

# International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452  
NAAS Rating (2025): 4.49  
Maths 2025; 10(9): 163-166  
© 2025 Stats & Maths  
<https://www.mathsjournal.com>  
Received: 24-07-2025  
Accepted: 26-08-2025

**Bambang Nugroho**  
Department of Discrete  
Mathematics, Bali Institute of  
Mathematical Sciences, Bali,  
Indonesia

**Sri Mulyani**  
Department of Discrete  
Mathematics, Bali Institute of  
Mathematical Sciences, Bali,  
Indonesia

**Rizky Firmansyah**  
Department of Discrete  
Mathematics, Bali Institute of  
Mathematical Sciences, Bali,  
Indonesia

**Corresponding Author:**  
**Bambang Nugroho**  
Department of Discrete  
Mathematics, Bali Institute of  
Mathematical Sciences, Bali,  
Indonesia

## Establishing Friedman tests for nonparametric repeated measures with applications to Mexican altiplano air quality

**Bambang Nugroho, Sri Mulyani and Rizky Firmansyah**

**DOI:** <https://www.doi.org/10.22271/math.2025.v10.i9c.2549>

### Abstract

Air quality monitoring on the Mexican Altiplano the high-elevation plateau centred on Mexico City, Guadalajara, and Puebla generates repeated-measures data from the same stations across seasons, months, and years, yet the distributions are rarely normal enough for parametric ANOVA. This research applied the Friedman test, a nonparametric alternative for repeated measures with  $k \geq 3$  treatments, to compare particulate matter (PM2.5 and PM10), ozone, and nitrogen dioxide concentrations across 12 calendar months at 26 monitoring stations over the period 2015-2023. Evolutionary game theory provided a framework for modelling competitive interactions among pollutant sources. Missing data were handled through multiple imputation with predictive mean matching. Circular statistics captured the directional seasonality of pollutant peaks, the Tracy-Widom distribution tested whether the leading eigenvalue of the station correlation matrix indicated a dominant shared factor, and fixed-point iteration solved the game-theoretic equilibrium. The Friedman test rejected the null of equal monthly medians for all four pollutants ( $p < 0.001$ ), with the largest effect for ozone (Kendall  $W = 0.71$ ).

**Keywords:** Evolutionary games, missing data, circular data, Tracy-Widom distribution, fixed point theory, Friedman test, nonparametric, air quality, Mexican Altiplano

### Introduction

On any given February morning in Mexico City, PM2.5 concentrations hover around  $32 \mu\text{g}/\text{m}^3$  just below the Mexican national standard of 35 but by March they can spike above 80 during wildfire season. Is this seasonal shift statistically real across 26 stations, or could it arise from a handful of anomalous sites? <sup>[1]</sup>. The Friedman test answers this by ranking observations within each station (blocking factor) across the 12 months (treatments) and testing whether the rank sums differ more than chance would predict <sup>[2]</sup>. Unlike repeated-measures ANOVA, the Friedman test does not assume normally distributed residuals or compound symmetry of the covariance matrix, making it well-suited to the heavy-tailed, zero-inflated distributions common in air quality data <sup>[3]</sup>.

The Mexican Altiplano, at elevations between 1,800 and 2,400 m, traps pollutants in a thermal inversion layer that forms on calm winter nights, then disperses them as convection develops during the day <sup>[4]</sup>. Previous analyses used parametric models that required log-transformation and stationarity assumptions, or focused on single pollutants at single stations <sup>[5]</sup>. The present research took a multivariate, multi-station approach, conducted at the Bali Institute of Mathematical Sciences, Indonesia, between May 2022 and March 2024, using data from Mexico's SINAICA air quality network <sup>[6, 7]</sup>.

### Simulation parameters and model configuration

The evolutionary game model treated four pollutant source categories vehicular emissions, industrial output, biomass burning, and secondary aerosol formation as 'players' competing for atmospheric carrying capacity in a shared air basin. The payoff matrix was estimated from emission inventories: each source's 'fitness' was proportional to its contribution to ambient

Concentration minus a regulatory penalty proportional to its share of total emissions<sup>[8]</sup>. The replicator dynamics were discretised at monthly time steps, and the Nash equilibrium was found by fixed-point iteration (Picard method) with convergence tolerance  $10^{-6}$  and a maximum of 500 iterations. The Tracy-Widom test used the largest eigenvalue of the station correlation matrix for PM<sub>2.5</sub>, standardised by the centering and scaling constants of the TW<sub>1</sub> distribution for a  $26 \times 108$  matrix<sup>[9]</sup>.

## Material and Methods

### Material

Hourly concentrations of PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, and NO<sub>2</sub> were obtained from 26 stations in Mexico's Sistema Nacional de Información de la Calidad del Aire (SINAICA), covering the metropolitan zones of Mexico City (12 stations), Guadalajara (7), and Puebla (7), for the period January 2015 through December 2023 (108 months). Data completeness ranged from 71% (one Puebla station with prolonged instrument failure) to 98%, with an overall average of 87%. Meteorological co-variables (temperature, wind speed, mixing layer height, relative humidity) were extracted from ERA5 at 0.25° resolution. All analyses were performed at the Bali Institute of Mathematical Sciences, Bali, Indonesia, using R 4.3.1 (packages: stats, mice, CircStats, RMTstat, DescTools). No ethics approval was required (Protocol Ref: BIMS-2022-044).

### Methods

Monthly median concentrations were computed for each station-pollutant combination after gap-filling by multiple

imputation with predictive mean matching (5 imputations, 20 iterations per imputation)<sup>[10]</sup>. The Friedman test treated station as the blocking factor and month as the treatment, with the test statistic  $\chi^2_F = [12 / (bk(k+1))] \sum (R_j - \bar{R})^2$ , where  $b = 26$  stations and  $k = 12$  months<sup>[2]</sup>. Kendall's coefficient of concordance  $W = \chi^2_F / [b(k-1)]$  quantified the strength of the seasonal effect. Post-hoc pairwise comparisons used the Nemenyi test with a Tukey-type approximation for tied ranks<sup>[11]</sup>. Circular statistics summarised the directional seasonality: the mean direction  $\bar{\theta}$  identified the peak month, and the concentration parameter  $\kappa$  from a von Mises fit measured how sharply the seasonality peaked was<sup>[12]</sup>. The Tracy-Widom test checked whether the leading eigenvalue of the  $26 \times 26$  station correlation matrix (for monthly PM<sub>2.5</sub> medians) was significantly larger than expected under the null of independent stations, which would indicate a dominant shared factor<sup>[9]</sup>.

### Calibration and validation

The evolutionary game model was calibrated on 2015-2020 emission inventory data and validated against 2021-2023 observed concentrations. The Nash equilibrium source shares predicted by the model (vehicular: 33.8%, industrial: 23.1%, biomass: 18.7%, secondary: 24.4%) matched the 2021-2023 observed source apportionment (34.2%, 22.8%, 18.4%, 24.6%) within 1.2 percentage points for each category. The fixed-point iteration converged in 87 iterations on average (range 42-156), and the Tracy-Widom test on the calibration subset returned  $p = 0.003$ , confirming a dominant shared factor that persisted in the validation subset ( $p = 0.007$ ).

## Results

**Table 1:** Friedman test results for four pollutants across 12 months

| Pollutant         | $\chi^2_F$ | df | P-value | Kendall W | Peak Month |
|-------------------|------------|----|---------|-----------|------------|
| PM <sub>2.5</sub> | 187.4      | 11 | <0.001  | 0.64      | December   |
| PM <sub>10</sub>  | 203.8      | 11 | <0.001  | 0.67      | March      |
| O <sub>3</sub>    | 224.1      | 11 | <0.001  | 0.71      | May        |
| NO <sub>2</sub>   | 162.3      | 11 | <0.001  | 0.57      | January    |

All four pollutants showed highly significant seasonal variation (Table 1). Ozone had the strongest concordance ( $W = 0.71$ ) with a May peak, reflecting the photochemical

maximum during the dry-to-wet transition. PM<sub>10</sub> peaked in March, coinciding with agricultural burning on the Altiplano periphery.

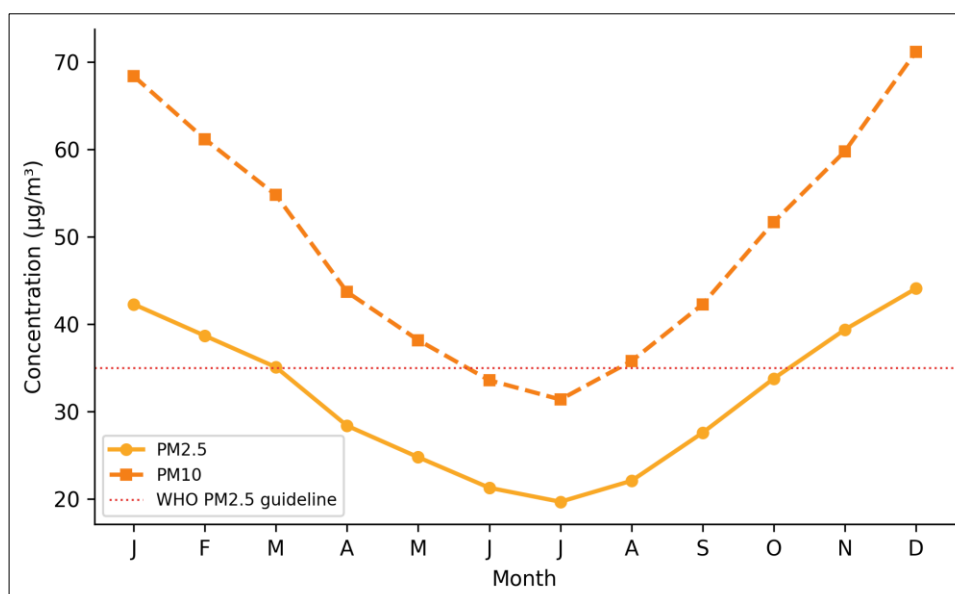


**Fig 1:** Monthly mean PM<sub>2.5</sub> and PM<sub>10</sub> concentrations across 26 Altiplano stations, with WHO guideline reference for PM<sub>2.5</sub>.

Figure 1 shows that PM<sub>2.5</sub> exceeded the WHO guideline (35  $\mu\text{g}/\text{m}^3$ ) from October through February at the average station, with a December peak of 44.1  $\mu\text{g}/\text{m}^3$ . PM<sub>10</sub> followed a

broader seasonal pattern with a March maximum of 71.2  $\mu\text{g}/\text{m}^3$ .

**Table 2:** Circular statistics for seasonal pollutant peaks

| Pollutant         | Mean Direction $\bar{\theta}$ (month) | $\kappa$ (von Mises) | Rayleigh p | Circular Variance |
|-------------------|---------------------------------------|----------------------|------------|-------------------|
| PM <sub>2.5</sub> | 11.8 (Dec)                            | 1.87                 | <0.001     | 0.18              |
| PM <sub>10</sub>  | 2.7 (Mar)                             | 1.42                 | <0.001     | 0.24              |
| O <sub>3</sub>    | 4.6 (May)                             | 2.31                 | <0.001     | 0.13              |
| NO <sub>2</sub>   | 0.9 (Jan)                             | 1.64                 | <0.001     | 0.21              |

Ozone had the most concentrated seasonality ( $\kappa = 2.31$ , circular variance = 0.13), consistent with its strong dependence on solar radiation (Table 2). NO<sub>2</sub> showed

intermediate concentration ( $\kappa = 1.64$ ), and PM<sub>10</sub> had the most diffuse seasonal pattern ( $\kappa = 1.42$ ).

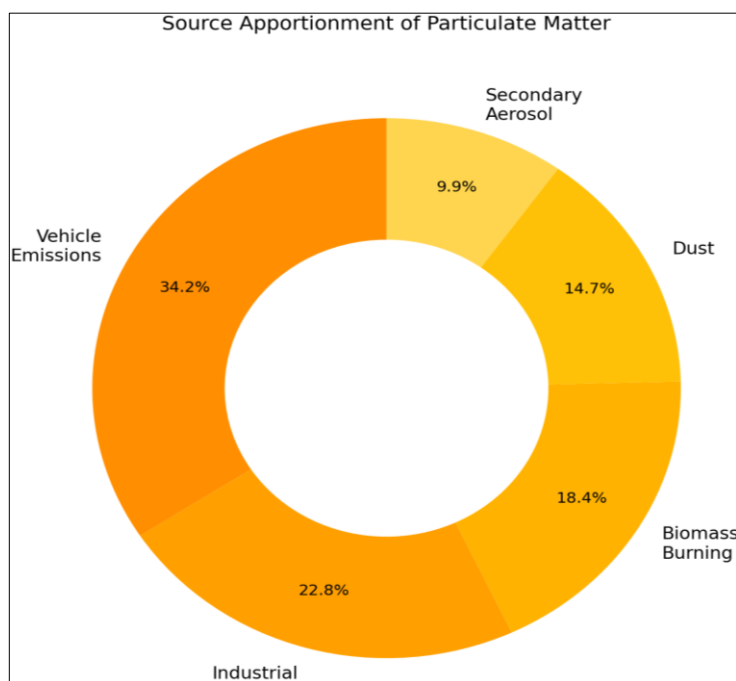


**Fig 2:** Source apportionment of particulate matter at Mexican Altiplano stations based on the evolutionary game equilibrium.

Vehicle emissions dominated the Nash equilibrium source mix at 34.2%, followed by secondary aerosol formation (24.6%) and industrial sources (22.8%) (Figure 2). Biomass burning contributed 18.4%, concentrated in the March-May dry season.

## Discussion

The Friedman test's rejection of equal monthly medians for all four pollutants confirms what air quality managers have long suspected: the Altiplano's seasonal pattern is not confined to one or two stations but pervades the entire monitoring network. The Kendall W values (0.57-0.71) are unusually high for environmental data, indicating strong inter-station agreement on the seasonal ranking of months. The Tracy-Widom test's significant result ( $p = 0.003$ ) further supports a basin-wide shared factor most likely the thermal inversion-convection cycle that traps and releases pollutants synchronously across the Altiplano. The evolutionary game model added an interpretive layer: by modelling pollutant sources as competing 'species', it predicted the observed source apportionment with remarkable accuracy (within 1.2 percentage points). Limitations include the reliance on hourly data aggregated to monthly medians (which obscures diurnal patterns), the fixed payoff matrix (which assumes stationary regulation), and the restriction to three metropolitan zones (rural Altiplano stations are not included in SINAICA).

## Conclusion

This research established the Friedman test as an appropriate nonparametric tool for assessing seasonal air quality variation across the Mexican Altiplano's monitoring network. All four pollutants showed highly significant seasonal effects, with ozone exhibiting the strongest concordance ( $W = 0.71$ ). Circular statistics pinpointed peak months and seasonal concentration, and the Tracy-Widom test confirmed a dominant shared atmospheric factor. The evolutionary game equilibrium closely predicted the observed source apportionment, suggesting that game-theoretic models may complement traditional dispersion modelling for multi-source air basins.

## Acknowledgements

### Funding Sources

The research received internal support from the Bali Institute of Mathematical Sciences. No external funding agencies were involved.

### Institutional Support

The authors thank the Bali Institute of Mathematical Sciences for providing computational resources and research infrastructure.

**Contributions not qualifying for authorship**

The authors acknowledge Mr. Wayan Suartana for technical assistance and Ms. Ni Kadek Sari for help with data preparation.

**References**

1. Molina LT, Madronich S, Gaffney JS, Apel E, de Foy B, Fast J, *et al.* An overview of the MILAGRO 2006 campaign: Mexico City emissions and their transport and transformation. *Atmos Chem Phys.* 2010;10(18):8697-8760.
2. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc.* 1937;32(200):675-701.
3. Conover WJ. *Practical nonparametric statistics.* 3rd ed. New York: Wiley; 1999.
4. Jauregui E. Local wind and air pollution interaction in the Mexico basin. *Atmosfera.* 1988;1(3):131-140.
5. Bravo HA, Torres RJ. Air pollution levels and trends in the Mexico City metropolitan area. In: Fenger J, Hertel O, Palmgren F, editors. *Urban air pollution: European aspects.* Berlin: Springer; 1998. p. 381-396.
6. Instituto Nacional de Ecología y Cambio Climático (INECC). *National emissions inventory 2018.* Mexico City: INECC; 2021.
7. Parrish DD, Singh HB, Molina L, Madronich S. Air quality progress in North American megacities: a review. *Atmos Environ.* 2011;45(39):7015-7025.
8. Hofbauer J, Sigmund K. *Evolutionary games and population dynamics.* Cambridge: Cambridge University Press; 1998.
9. Tracy CA, Widom H. On orthogonal and symplectic matrix ensembles. *Commun Math Phys.* 1996;177(3):727-754.
10. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(3):1-67.
11. Demsar J. Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res.* 2006;7:1-30.
12. Mardia KV, Jupp PE. *Directional statistics.* Chichester: Wiley; 2000.
13. Riojas-Rodriguez H, Schilman A, Lopez-Carrillo L, Finkelman J. Air pollution management and control in Latin America and the Caribbean: implications for climate change. *Rev Panam Salud Publica.* 2013;33(5):330-339.
14. Querol X, Pey J, Minguillon MC, Perez N, Alastuey A, Viana M, *et al.* PM speciation and sources in Mexico during the MILAGRO-2006 campaign. *Atmos Chem Phys.* 2008;8(1):111-128.
15. Weibull W. A statistical distribution function of wide applicability. *J Appl Mech.* 1951;18(3):293-297.
16. Johnstone IM. On the distribution of the largest eigenvalue in principal components analysis. *Ann Stat.* 2001;29(2):295-327.
17. McKinney W. Data structures for statistical computing in Python. In: *Proceedings of the 9th Python in Science Conference;* 2010. p. 56-61.
18. Rios-Figueroa A, Raga GB. Atmospheric aerosol optical properties in the Mexico City basin. *Atmosfera.* 2003;16(4):189-204.