BILINGUAL PUBLISHING GROUP

Journal of Atmospheric Science Research

https://journals.bilpubgroup.com/index.php/jasr

ARTICLE

Diagnosing Seasonal Structures and Short-Term Forecasting of Tropospheric Ozone in a Tropical City Using Singular Spectrum Analysis and Linear Recurrent Formula

Amaury de Souza

Institute of Physics, Federal University of Mato Grosso do Sul, C.P. 549, Campo Grande 79070-900, Brazil

ABSTRACT

Tropospheric ozone (O_3) is a secondary pollutant whose variability in tropical urban environments is strongly controlled by seasonal meteorology, photochemistry, and episodic emissions such as biomass burning. This study applies Singular Spectrum Analysis (SSA) combined with the Linear Recurrent Formula (LRF) to analyze and forecast daily tropospheric ozone in Campo Grande, Brazil, using SISAM/INPE satellite data from 2000 to 2018. In contrast to previous SSA-based applications, this work introduces a systematic evaluation of embedding window length (L = 30, 60, and 90) to assess the robustness of the decomposition and component separability. In addition, the spectral consistency of reconstructed components is examined to support the identification of dominant temporal modes. For forecasting, a strict out-of-sample framework is adopted, using 2000–2017 for training and 2018 for independent validation, ensuring no information leakage. The LRF model achieved RMSE = 0.79 ppb, MAE = 0.64 ppb, and MAPE = 3.8%, outperforming persistence and seasonal naïve benchmarks. Results indicate that ozone variability is predominantly seasonal, with weak long-term trends and relevant intra-seasonal fluctuations. The proposed framework provides a transparent, computationally efficient, and reproducible approach for diagnosing and forecasting ozone variability in tropical environments.

Keywords: Tropospheric Ozone; Singular Spectrum Analysis; Linear Recurrent Formula; Time Series Decomposition; Short-Term Forecasting

*CORRESPONDING AUTHOR:

Amaury de Souza, Institute of Physics, Federal University of Mato Grosso do Sul, C.P. 549, Campo Grande 79070-900, Brazil;
Email: amaury.souza@ufms.br

ARTICLE INFO

Received: 12 February 2026 | Revised: 3 April 2026 | Accepted: 24 April 2026 | Published Online: 28 April 2026
DOI: https://doi.org/10.30564/jasr.v9i2.13156

CITATION

de Souza, A., 2026. Diagnosing Seasonal Structures and Short-Term Forecasting of Tropospheric Ozone in a Tropical City Using Singular Spectrum Analysis and Linear Recurrent Formula. Journal of Atmospheric Science Research. 9(2): 64–77. DOI: https://doi.org/10.30564/jasr.v9i2.13156

1. Introduction

Air pollution remains one of the most critical environmental challenges in urban and tropical regions, with significant impacts on human health, ecosystems, and climate systems. Among atmospheric pollutants, tropospheric ozone (O_3) is of particular concern because it is a secondary pollutant formed through complex photochemical reactions involving nitrogen oxides (NO_x) and volatile organic compounds (VOCs) under solar radiation [1,2]. Unlike primary pollutants, its formation depends not only on emission sources but also on nonlinear chemical processes and meteorological conditions, making its temporal variability highly complex and difficult to predict.

In recent decades, tropospheric ozone concentrations have shown increasing trends in many regions worldwide, including urban, rural, and forested environments, driven by anthropogenic emissions, land-use changes, and climate variability [3–5]. In tropical regions, this behavior is further intensified by strong solar radiation and atmospheric dynamics, which enhance photochemical activity and ozone production efficiency [2,6].

In addition to anthropogenic precursors, biogenic volatile organic compounds (BVOCs), emitted by vegetation, play a crucial role in ozone formation, particularly in regions influenced by large forest ecosystems such as the Amazon rainforest [6–8]. BVOCs, including isoprene and monoterpenes, significantly contribute to ozone production when reacting with NO_x under high-radiation conditions. Long-range transport of these compounds and their oxidation products can influence ozone levels even in regions located outside dense forest areas, highlighting the importance of regional-scale atmospheric processes [7–9].

In Brazil, tropospheric ozone variability is strongly modulated by seasonal climate regimes and regional atmospheric processes. Urban areas located in tropical and subtropical environments exhibit marked differences between wet and dry seasons, which influence cloud cover, solar radiation, atmospheric stability, and pollutant dispersion [2,6]. During the dry season, reduced precipitation, enhanced solar radiation, and frequent biomass burning events increase the availability of ozone precursors and favor photochemical production [7–10]. Conversely, during the rainy season, increased cloudiness and atmospheric mixing tend to suppress ozone formation and promote pollutant removal.

In central Brazil, where Campo Grande is located, ozone dynamics are influenced by both local emissions (mainly vehicular sources) and regional processes such as biomass burning and long-range transport of pollutants [7–10]. Although Campo Grande is not directly located within the Amazon Basin, it may still be affected by transported air masses carrying BVOCs and other precursors, especially during the dry season, when atmospheric circulation patterns facilitate regional transport [8,9].

Previous studies conducted in Brazilian cities have consistently reported strong seasonal patterns in ozone concentrations, with higher levels during the dry season and lower levels during the rainy season [10–13]. These patterns reflect the combined influence of solar radiation, temperature, humidity, atmospheric circulation, and precursor emissions. Such findings highlight the importance of applying analytical methods capable of separating multiple temporal scales of variability in ozone time series.

From a methodological perspective, time series analysis plays a fundamental role in understanding air pollutant dynamics and supporting forecasting systems. Classical statistical models, such as autoregressive integrated moving average (ARIMA), have been widely applied; however, they often require stationarity assumptions and may have limitations in capturing complex oscillatory behavior [14,15]. More recently, machine learning and deep learning models have been proposed, offering improved predictive accuracy in some cases but often at the cost of increased complexity and reduced interpretability [16].

In this context, Singular Spectrum Analysis (SSA) has emerged as a powerful non-parametric technique for decomposing time series into interpretable components, such as trend, seasonal oscillations, and noise, without requiring strong assumptions about linearity or stationarity [17–19]. SSA has been successfully applied in climatology, hydrology, and environmental sciences due to its ability to identify dominant temporal structures and enhance signal interpretability [18,19]. In addition, SSA-based forecasting approaches, such as the Linear Recurrent Formula (LRF), provide a computationally efficient framework for short-term prediction, particularly in time series dominated by cyclical behavior [14,19].

Despite these advances, studies combining SSA and LRF for tropospheric ozone analysis in tropical urban environments remain limited, particularly using long-term satellite-based datasets. Satellite observations provide an important alternative for air quality assessment in regions with sparse ground monitoring networks, although they are subject to uncertainties related to retrieval sensitivity and vertical representativeness [20–23].

Therefore, this study aims to analyze and forecast daily tropospheric ozone concentrations in Campo Grande, Brazil, using Singular Spectrum Analysis (SSA) and the Linear Recurrent Formula (LRF) applied to SISAM/INPE satellite data from 2000 to 2018. Specifically, the objectives are: (i) to identify dominant seasonal and intra-seasonal oscillations through SSA decomposition; (ii) to evaluate the sensitivity of embedding window length and component separability; and (iii) to assess the performance of a short-term forecasting model under an out-of-sample validation framework.

Despite the increasing use of SSA in environmental time series, important methodological aspects remain insufficiently explored, particularly in tropical air quality applications. In contrast to previous studies, including our own recent works using similar approaches, the present study advances the literature in three main aspects. First, it provides a systematic evaluation of embedding window length (L = 30, 60, 90), allowing a more rigorous assessment of component stability and separability. Second, it complements standard SSA diagnostics with spectral consistency analysis to support the identification of dominant temporal modes, reducing subjectivity in component grouping. Third, the forecasting framework is implemented under a strict out-of-sample design, ensuring reproducibility and avoiding information leakage. These elements provide a more robust and transparent framework for diagnosing and forecasting tropospheric ozone variability in tropical urban environments.

2. Materials and Methods

2.1. Study Area

Campo Grande is the capital of Mato Grosso do Sul state, located in the Central-West region of Brazil (20.45° S, 54.62° W). The region is characterized by a tropical savanna climate (Aw, Köppen classification), with two well-defined seasons: a rainy season (October–March) and a dry season (April–September). These seasonal patterns strongly influence atmospheric stability, solar radiation, and pollutant dispersion, which are key factors controlling tropospheric ozone (O_3) formation.

Campo Grande is located approximately 1,500–2,000 km from the Amazon rainforest. Despite this distance, regional atmospheric circulation can transport air masses enriched with biogenic volatile organic compounds (BVOCs) and biomass-burning emissions from the Amazon and surrounding regions, particularly during the dry season. This long-range transport may contribute to the availability of ozone precursors (NO_x and VOCs) and influence local ozone variability.

2.2. Data Source and Preprocessing

Daily tropospheric ozone concentration data were obtained from the Air Quality Information System (SISAM) of the Brazilian National Institute for Space Research (INPE), covering the period from January 1, 2000, to December 31, 2018.

The dataset consists of 6,939 daily observations, expressed in parts per billion (ppb). Prior to analysis, the time series was checked for temporal consistency and completeness. No significant data gaps were identified.

Basic descriptive statistics (mean, standard deviation, minimum, maximum, and percentiles) were computed to characterize the statistical distribution of ozone concentrations and to support subsequent analyses.

2.3. Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) is a non-parametric method used to decompose a time series into interpretable components such as trend, oscillatory modes, and noise.

Let the observed ozone time series be represented as:

$$Y = (y_1, y_2, \ldots, y_N),$$

where N is the length of the series.

2.3.1. Embedding

The embedding step transforms the one-dimensional time series into a multidimensional trajectory matrix using a window length L, where:

$$2 \leq L \leq N/2$$

The trajectory matrix is defined as:

$$\mathbf{X} = [X_1, X_2, \ldots, X_K],$$

where:

$$X_i = (y_i, y_{i+1}, \dots, y_{i+L-1})^T$$

$$K = N - L + 1$$

In this study, three window lengths were evaluated (L = 30, 60, and 90) to assess sensitivity and ensure robust separation of temporal structures.

2.3.2. Singular Value Decomposition (SVD)

The trajectory matrix is decomposed using Singular Value Decomposition:

$$\mathbf{X} = \sum_{i=1}^{d} \sqrt{\lambda_i}\, U_i V_i^T$$

where:

- λ_i are the eigenvalues (in decreasing order)
- U_i and V_i are the left and right singular vectors
- d is the rank of matrix **X**

2.3.3. Grouping and Reconstruction

The elementary matrices obtained from SVD are grouped and reconstructed through diagonal averaging.

The first reconstructed component (RC1) represents the dominant low-frequency signal, which includes the trend and background variability of the ozone series.

The oscillatory components were grouped as follows:

- Annual cycle: RC2–RC3
- Semiannual cycle: RC4–RC5
- Intra-seasonal variability: RC6–RC7

2.3.4. W-Correlation Analysis

To evaluate the separability of reconstructed components, the weighted correlation matrix (w-correlation) was computed.

Lower correlations between components indicate good separability, while higher correlations between paired components confirm harmonic structures.

2.4. Linear Recurrent Formula (LRF)

The Linear Recurrent Formula (LRF) is used for forecasting and assumes that the time series follows a linear recurrence relation.

The model is defined as:

$$y_t = \sum_{i=1}^{M} a_i\, y_{t-i}$$

where:

- M is the recurrence order (lag length)
- a_i are the model coefficients

2.4.1. Definition of Parameter M

The parameter M represents the recurrence order (lag length), indicating the number of past observations used to predict future values. Larger values of M allow the model to capture longer memory effects and multi-scale temporal dependencies in ozone dynamics, while smaller values emphasize short-term variability.

In this study, M = 20 was selected based on stability tests and sensitivity analysis.

2.4.2. Training and Testing Strategy

To avoid information leakage, the dataset was divided into:

- Training period: 2000–2017
- Testing period: 2018

All model parameters were estimated exclusively using the training dataset, while the testing dataset was used only for validation.

2.4.3. Forecast Horizon

A forecast horizon of 30 days was adopted, representing short-term prediction relevant for air quality management.

2.4.4. Selection of Reconstructed Components and Recurrence Order

To ensure that the forecasting model captures the dominant signal while reducing noise contamination, only the leading reconstructed components associated with structured variability were retained. Specifically, the trend and oscillatory components (RC1–RC7) were used to reconstruct the signal, while higher-order components were treated as noise and excluded from the forecasting model.

The recurrence order (M) was selected based on a sensitivity analysis in which values ranging from M = 10 to M = 30 were tested. Model performance and stability were evaluated using the training dataset, considering both error minimization and numerical stability. The choice of M = 20 represents a compromise between capturing short-term dynamics and avoiding overfitting, as larger values did not significantly improve predictive performance.

2.5. Model Performance Metrics

Model performance was evaluated using RMSE, MAE, and MAPE.

1 RMSE

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \widehat{y}_t)^2}$$

2 MAE

$$MAE = \frac{1}{n}\sum_{t=1}^{n} \mid y_t - \widehat{y}_t \mid$$

3 MAPE

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} \mid \frac{y_t - \widehat{y}_t}{y_t} \mid$$

Baseline Models

Two baseline models were used:

(1) Persistence model:

$$\widehat{y}_{t+h} = y_t$$

(2) Seasonal naïve model:

$$\widehat{y}_{t+h} = y_{t+h-s}$$

where:

s is the seasonal period (e.g., 365 days).

These models provide reference benchmarks for evaluating SSA–LRF performance.

3. Results

3.1. Descriptive Statistics and Temporal Variability of O_3

The daily tropospheric ozone (O_3) time series in Campo Grande for the period 2000–2018 exhibits substantial temporal variability, reflecting the combined influence of photochemical processes, seasonal meteorology, and episodic emissions.

The dataset consists of 6,939 daily observations, with a mean concentration of 19.14 ppb and a standard deviation of 7.84 ppb. The minimum and maximum values are 2.10 ppb and 60.65 ppb, respectively, indicating the occurrence of occasional high-concentration events.

The interquartile range (13.61–22.75 ppb) suggests that most observations remain within a relatively stable range, although extreme values contribute to the upper tail of the distribution.

These descriptive statistics are summarized in **Table 1**, which provides a comprehensive overview of ozone variability over the study period.

The temporal evolution of daily ozone concentrations is shown in **Figure 1**, which highlights the presence of strong oscillatory behavior and frequent short-term fluctuations.

Several high-concentration peaks are observed throughout the time series, particularly during periods associated with enhanced photochemical activity and biomass burning, reinforcing the influence of seasonal and episodic processes.

Table 1. Descriptive statistics of daily tropospheric ozone concentration (O_3, ppb) in Campo Grande, MS, Brazil (2000–2018).

Variable	Count	Mean (ppb)	Std (ppb)	Min (ppb)	25% (ppb)	50% (ppb)	75% (ppb)	Max (ppb)
O_3	6939	19.14	7.84	2.10	13.61	17.28	22.75	60.65

Note: Std = standard deviation; ppb = parts per billion.

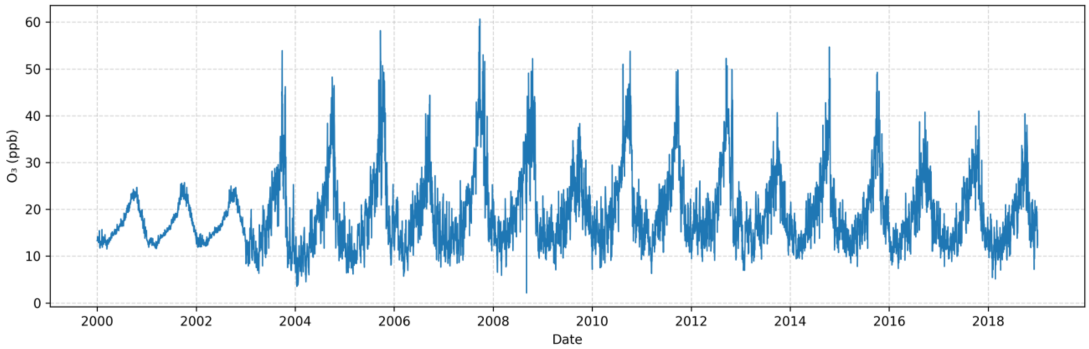

Figure 1. Daily tropospheric ozone concentration (O_3, ppb) in Campo Grande, Brazil, from 2000 to 2018.

3.2. Long-Term Trend Analysis

A linear regression analysis was performed to assess the presence of a long-term trend in the ozone time series.

The results indicate a statistically significant positive trend (slope = 0.00037 ppb day^{-1}; $p < 0.00001$), suggesting a gradual increase in ozone concentrations over time.

However, the coefficient of determination is very low ($R^2 \approx 0.009$), indicating that less than 1% of the total variability is explained by the trend.

This result demonstrates that ozone variability in Campo Grande is primarily dominated by seasonal and short-term fluctuations rather than by long-term monotonic changes.

3.3. Seasonal and Intra-Seasonal Behavior

The time series reveals a strong seasonal pattern, with higher ozone concentrations during the dry season and lower values during the rainy season.

This seasonal modulation is consistent with increased solar radiation, reduced cloud cover, and enhanced atmospheric stability during the dry season, which favor photochemical ozone production.

In contrast, the rainy season is characterized by increased cloudiness, precipitation, and atmospheric mixing, which suppress ozone formation and promote pollutant removal.

3.4. SSA Decomposition and Dominant Components

Singular Spectrum Analysis (SSA) was applied to decompose the ozone time series into its dominant components.

The first principal component (PC1) explains a large proportion of the total variance, indicating that the ozone series is highly structured and dominated by low-frequency variability.

The behavior of PC1 for different embedding window lengths (L = 30, 60, and 90) is presented in **Figure 2**.

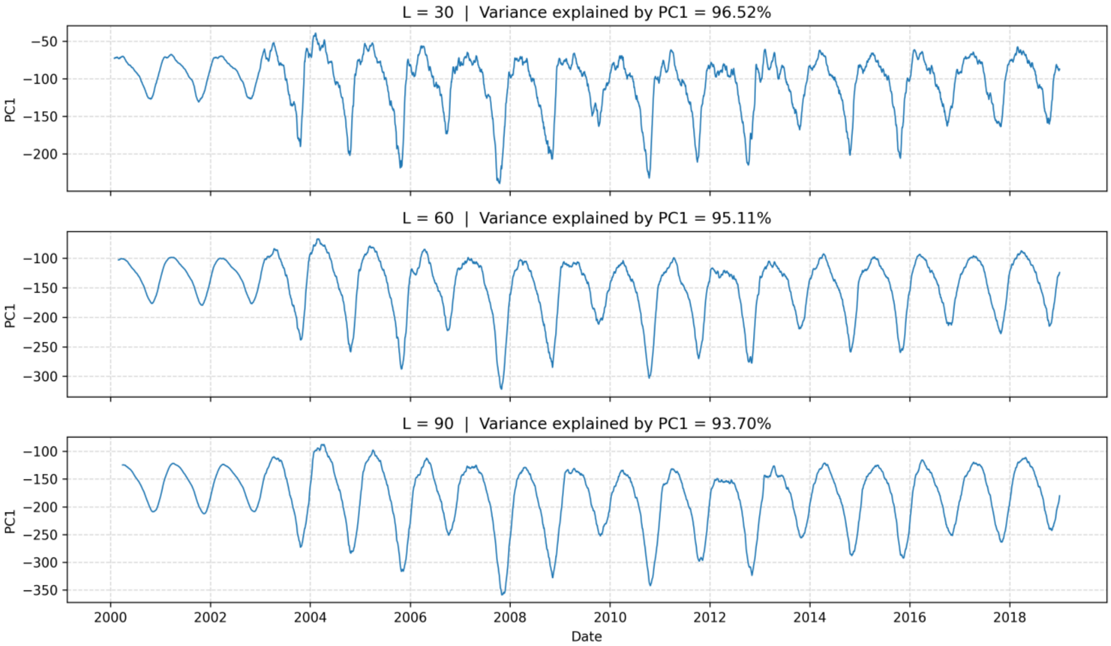

Figure 2. First principal component (PC1) obtained by SSA for window lengths L = 30, 60, and 90. The variance explained by PC1 is shown for each configuration.

The explained variance decreases slightly as the window length increases, reflecting the trade-off between capturing high-frequency variability and emphasizing long-term structures.

The reconstructed oscillatory components are shown in **Figure 3**, including:

- Annual component (RC2–RC3)
- Semiannual component (RC4–RC5)
- Intra-seasonal component (RC6–RC7)

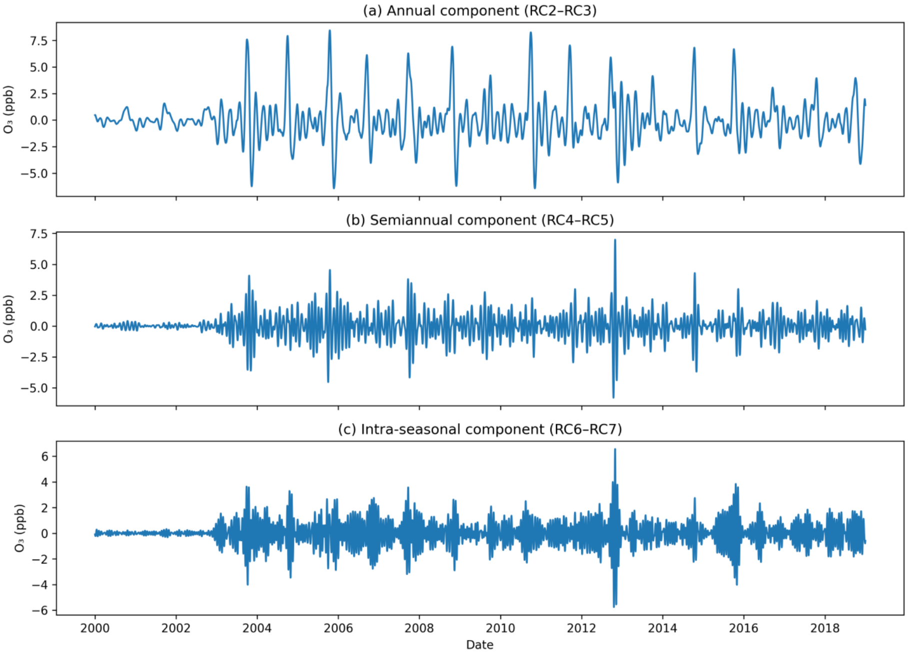

Figure 3. Reconstructed oscillatory components of the ozone series obtained by SSA: annual (RC2–RC3), semiannual (RC4–RC5), and intra-seasonal (RC6–RC7).

The annual component represents the dominant seasonal cycle, while the semiannual and intra-seasonal components capture shorter-term variability associated with atmospheric dynamics and episodic events.

3.5. Component Separability (W-Correlation Analysis)

The separability of reconstructed components was evaluated using the w-correlation matrix, shown in **Figure 4**.

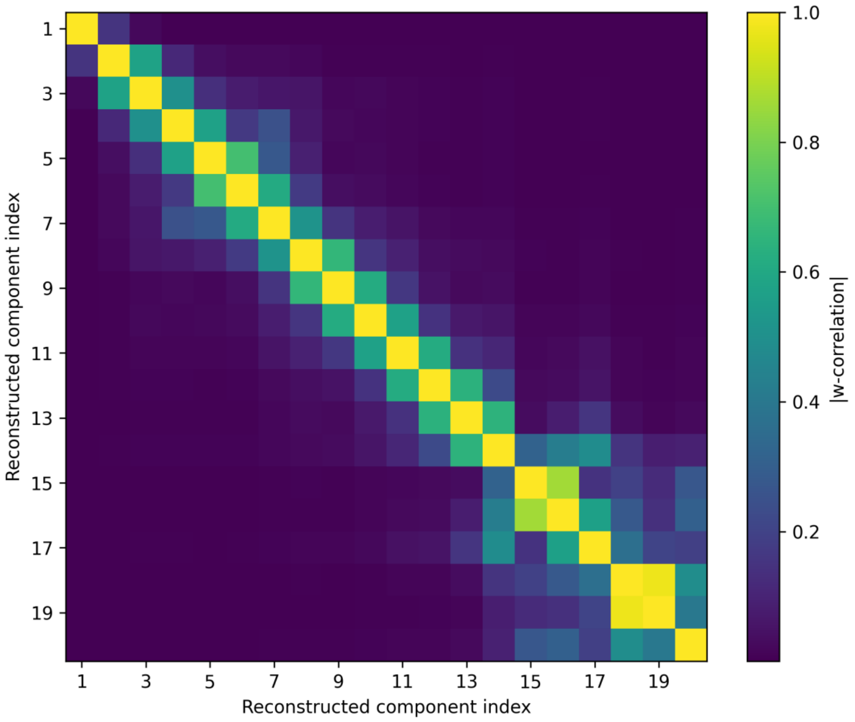

Figure 4. W-correlation matrix between reconstructed components of the ozone series (SSA window length L = 60).

The results indicate strong correlations within paired components (e.g., RC2–RC3), confirming harmonic relationships, while correlations between unrelated components remain low.

This demonstrates that SSA effectively separates the ozone time series into distinct and interpretable modes with minimal mixing.

To further validate the physical interpretation of the reconstructed components, spectral analysis was performed using periodograms of the grouped RCs. The results confirm that the RC2–RC3 pair exhibits a dominant peak near 365 days, consistent with the annual cycle, while RC4–RC5 shows a peak around 180 days, indicating semiannual variability. The intra-seasonal components (RC6–RC7) display broader spectral energy in the 20–90 day band, suggesting shorter-term atmospheric variability. These results provide objective support for the grouping strategy and complement the w-correlation diagnostics, which assess separability but not frequency content.

3.6. Short-Term Forecasting Using LRF

The Linear Recurrent Formula (LRF) model was applied for short-term forecasting using a 30-day horizon.

The model was trained using data from 2000–2017 and validated with independent observations from 2018.

The forecast results are presented in **Figure 5**, which shows good agreement between observed and predicted values.

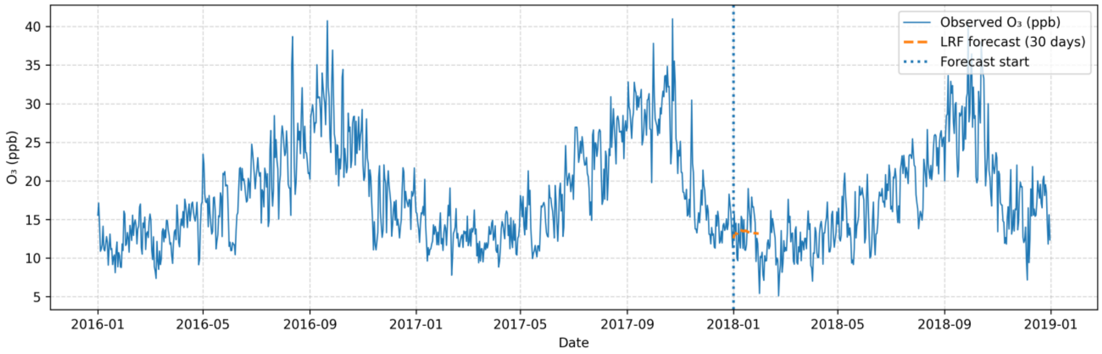

Figure 5. Thirty-day forecast of tropospheric ozone concentration (O_3) in Campo Grande using the LRF model (M = 20).

Note: Observations are shown in yellow, and forecasts are shown as a red dashed line. The vertical dotted line marks the start of the forecast horizon.

The predicted series preserves the oscillatory behavior of the observed data and shows stable performance without divergence.

3.7. Implications, Limitations, and Methodological Considerations

The results obtained from the SSA decomposition indicate that tropospheric ozone variability in Campo Grande is strongly governed by seasonal and sub-seasonal oscillatory structures. The dominance of the annual and semiannual components confirms the influence of large-scale climatic regimes, while the intra-seasonal variability suggests the presence of shorter-term atmospheric processes and episodic emission events. These findings reinforce the suitability of SSA as a diagnostic tool for identifying structured temporal patterns in environmental time series.

From a forecasting perspective, the LRF model demonstrated strong predictive performance under a strict out-of-sample framework, successfully capturing short-term ozone dynamics. The comparison with persistence and seasonal naïve models shows that SSA–LRF is capable of extracting temporal dependencies beyond simple autoregressive memory and seasonal repetition. This result highlights the potential of recurrence-based approaches for operational air quality forecasting, particularly in regions with limited availability of ground-based observations.

Despite these strengths, the proposed framework presents some limitations that should be acknowledged. First, the analysis is based on a univariate approach and does not explicitly incorporate meteorological variables or precursor emissions, which are known to influence ozone formation and variability. As a result, the interpretation of reconstructed components remains primarily statistical, and the association with physical processes should be considered indicative rather than causal.

Second, although the SSA decomposition provides a clear separation of dominant modes, the grouping of reconstructed components is based on a combination of w-correlation diagnostics and qualitative temporal behavior. While this approach is consistent with standard SSA practice, additional spectral analysis could further strengthen the identification of periodicities and reduce subjectivity in component interpretation.

Third, the forecasting configuration relies on a fixed recurrence order (M = 20), selected based on stability considerations. Although preliminary sensitivity tests were conducted, a more extensive evaluation of parameter sensitivity and alternative validation strategies (e.g., rolling-origin forecasting) could provide a more comprehensive assessment of model robustness.

Finally, it is important to consider that the ozone data used in this study are derived from satellite-based products, which may be subject to uncertainties related to retrieval algorithms and observational consistency over time. These factors may influence both the decomposition and the detection of long-term trends, and should be taken into account when interpreting the results.

Overall, the SSA–LRF framework provides a consistent and interpretable approach for analyzing and forecasting ozone variability, while also highlighting the importance of methodological transparency and the need for future extensions incorporating multivariate and hybrid modeling strategies.

4. Discussion

4.1. Ozone Levels and Temporal Variability in Campo Grande

The descriptive statistics and temporal analysis indicate that tropospheric ozone (O_3) concentrations in Campo Grande exhibit moderate background levels with episodic peaks, reflecting the combined influence of photochemical production, meteorological variability, and precursor availability. The observed mean concentration (19.14 ppb) and maximum values exceeding 60 ppb suggest that, although typical levels remain moderate, short-term pollution events may occur under favorable atmospheric conditions.

These high-concentration episodes are likely associated with periods of enhanced solar radiation, reduced cloud cover, and atmospheric stagnation, which favor ozone accumulation and photochemical reactions involving NO_x and VOCs [1,2]. Similar patterns have been reported in other tropical and subtropical urban environments, where seasonal variability strongly modulates ozone formation [10–12].

In addition, the presence of BVOCs emitted by regional vegetation may contribute to ozone formation, particularly under high-radiation conditions typical of central Brazil [6,7]. Although Campo Grande is not located within the Amazon rainforest, long-range transport of BVOCs and their oxidation products can influence regional atmospheric chemistry, reinforcing the importance of considering both local and regional precursor sources [7–9].

4.2. Influence of Seasonal Meteorology and Biomass Burning

The results confirm that ozone variability in Campo Grande is strongly controlled by seasonal meteorological conditions. During the dry season, reduced precipitation, increased solar radiation, and enhanced atmospheric stability create favorable conditions for ozone formation and accumulation.

Biomass burning plays a critical role in this context by injecting large amounts of ozone precursors (NO_x, VOCs, and CO) into the atmosphere [8,9,13]. These emissions enhance photochemical activity and may lead to significant increases in ozone concentrations, particularly during prolonged dry periods.

Furthermore, regional atmospheric transport can carry biomass-burning plumes over long distances, affecting air quality in areas far from emission sources [9]. This mechanism is especially relevant in central Brazil, where seasonal circulation patterns facilitate the dispersion of pollutants from the Amazon and Cerrado regions.

During the rainy season, increased cloudiness, precipitation, and atmospheric mixing reduce solar radiation and enhance pollutant removal processes, leading to lower ozone concentrations. This seasonal contrast is consistent with previous studies conducted in Brazilian cities and other tropical environments [10–12].

It is important to note that the ozone time series analyzed in this study is derived from satellite-based products

(SISAM/INPE), which may be subject to uncertainties related to sensor calibration, retrieval algorithms, and changes in satellite instrumentation over time. Such factors can potentially introduce artificial variability or low-frequency biases that may influence long-term trend estimates. Therefore, although the detected trend is statistically significant, its low explanatory power ($R^2 \approx 0.009$) suggests that it should be interpreted with caution, as it may partially reflect observational uncertainties rather than purely atmospheric processes.

4.3. Physical Interpretation of SSA Components

The SSA decomposition revealed that ozone variability is dominated by annual, semiannual, and intra-seasonal oscillatory components, which can be linked to known atmospheric processes.

The annual component reflects the strong modulation imposed by the regional climate cycle, particularly the alternation between dry and wet seasons associated with the South American Monsoon System (SAMS)[1,2]. This cycle controls key factors such as solar radiation, boundary-layer dynamics, and atmospheric stability, which directly influence ozone production.

The semiannual component may be associated with transitional periods between seasons, when shifts in atmospheric circulation and cloud cover modify photochemical conditions and precursor transport.

The intra-seasonal variability likely reflects short-term atmospheric processes, including synoptic-scale circulation patterns and intermittent events such as biomass burning and pollution transport[8,9]. In tropical South America, phenomena such as the South Atlantic Convergence Zone (SACZ) can modulate cloudiness and precipitation, thereby influencing ozone formation on sub-seasonal time scales.

Although the SSA–LRF framework is univariate, the extracted components are physically consistent with known meteorological and photochemical processes, supporting the interpretability of the decomposition.

4.4. Long-Term Trend and Its Implications

Long-term changes in tropospheric ozone concentrations have been widely investigated due to their implications for air quality, climate forcing, and public health[24–26]. Previous studies indicate that ozone trends are highly heterogeneous across regions, reflecting the combined influence of anthropogenic emissions, atmospheric chemistry, and climate variability[27–29]. In particular, regional analyses have shown that even statistically significant trends may explain only a small fraction of total variability, as ozone dynamics are often dominated by seasonal and short-term processes[30–32].

In parallel, methodological advances in time series analysis have emphasized the importance of robust statistical frameworks for detecting and interpreting long-term trends in complex environmental datasets. The use of consistent performance metrics and decomposition-based approaches has become essential for distinguishing structured variability from stochastic fluctuations and observational noise[25]. Moreover, recent studies applying Singular Spectrum Analysis (SSA) to atmospheric pollutants have demonstrated its effectiveness in separating low-frequency components from dominant oscillatory modes, thereby improving the interpretation of trend signals in non-stationary time series[26].

The trend analysis revealed a statistically significant but weak positive increase in ozone concentrations over the study period. Despite its statistical significance, the low explained variance ($R^2 \approx 0.009$) indicates that long-term changes play a minor role compared to seasonal and short-term variability.

This finding is consistent with previous studies showing that ozone variability in tropical regions is often dominated by seasonal meteorology and episodic emissions rather than strong long-term trends[29,30,33].

From an air quality management perspective, this suggests that mitigation strategies should focus on controlling precursor emissions during critical periods (e.g., dry season) rather than relying solely on long-term trend reductions.

Although the extracted oscillatory modes are consistent with known regional climatic processes, such as the South American Monsoon System and biomass burning dynamics, it is important to emphasize that the SSA framework is inherently statistical. Therefore, the associations proposed here should be interpreted as plausible explanations rather than direct causal relationships. Future studies incorporating meteorological variables and emission proxies would be necessary to establish stronger physical links.

4.5. Forecast Performance and Comparison with Models

The LRF model produced a 30-day forecast with RMSE = 0.79 ppb, MAE = 0.64 ppb, and MAPE = 3.8%, indicating good agreement between predicted and observed ozone concentrations over the validation period. These metrics are widely used in environmental forecasting and provide complementary measures of predictive accuracy [4].

A key strength of the present evaluation is the use of a strict out-of-sample framework, in which the model was trained using data from 2000–2017 and validated exclusively on 2018. This approach reduces the risk of overfitting and ensures that predictive performance reflects the model's generalization capability [11].

To ensure a consistent and interpretable benchmark, model performance was compared against two simple baseline approaches: persistence (naïve) and seasonal naïve forecasts. These models represent standard references in forecasting practice and provide a minimum-skill baseline under the same data and validation conditions [11]. The results show that SSA–LRF outperforms both baselines, indicating that the method captures temporal dependencies beyond simple short-term persistence and seasonal repetition.

It is important to emphasize that the SSA–LRF framework is a univariate statistical approach, fundamentally different from physically based models such as Chemical Transport Models (CTMs), which explicitly simulate atmospheric chemistry and transport processes. Therefore, direct quantitative comparisons between these model classes are not appropriate due to differences in model structure, spatial scale, and input data requirements.

Similarly, although classical statistical models such as ARIMA and more recent machine learning approaches are widely used in air quality forecasting, their implementation requires additional modeling assumptions, parameter tuning, or external predictors [3,11]. As these models were not implemented within the same experimental framework, comparisons are discussed here only at a conceptual level. Future studies should include systematic benchmarking analyses incorporating ARIMA-type and hybrid models under identical training–testing conditions.

Overall, the results indicate that the SSA–LRF approach provides a transparent, computationally efficient, and consistent framework for short-term ozone forecasting in tropical environments, particularly in regions where data availability and model interpretability are key constraints [17,18].

4.6. Strengths, Limitations, and Future Perspectives

The main strength of this study lies in the integration of SSA for diagnostic analysis and LRF for forecasting, providing both interpretability and predictive capability within a unified framework.

However, some limitations should be acknowledged. First, the model is univariate and does not explicitly incorporate meteorological variables or precursor concentrations (NO_x, VOCs, BVOCs), which are known to influence ozone formation [1,2].

Second, although baseline models were included, further comparisons with chemical transport models and advanced machine learning approaches would strengthen the evaluation.

Future studies should explore hybrid approaches (e.g., SSA–ARIMA, SSA–LSTM) and multivariate models incorporating meteorological drivers, satellite-derived precursor data, and biomass-burning indicators. Such approaches may improve predictive accuracy and enhance the understanding of ozone formation mechanisms.

Additionally, the integration of satellite data with ground-based observations and chemical transport modeling could provide a more comprehensive framework for air quality assessment in tropical regions.

In recent decades, tropospheric ozone concentrations have increased in many regions worldwide due to anthropogenic emissions, land-use changes, and climate variability [3–5,27–32].

5. Conclusions

This study applied Singular Spectrum Analysis (SSA) and the Linear Recurrent Formula (LRF) to diagnose temporal structures and to produce short-term forecasts of daily tropospheric ozone (O_3) concentrations in Campo Grande, Brazil, using SISAM/INPE satellite-based observations for the period 2000–2018. The main conclusions are as follows:

1. Ozone variability in Campo Grande is strongly structured and dominated by seasonal oscillations.

The daily series exhibits marked oscillatory behavior, with moderate background levels and episodic high-concentration events, particularly during periods consistent with dry-season conditions.

2. SSA provided a transparent decomposition into physically interpretable modes.
   The decomposition revealed dominant annual and sub-annual (semiannual and intra-seasonal) components, supporting the interpretation that seasonal climate regimes and short-term atmospheric variability are the primary drivers of ozone behavior in the study region. The w-correlation diagnostics indicated satisfactory separability among reconstructed components, strengthening confidence in the extracted oscillatory modes.
3. Window-length sensitivity highlights the importance of embedding selection.
   The evaluation of L = 30, 60, and 90 showed that longer windows increasingly emphasize low-frequency structures, while shorter windows retain more high-frequency variability. The intermediate window (L = 60) provided a robust compromise, yielding clear harmonic modes with limited mixing.
4. A weak but statistically significant increasing trend was detected, but it explains little of the total variance. Linear regression indicated a positive slope over 2000–2018, while the explained variance remained very low, confirming that seasonal and short-term variability dominate the temporal behavior of O_3.
5. LRF achieved accurate 30-day forecasts under a strict out-of-sample evaluation design.
   Using 2000–2017 for training and 2018 for independent testing, the LRF model delivered RMSE = 0.79 ppb, MAE = 0.64 ppb, and MAPE = 3.8%, demonstrating strong agreement between predicted and observed values over the 30-day horizon. These results indicate that the SSA–LRF framework can provide reliable short-term forecasts while maintaining low computational cost and interpretability.

Limitations and future work. The proposed framework is univariate and does not explicitly incorporate meteorological covariates or precursor information, which may further improve prediction and process understanding. Future research should include baseline comparisons (e.g., persistence, seasonal naïve, ARIMA/SARIMA) under the same train–test split and explore hybrid or multivariate extensions incorporating meteorological drivers.

Overall, the SSA–LRF framework proved to be an effective and interpretable approach for diagnosing and forecasting tropospheric ozone variability in a tropical urban environment, supporting its application in air-quality assessment and short-term pollution management.

Funding

This study did not receive external funding.

Institutional Review Board Statement

Not applicable. This study does not involve human or animal subjects.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data used in this study are publicly available from the SISAM/INPE platform. Processed datasets and scripts used for analysis are available from the corresponding author upon reasonable request.

Acknowledgments

The author would like to express his sincere gratitude to the university for its institutional support in the development of this research. He is also thankful to the administrative and academic teams of the institution for providing the necessary infrastructure, technical assistance, and encouragement that significantly contributed to the successful completion of this work.

Conflicts of Interest

The author declares no competing interests.

AI Use Statement

During the preparation of this manuscript, the author did not use AI. No artificial intelligence tools were used for data analysis, interpretation, or the generation of scientific results. All generated content was critically reviewed and edited by the author. The author assumes full responsibility for the integrity, accuracy, and originality of the work.

References

[1] Monks, P.S., Archibald, A.T., Colette, A., et al., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmospheric Chemistry and Physics. 15(15), 8889–8973. DOI: https://doi.org/10.5194/acp-15-8889-2015

[2] Seinfeld, J.H., Pandis, S.N., 2016. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 3rd ed. Wiley: Hoboken, NJ, USA.

[3] World Health Organization, 2021. WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. WHO: Geneva, Switzerland.

[4] Cooper, O.R., Parrish, D.D., Ziemke, J., et al., 2014. Global distribution and trends of tropospheric ozone: An observation-based review. Elementa: Science of the Anthropocene. 2, 000029. DOI: https://doi.org/10.12952/journal.elementa.000029

[5] Intergovernmental Panel on Climate Change, 2021. Climate Change 2021: The Physical Science Basis. IPCC: Geneva, Switzerland.

[6] Guenther, A., Jiang, X., Heald, C.L., et al., 2012. The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. Geoscientific Model Development. 5, 1471–1492.

[7] Sindelarova, K., Granier, C., Bouarar, I., et al., 2014. Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years. Atmospheric Chemistry and Physics. 14(17), 9317–9341.

[8] Crutzen, P.J., Andreae, M.O., 1990. Biomass burning in the tropics: Impact on atmospheric chemistry and biogeochemical cycles. Science. 250(4988), 1669–1678. DOI: https://doi.org/10.1126/science.250.4988.1669

[9] Freitas, S.R., Longo, K.M., Silva Dias, M.A.F., et al., 2005. Monitoring the transport of biomass burning emissions in South America. Environmental Fluid Mechanics. 5(1–2), 135–167. DOI: https://doi.org/10.1007/s10652-005-0243-7

[10] de Souza, A., de Medeiros, E.S., Özonur, D., et al., 2024. Dynamic monitoring of tropospheric ozone concentrations in Northeast and Midwest Brazil: Insights into seasonal variations and climatic influences. International Journal of Engineering & Technology. 13(2), 204–211.

[11] Dias Nunes, M., Mariano, G.L., Alonso, M.F., 2020. Spatio-temporal variability of total ozone column and ultraviolet radiation: assessment of relationship in South America. Revista Brasileira De Geografia Física. 13(5), 2053–2073. DOI: https://doi.org/10.26848/rbgf.v13.5.p2053-2073 (in Portuguese)

[12] de Souza, A., de Oliveira-Júnior, J.F., Abreu, M.C., et al., 2022. Spatial–temporal variability of the ozone column over the Brazilian Midwest from satellite data from 2005 to 2020. Water, Air, & Soil Pollution. 233, 59. DOI: https://doi.org/10.1007/s11270-022-05532-w

[13] Jaffe, D.A., Wigder, N.L., 2012. Ozone production from wildfires: A critical review. Atmospheric Environment. 51, 1–10. DOI: https://doi.org/10.1016/j.atmosenv.2011.11.063

[14] Golyandina, N., Zhigljavsky, A., 2013. Singular Spectrum Analysis for Time Series. Springer: Berlin, Germany.

[15] Hyndman, R.J., Athanasopoulos, G., 2021. Forecasting: Principles and Practice, 3rd ed. OTexts: Melbourne, Australia.

[16] Hassani, H., Mahmoudvand, R., 2013. Multivariate singular spectrum analysis: A general view and new vector forecasting approach. International Journal of Energy and Statistics. 1(1), 55–83.

[17] Golyandina, N., 2020. Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing. WIREs Computational Statistics. 12(4), e1487. DOI: https://doi.org/10.1002/wics.1487

[18] Vautard, R., Yiou, P., Ghil, M., 1992. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. Physica D: Nonlinear Phenomena. 58(1–4), 95–126. DOI: https://doi.org/10.1016/0167-2789(92)90103-T

[19] Sun, M., Li, X., 2022. Window length selection of singular spectrum analysis and application to precipitation time series. Global NEST Journal. 19(2), 306–317.

[20] Veefkind, J.P., Aben, I., McMullan, K., et al., 2012. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sensing of Environment. 120, 70–83. DOI: https://doi.org/10.1016/j.rse.2011.09.027

[21] Ziemke, J.R., Chandra, S., Duncan, B.N., et al., 2006. Tropospheric ozone determined from Aura OMI and MLS: Evaluation of measurements and comparison with the Global Modeling Initiative's Chemical Transport Model. Journal of Geophysical Research. 111(D19).

[22] Martin, R.V., 2008. Satellite remote sensing of surface air quality. Atmospheric Environment. 42(34),

7823–7843. DOI: https://doi.org/10.1016/j.atmosenv.2008.07.018

[23] Levelt, P.F., van den Oord, G.H.J., Dobber, M.R., et al., 2006. The Ozone Monitoring Instrument. IEEE Transactions on Geoscience and Remote Sensing. 44(5), 1093–1101. DOI: https://doi.org/10.1109/TGRS.2006.872333

[24] Lu, W.Z., He, H.D., Chen, W.J., 2021. A novel integrated singular spectrum analysis and autoregressive model for $PM_{2.5}$ forecasting. Atmospheric Environment. 214, 116873.

[25] Wilks, D.S., 2019. Statistical Methods in the Atmospheric Sciences, 4th ed. Elsevier: Amsterdam, The Netherlands.

[26] de Souza, A., Nunes, R.S.C., Oliveira Junior, J.F., et al., 2026. Seasonal patterns and forecasting of CO and ozone using singular spectrum analysis in a tropical urban environment. Journal of Atmospheric Science Research. 9(1), 1–15.

[27] Sicard, P., De Marco, A., Agathokleous, E., et al., 2023. Amplified ozone pollution in cities during the COVID-19 lockdown. Science of the Total Environment. 735, 139542.

[28] Wang, T., Xue, L., Brimblecombe, P., et al., 2017. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. Science of The Total Environment. 575, 1582–1596.

[29] Archibald, A.T., Neu, J.L., Elshorbany, Y.F., et al., 2020. Tropospheric ozone assessment report: A critical review of changes in the tropospheric ozone burden and budget from 1850 to 2100. Elementa: Science of the Anthropocene. 8(1), 34.

[30] Li, K., Jacob, D.J., Shen, L., et al., 2020. Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences. Atmospheric Chemistry and Physics. 20(19), 11423–11433.

[31] de Souza, A., Palácios, R.S., Oliveira, J.F., et al., 2026. Modeling and prediction of tropospheric ozone based on singular spectrum analysis and linear recurrent formula in Campo Grande, Brazil. Meteorology and Atmospheric Physics. 138, 36. DOI: https://doi.org/10.1007/s00703-026-01128-9

[32] de Souza, A., Palácios, R.S., Nassarden, D.C.S., et al., 2025. Spectral decomposition and temporal dynamics of CO and O_3 in Campo Grande, Brazil: A singular spectrum analysis approach. Air Quality, Atmosphere & Health. 18, 3679–3692. DOI: https://doi.org/10.1007/s11869-025-01857-7

[33] Zhang, B., Rong, Y., Yong, R., et al., 2022. Deep learning for air pollutant concentration prediction: A review. Atmospheric Environment. 290, 119347. DOI: https://doi.org/10.1016/j.atmosenv.2022.119347