the decomposition of the series into its three fundamental components: trend, seasonality, and noise. A DataFrame was structured with date, time, concentration, season, and year, transforming the date column to datetime format and establishing it as the time index. Concentrations were converted to numerical values, handling conversion errors, and a period of 24 was defined to represent the daily cycle. Each component was extracted separately and aligned with the original series, facilitating a detailed analysis of long-term variations, recurring hourly patterns, and random fluctuations.

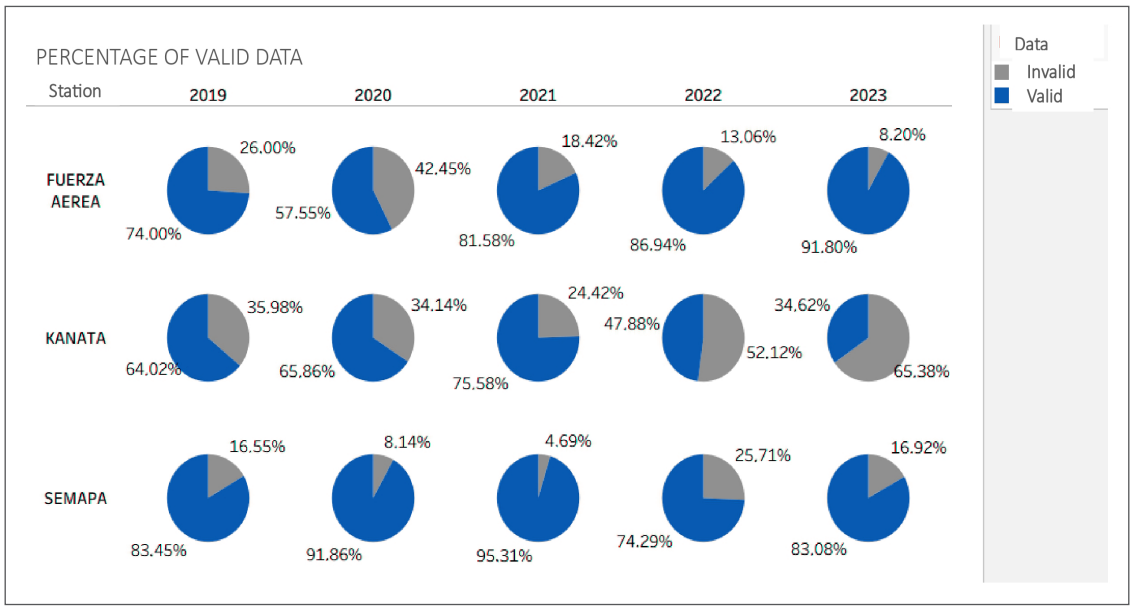
2.4. Development of visualization dashboards

For the exploratory analysis of PM10 concentrations, line graphs, histograms, box plots, and heat maps were used in Tableau®. These representations allowed for the identification of trends, seasonality, dispersion, and periods of high concentrations. In this study, 'critical moments' are defined as hourly or seasonal intervals in which the series shows relatively high concentrations and/or exceedances with respect to reference lines established based on the RMCA (Bolivia, 1995), NB 62011:2018 (IBNORCA, 2018), and WHO guideline values (2021).

3. RESULTS AND DISCUSSION

Once the time series was structured, data cleaning was performed using the percentile method with window functions, selected for its ability to remove outliers without significantly altering the overall behavior of the series. Subsequently, linear interpolation, polynomial interpolation, and moving average methods were compared. The moving average was considered the most suitable procedure for descriptive analysis because it best preserved the visual continuity of the series in short gaps, while linear interpolation was reserved for time decomposition, since this analysis required a complete sequence of observations.

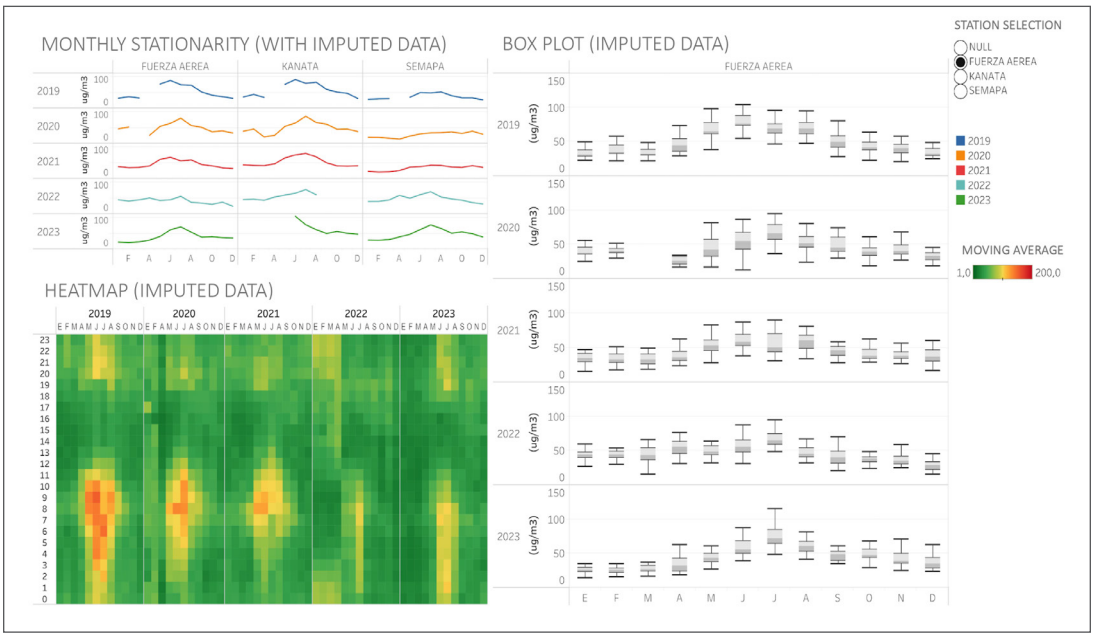
Figure 2. Percentage of valid data after imputation by moving average



Source: Cortez, (2025)

During the period 2019–2023, the three stations showed a marked seasonal pattern: PM10 concentrations tended to be higher between June and September and lower between December and March. This behavior is consistent with the reduced atmospheric dispersion characteristic of the dry winter season in the Cochabamba Valley. Station E1 exhibited the greatest variability in 2019–2020 before stabilizing; Station E2 maintained intermediate dispersion with occasional episodes of high concentration; and Station E3 generally registered the lowest relative variability.

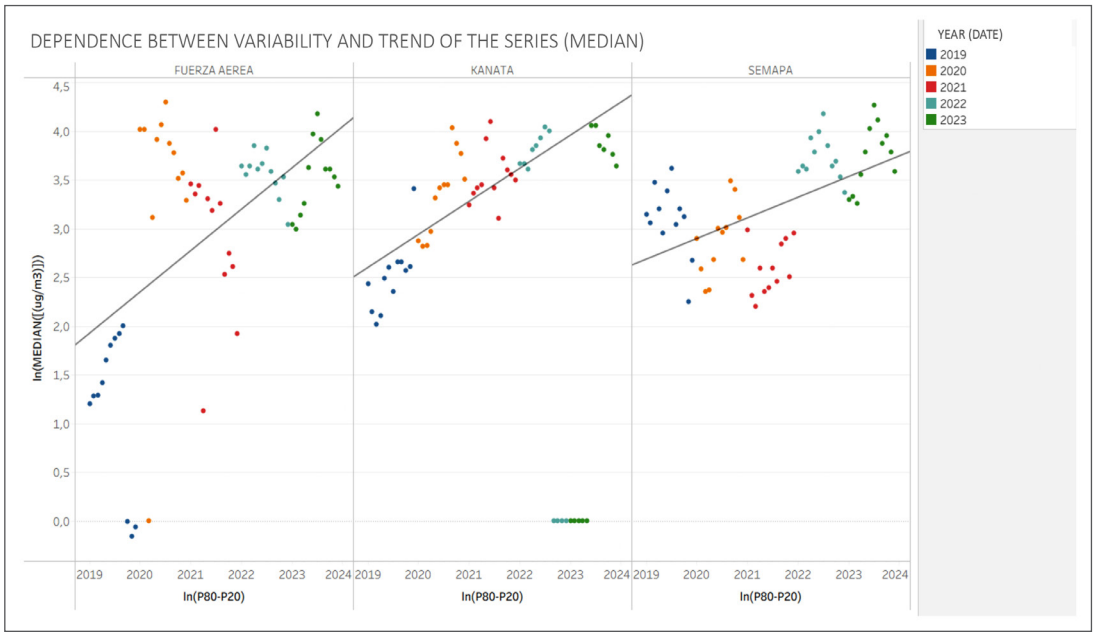
Figure 3. Visualization dashboard seasonal behavior



Source: Cortez, (2025)

When exploring the monthly relationship between the median and its dispersion, it was observed that some isolated years exhibited behaviors consistent with a multiplicative scheme; however, the 2019-2023 period was better suited to an additive model (Alam et al., 2015). Consequently, the classical additive decomposition was considered more appropriate to summarize the overall dynamics of the series.

Figure 4. Dependence of variability and series level (median)

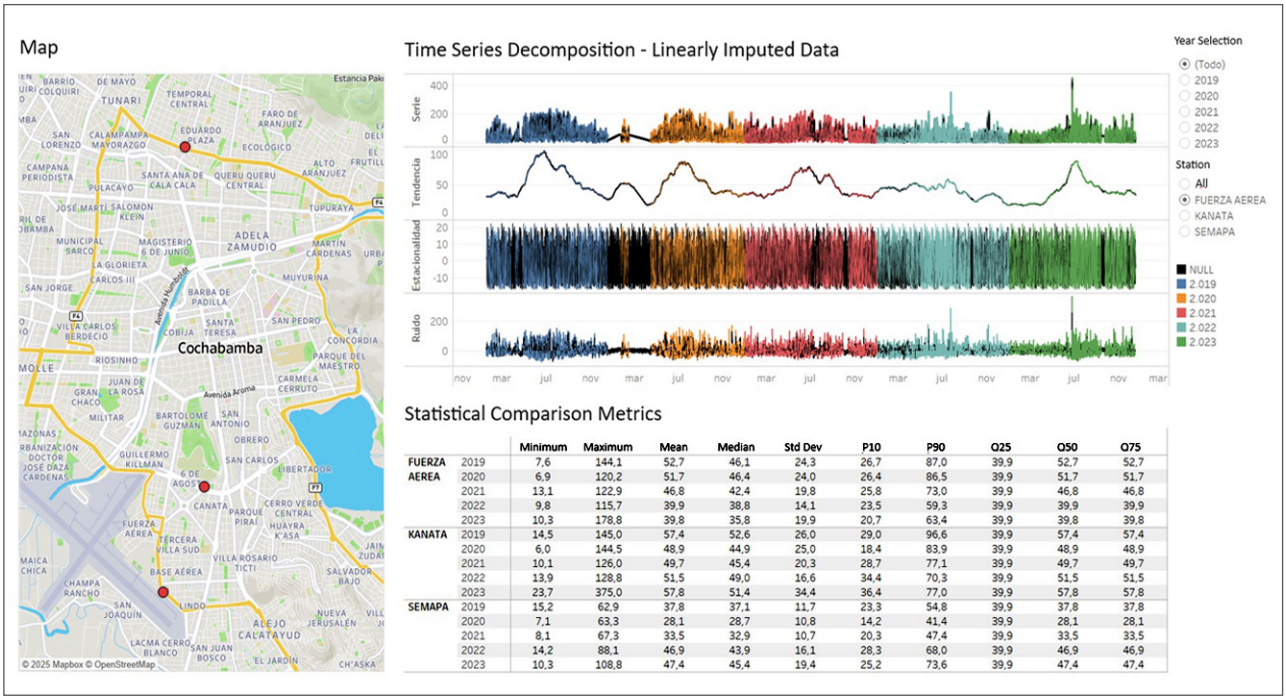


Source: Cortez, (2025)

Using additive decomposition with linear interpolation and a 720-hour smoothing, three main components were isolated:

- Trend, with a decline after 2020 in E1, seasonal fluctuations in E2 and a moderate increase in E3 from 2022;
- Seasonality, with recurring daily and monthly cycles in all three seasons;
- Noise or irregular component, predominantly random, with a greater presence of outliers in some periods of 2022 and 2023.

Figure 5. Time series decomposition and statistical comparison metrics



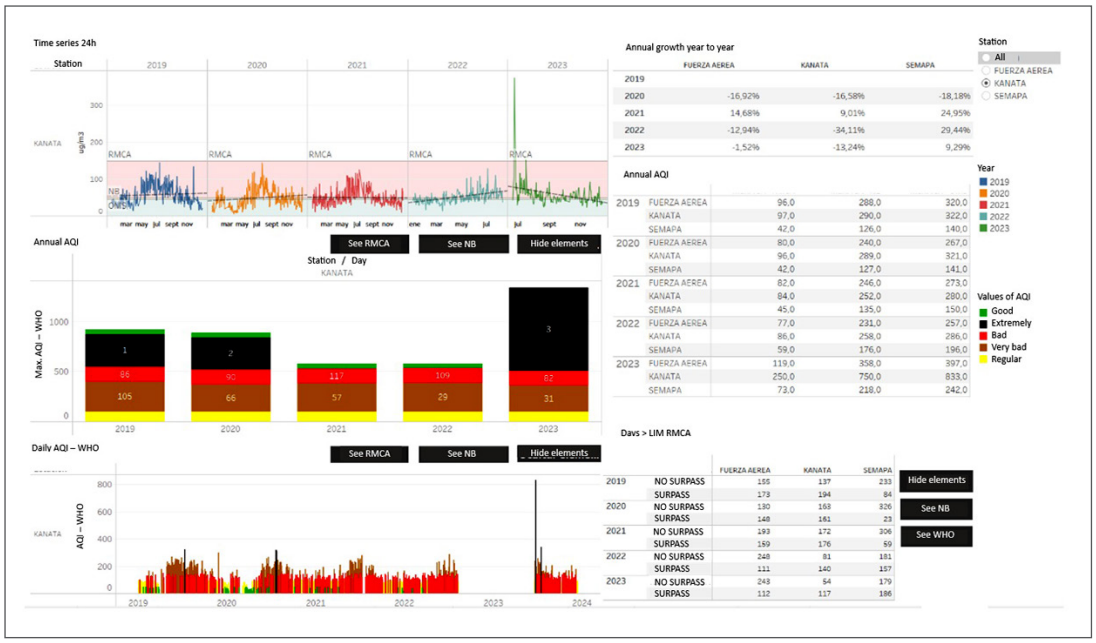
		Minimum	Maximum	Mean	Median	Std Dev	P10	P90	Q25	Q50	Q75
FUERZA AEREA	2019	7,6	144,1	52,7	46,1	24,3	26,7	87,0	39,9	52,7	52,7
	2020	6,9	120,2	51,7	46,4	24,0	26,4	86,5	39,9	51,7	51,7
	2021	13,1	122,9	46,8	42,4	19,8	25,8	73,0	39,9	46,8	46,8
	2022	9,8	115,7	39,9	38,8	14,1	23,5	59,3	39,9	39,9	39,9
	2023	10,3	178,8	39,8	35,8	19,9	20,7	63,4	39,9	39,8	39,8
KANATA	2019	14,5	145,0	57,4	52,6	26,0	29,0	96,6	39,9	57,4	57,4
	2020	6,0	144,5	48,9	44,9	25,0	18,4	83,9	39,9	48,9	48,9
	2021	10,1	126,0	49,7	45,4	20,3	28,7	77,1	39,9	49,7	49,7
	2022	13,9	128,8	51,5	49,0	16,6	34,4	70,3	39,9	51,5	51,5
	2023	23,7	375,0	57,8	51,4	34,4	36,4	77,0	39,9	57,8	57,8
SEMAPA	2019	15,2	62,9	37,8	37,1	11,7	23,3	54,8	39,9	37,8	37,8
	2020	7,1	63,3	28,1	28,7	10,8	14,2	41,4	39,9	28,1	28,1
	2021	8,1	67,3	33,5	32,9	10,7	20,3	47,4	39,9	33,5	33,5
	2022	14,2	88,1	46,9	43,9	16,1	28,3	68,0	39,9	46,9	46,9
	2023	10,3	108,8	47,4	45,4	19,4	25,2	73,6	39,9	47,4	47,4

Source: Cortez, (2025)

The heat maps showed that, without imputation, the temporal interpretation was fragmented by gaps in the data series. Using a moving average, time periods with relatively high concentrations during the winter were more clearly observed, especially between 03:00 and 10:00 in E2. The inconsistencies detected in E1 during 2022 suggest possible recording or operational problems with the monitoring system, rather than a conclusive change in the pollutant's dynamics.

In addition, histograms and bar graphs were generated to quantify the frequency of days within different Air Quality Index (AQI) ranges and to compare the observed concentrations with the limits of the RMCA, NB 62011:2018, and the WHO. This comparison allows the results to be contextualized from a regulatory perspective without extrapolating the findings to causal attributions not evaluated in the study.

Figure 6. Trend and permissible limits



Source: Cortez, (2025)

4. CONCLUSIONS

- A PM10 database was compiled for the period 2019–2023, corresponding to three automatic stations of the MoniCA Network in Cochabamba. The existence of discontinuities and differences in the proportion of valid data between stations confirmed that the sufficiency of information is a central condition for the interpretation of air quality time series.

- The debugging and descriptive analysis routines showed that the percentile method with window functions was useful for handling outliers, while the comparison between linear interpolation, polynomial interpolation, and moving average allowed for the selection of different procedures depending on the analytical objective. The moving average proved more convenient for the descriptive exploration of short gaps, and linear interpolation was functional for decomposition when complete continuity was required.

- Additive decomposition allows the separation of trend, seasonality, and noise components, showing a consistent temporal pattern across the three seasons and supporting the use of an additive approach to summarize the overall dynamics of PM10 in the period studied.

- At all three stations, a higher frequency of elevated PM10 concentrations was identified during the dry and winter periods, as well as spatial differences in the variability of these concentrations. The Parque Kanata station showed the highest values and greatest dispersion, while SEMAPA exhibited comparatively more stable behavior.

- The comparison with the limits of the RMCA, NB 62011:2018, and the WHO showed that the interpretation of risk changes depending on the standard applied. In particular, the WHO reference showed a greater number

of exceedances, highlighting the usefulness of including multiple regulatory references in environmental visualization platforms.

5. RECOMMENDATIONS

- It is recommended to strengthen the monitoring Quality Assurance and Quality Control (QA/QC) program through periodic calibrations, preventive and corrective maintenance, systematic data validation, and explicit documentation of measurement uncertainty. Furthermore, integrating meteorological variables and deepening station characterization is advisable to enhance future interpretations of the observed patterns.

- The integration of Tableau® with Python/ TabPy has proven to be a technically useful alternative for visualizing large time series, calculating descriptive metrics, and dynamically presenting comparative results. Its main contribution lies in facilitating the analytical exploration of available information within the limitations of database quality and continuity.

6. ACKNOWLEDGMENTS

We express our sincere gratitude to the MoniCA Network of Cochabamba, which, through the Cochabamba Municipal Government, provided data from its automatic particulate matter (PM10) monitoring system. Likewise, we acknowledge the valuable collaboration of the MoniCA Network's Air Quality Reference Laboratory for its willingness and support in providing supplementary PM10 monitoring data. The information provided by both institutions was fundamental to the development and robustness of this study.

This article derives from the undergraduate project of the Environmental Engineering Program developed by Mishel Cortez Apaza, “Development of a platform for visualization and analysis of time series of air quality for the municipality of Cochabamba”, of the Faculty of Engineering of UMSA.

7. LIST OF SYMBOLS AND ABBREVIATIONS

PM_{10}	Suspended particles smaller than 10 microns
NO_2	Nitrogen dioxide
CO	Carbon monoxide
O_3	Tropospheric ozone
E1	Fuerza Aérea (Air Force)
E2	Parque Kanata (Kanata Park)
E3	SEMAPA
IQR	Interquartile Range
TabPy	Tableau Python Server – An extension that connects Tableau® with a Python environment
DAX	Data Analysis Expressions – Power BI® 's language of formulas and expressions
RMCA	Air Quality Regulations (Bolivia, DS No. 24176)
NB 62011:2018	Bolivian Standard “Air Quality – Outdoor Atmospheric Pollutants – Maximum Permissible Limits”
WHO	World Health Organization
AQI	Air Quality Index
SNIA	National Environmental Information System (Bolivia)
SDC	Swiss Agency for Development and Cooperation

8. REFERENCES

Acosta, N. (2022). *Informe Anual de la Calidad de Aire - Departamento de Cochabamba*. GADC.

ADS. (26 de noviembre de 2024). *Understanding the Interquartile Range (IQR) for Better Data Analysis*. Around Data Science: arounddatascience.com

Aggarwal, C. (2016). *An Introduction to Outlier Analysis, Outlier Analysis*. Springer, 1-34. doi.org

Alam, M., Rahman, M., Sumy, S., & Parh, Y. (2015). Time Series Decomposition and Seasonal Adjustment. *Global Journal of Science Frontier Research: F Mathematics and Decision Sciences*, XV (9), 10-20. doi.org

Bascopé, D. (2001). La Red de Monitoreo de la Calidad del Aire de Cochabamba (Red MoniCA*). Acta Nova,* I (3), 283-291.

Bolivia (1995). *Reglamento en Materia de Contaminación Atmosférica (DS N° 24176). FAOLEX: c*hrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://faolex.fao.org/docs/pdf/bol179630.pdf

Cortez M., (2025). *Desarrollo de una plataforma de visualización y análisis de series temporales de calidad del aire para el municipio de Cochabamba*. Facultad de Ingeniería-UMSA.

El País. (8 de septiembre de 2021). *El aire contaminado mata a 7 millones de personas al año*. elpais.bo?

Esparza, C. (2004). *Series Temporales*. slideshare.net

IBNORCA. (2018). *Calidad del aire – Contaminantes atmosféricos en exterior - Límites máximos permisibles*. *NB 62011:2018*. Bolivia: IBNORCA.

MMAyA. (2021). Programa Nacional de Gestión de Calidad del Aire - SNIA. Red de Monitoreo de Calidad del Aire: chrome-extension://efaidnbm+nnnibpcajpcglclefindmkaj/mmaya.gob.bo

Röbbel, N. (s.a.). *Los espacios verdes: un recurso indispensable para lograr una salud sostenible en las zonas urbanas*. Crónica ONU: un.org

WHO-OMS. (22 de septiembre de 2021). *Directrices mundiales de la OMS sobre la calidad del aire*. Organización Mundial de la Salud: https://www.who.int/es/news-room/questions-and-answers/item/who-global-air-quality-guidelines

WHO-OMS. (24 de octubre de 2024). *Contaminación del aire ambiente (exterior) y salud*. who.int