BILINGUAL PUBLISHING GROUP

Journal of Atmospheric Science Research

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

ARTICLE

Seasonal Patterns and Forecasting of CO and Ozone Using Singular Spectrum Analysis in a Tropical Urban Environment

Amaury de Souza [1] ***, Raquel Soares Casaes Nunes*** [2*] ***, José Francisco de Oliveira Junior*** [3] ***, Ivana Pobocikova*** [4] ***, Sianny Vanessa da Silva Freitas*** [5] ***, Kelvy Rosalvo Alencar Cardoso*** [3]

[1] *Institute of Physics, University of Mato Grosso do Sul, Campo Grande 79070-900, Brazil*

[2] *Saude-Decania Science Center, Federal University of Rio de Janeiro, Rio de Janeiro 21941-901, Brazil*

[3] *Institute of Atmospheric Sciences (ICAT), University of Alagoas, Maceió 57072-900, Brazil*

[4] *Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Zilina, 010 26 Zilina, Slovakia*

[5] *Institute of Tropical Diseases, Federal University of Pará, Belém 66075-900, Brazil*

ABSTRACT

Singular Spectrum Analysis (SSA) was applied to daily time series of carbon monoxide (CO) and ozone (O_3) observed between 2000 and 2018 in Campo Grande, MS, Brazil, to identify seasonal patterns, long-term variability, and evaluating the predictive capacity of the technique. The methodology involved the decomposition of the series into structural components and subsequent prediction using the Linear Recurrence Formula (LRF). The analysis revealed strong and persistent annual seasonality for both pollutants, particularly for CO, whose maximum concentrations occur between August and October, coinciding with the dry season and intensified biomass-burning activity. SSA proved effective in extracting low-frequency components, including trend and seasonal cycles, providing a clear representation of the dominant temporal structure of both pollutants. Forecasting results indicated that SSA-LRF successfully reproduced the main seasonal behavior of O_3, while daily prediction skill remained limited, as reflected by negative R^2 values during the validation period. For CO, the highly irregular and episodic nature of fire-related peaks resulted in larger forecast errors and reduced predictive skill. These results highlight that univariate SSA is more suitable for reconstructing and predicting low-frequency pollutant

*CORRESPONDING AUTHOR:

Raquel Soares Casaes Nunes, Saude-Decania Science Center, Federal University of Rio de Janeiro, Rio de Janeiro 21941-901, Brazil; Email: quelcasaes@micro.ufrj.br

ARTICLE INFO

Received: 1 November 2025 | Revised:17 December 2025 | Accepted: 25 December 2025 | Published Online: 2 January 2026
DOI: https://doi.org/10.30564/jasr.v9i1.12273

CITATION

de Souza, A., Nunes, R.S.C., de 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. DOI: https://doi.org/10.30564/jasr.v9i1.12273

dynamics than short-term daily variability. The findings demonstrate that SSA is a robust exploratory and decomposition tool for air-quality time series in tropical environments, particularly for identifying seasonal and structural patterns. For operational forecasting of pollutants with strong volatility, such as CO, hybrid approaches combining SSA with statistical or machine-learning models are recommended to improve predictive performance.

Keywords: Singular Spectrum Analysis; Air Pollution; CO; O_3; Seasonal Forecast

1. Introduction

Ozone (O_3) and carbon monoxide (CO) are key atmospheric pollutants with significant implications for urban air quality, climate dynamics, and human health. In the troposphere, O_3 is produced through photochemical reactions involving precursors such as CO, NO_x, and volatile organic compounds, modulated by solar radiation and meteorological conditions[1–5]. In tropical South America, the temporal behavior of CO and O_3 is strongly influenced by the seasonal contrast between wet and dry periods, biomass-burning emissions, and regional atmospheric circulation, resulting in well-defined annual cycles and notable interannual variability[6–10]. In particular, dry-season conditions favor reduced cloud cover, enhanced solar radiation, and lower planetary boundary layer heights, which jointly intensify pollutant accumulation and photochemical activity.

Environmental time-series analysis has increasingly adopted non-parametric approaches capable of extracting trends, oscillations, and noise from nonlinear or noisy datasets. Among these techniques, Singular Spectrum Analysis (SSA) stands out for its ability to decompose short and complex time series without requiring strong statistical assumptions[11–15]. SSA has been widely applied in climatology, hydrology, and geophysics for detecting periodic modes, identifying regime shifts, and improving predictability[16–20]. Its appeal lies in the capacity to isolate physically interpretable components—such as trend, seasonal cycles, and intra-seasonal oscillations—while remaining robust to noise and nonstationarity. However, despite these advantages, SSA remains underexplored in the analysis of long-term pollutant time series in tropical environments, where atmospheric variability is strongly shaped by biomass-burning events and episodic emissions.

A critical yet often overlooked aspect of SSA is the choice of window length (L), which directly affects component separability, spectral resolution, and forecasting performance. Although several studies using SSA or SSA–hybrid models—such as SSA–ARIMA, SSA–LSTM, and SSA–GARCH—have demonstrated promising results in environmental forecasting[20–22], most adopt fixed or heuristic values for L, with limited assessment of how window-length selection influences decomposition quality or predictive accuracy. This methodological gap is particularly relevant for pollutant time series affected by both photochemical processes and irregular fire-driven emissions, as commonly observed in tropical regions, where multiple temporal scales coexist.

Hybrid modeling frameworks integrating SSA with classical statistical or machine-learning approaches have advanced pollutant forecasting, especially for series exhibiting strong seasonality and nonlinear behavior[20–22]. Nevertheless, the effectiveness of such hybrid models remains closely linked to the internal structure of SSA decomposition, including the ability to properly separate low-frequency components from high-frequency variability. Consequently, understanding how window-length sensitivity alters component identification is essential for improving reconstruction and prediction, particularly when dealing with pollutants characterized by multiscale variability and episodic extremes.

In this context, the present study applies SSA to daily CO and O_3 concentrations measured in Campo Grande, Brazil, from 2000 to 2018. Campo Grande is located in central Brazil and is strongly influenced by seasonal biomass burning, regional transport of smoke plumes, and marked wet–dry climatic contrasts. These characteristics make it an ideal case study for assessing the performance of SSA in a tropical urban environment affected by both regular seasonal forcing and irregular emission events. The study pursues two main objectives:

(i) To identify dominant temporal patterns—including seasonal cycles, synoptic-scale oscillations, and long-term variability; and

(ii) To evaluate the forecasting performance of SSA using the Linear Recurrence Formula (LRF), complemented by hybrid SSA–ARIMA/SARIMA models under a rolling-origin validation scheme.

Despite the growing body of research on air-pollution forecasting using advanced statistical and machine-learning techniques, comparatively fewer studies have emphasized the importance of understanding the intrinsic temporal structure of pollutant time series prior to model implementation. In many cases, predictive performance is assessed without a detailed evaluation of scale-dependent variability, component separability, or the physical interpretability of extracted modes. This limitation is particularly evident in tropical environments, where pollutant dynamics are shaped by the superposition of regular seasonal forcing and irregular, event-driven emissions associated with biomass burning. Under such conditions, methodological choices—such as the selection of SSA window length and the criteria for component grouping—can substantially influence both reconstruction quality and forecasting outcomes. By explicitly addressing these aspects, the present study contributes not only to pollutant-specific analysis but also to broader methodological discussions in environmental time-series research. The emphasis on transparency, scale awareness, and physical interpretability responds directly to recent calls for more robust and reproducible statistical frameworks in air-quality studies, especially in regions characterized by strong climatic seasonality and emission intermittency.

The novelty of this work lies in its systematic evaluation of window-length sensitivity, quantitative assessment of separability using w-correlation metrics [23,24], and objective criteria for principal component selection. By applying SSA to pollutant time series strongly influenced by biomass-burning activity, this study advances the methodological understanding of SSA in environmental sciences and provides new insights into the multiscale atmospheric behavior of CO and O_3 in a tropical urban environment.

2. Methodology and Data

2.1. Study Area and Data

Daily concentrations of ozone (O_3, ppb) and carbon monoxide (CO, ppb) for 2000–2018 were obtained from the Air Quality Information System (SISAM/INPE), maintained by the National Institute for Space Research (INPE). Campo Grande, located in central Brazil, is characterized by a tropical climate with a marked wet–dry seasonal cycle and recurrent biomass-burning activity, which strongly influences pollutant levels [3,4]. Data was quality-controlled and aggregated to daily means. Stationarity was evaluated using the Augmented Dickey–Fuller test, confirming stochastic seasonality and long-term variability appropriate for non-parametric decomposition.

2.2. Singular Spectrum Analysis (SSA)

SSA was applied following the classical four-step framework documented in foundational SSA literature [5–9,14–16]:

1. **Embedding:**
   The original series is mapped into a trajectory matrix constructed using window length (L) [8].
2. **Decomposition:**
   Singular value decomposition (SVD) is applied to obtain eigenvalues, eigenvectors (EOFs), and principal components (PCs) [7–10].
3. **Grouping:**
   Components are grouped based on spectral similarity, relative contribution to variance, and w-correlation structure [23,25].
4. **Reconstruction:**
   Component groups (trend, seasonal, high-frequency noise) are recombined using diagonal averaging to recover reconstructed time series [8,10].

SSA is particularly suitable for nonlinear, noisy environmental series because it does not require parametric assumptions about pollutant dynamics [11,15,16].

2.3. Window-Length Selection

Window length (L) strongly influences spectral resolution and component separability in SSA, as emphasized by Golyandina and Zvonarev [23], Sun and Li [24]. We evaluated candidate values L = 6, 12, 60, 114 using:

- reconstruction error (RMSE);
- mean w-correlation between reconstructed components [24];

- consistency with dominant periodicities identified via periodogram.
- interpretability of extracted modes [24].

Short windows (L = 6–12) captured seasonal cycles but exhibited stronger mode mixing. The long window L = 114 ≈ N/2, frequently recommended for maximizing separability in SSA applications [4,24,25], provided the lowest reconstruction error and clearest distinction between trend, seasonal, and stochastic components. Thus, L = 114 was used for the main decomposition and forecasting.

2.4. Selection of SSA Components

Principal components were selected based on:

- cumulative explained variance;
- eigenvalue decay (scree plot) [7–10];
- spectral interpretation of PCs (annual, semiannual, intra-seasonal modes);
- w-correlation < 0.30, indicating noise-dominated behavior [23].

For both pollutants, PC1 and PC2 accounted for more than 90% of total variance and corresponded to interpretable physical modes (trend + annual cycle). Higher PCs represented noise or irregular fire-driven fluctuations.

2.5. Forecasting Using SSA-LRF and Hybrid Models

Forecasts were generated using the Linear Recurrence Formula (LRF) derived from SSA, following the forecasting framework described by Golyandina and Shapoval [13]. To improve predictive accuracy—especially for pollutants with irregular or nonlinear variability—we adopted hybrid modeling extensions inspired by recent environmental forecasting studies [20–22]:

- SSA–ARIMA for long-term components;
- SSA–SARIMA for seasonal components;
- ARMA modeling of high-frequency residuals.

These hybrid models exploit SSA's decomposition ability while enabling the capture of short-term or heteroscedastic dynamics that SSA-LRF alone cannot represent.

2.6. Validation Strategy

Forecast performance was assessed using a rolling-origin cross-validation approach commonly recommended for environmental time-series prediction [20,26,27].

- Training windows expanded progressively (e.g., 2000–2011 → test 2012–2013; … → 2000–2015 → test 2016–2018).

The following metrics were computed:

- RMSE, MAE, MAPE, and R^2.

This validation scheme provides a robust assessment of forecasting skill in series affected by both seasonal structure and episodic events such as fire emissions.

2.7. Methodological Contribution

The methodological contributions of this study include:

- A systematic evaluation of window-length sensitivity, rarely addressed in air-pollution SSA studies;
- Quantitative separability assessment via w-correlation [24];
- A clear, objective framework for principal component selection;
- Integration of SSA with ARIMA/SARIMA hybrid structures, consistent with recent advances in environmental forecasting [20–22];
- Application of SSA to pollutant time series strongly influenced by fire activity—a sparse context in current SSA literature.

3. Results

3.1. Seasonal Behavior of CO and O_3

Figure 1 presents the monthly boxplots of daily CO and O_3 concentrations for the period 2000–2018, providing an overview of the seasonal distribution, central tendency, and dispersion of both pollutants. The boxplots clearly highlight the pronounced dry-season enhancement of CO concentrations between August and October, characterized by high medians, wide interquartile ranges, and numerous extreme outliers associated with biomass-burning events. In contrast, O_3 exhibits a more regular and symmetric seasonal pattern,

with lower variability and a gradual increase toward the late dry season, reflecting the dominant role of photochemical production under enhanced solar radiation and reduced cloud cover.

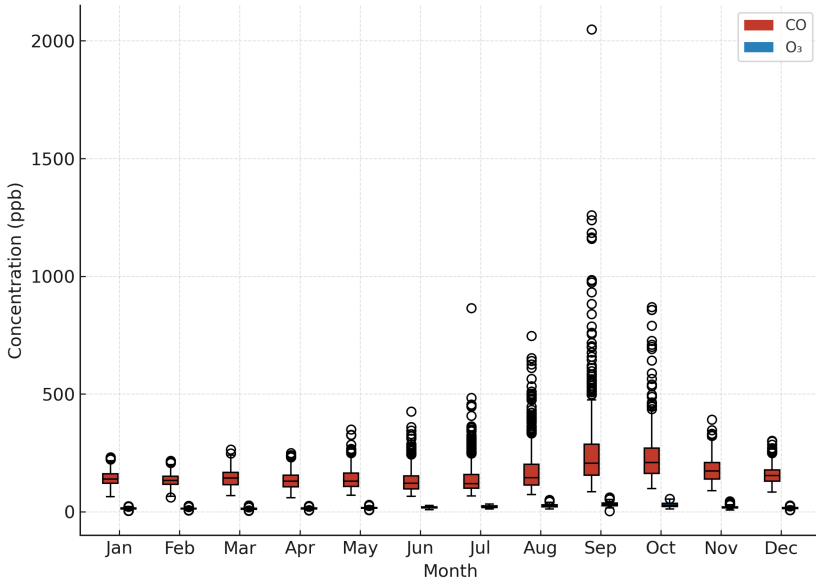


Figure 1. Monthly boxplots of CO and O_3 concentrations (2000–2018), highlighting higher values during the dry season.

Both pollutants exhibit a clear seasonal cycle. CO peaks sharply between August and October, showing wide dispersion and numerous outliers associated with biomass-burning episodes and reduced dispersion during the dry season. O_3 presents a smoother annual pattern, with minima from January to April and gradual increases toward late winter due to enhanced photochemical activity.

Figure 2 shows the monthly mean concentrations of CO and O_3 together with their associated variability (±1 standard deviation), reinforcing the seasonal contrasts identified in the boxplot analysis. The results indicate a pronounced seasonal amplitude for CO, with substantially higher mean values during the dry season, reflecting the cumulative effect of biomass-burning emissions and reduced atmospheric dispersion. In contrast, O_3 exhibits a smoother and more stable annual cycle, with moderate variability and a gradual increase toward the late dry season, consistent with enhanced photochemical production under favorable meteorological conditions.

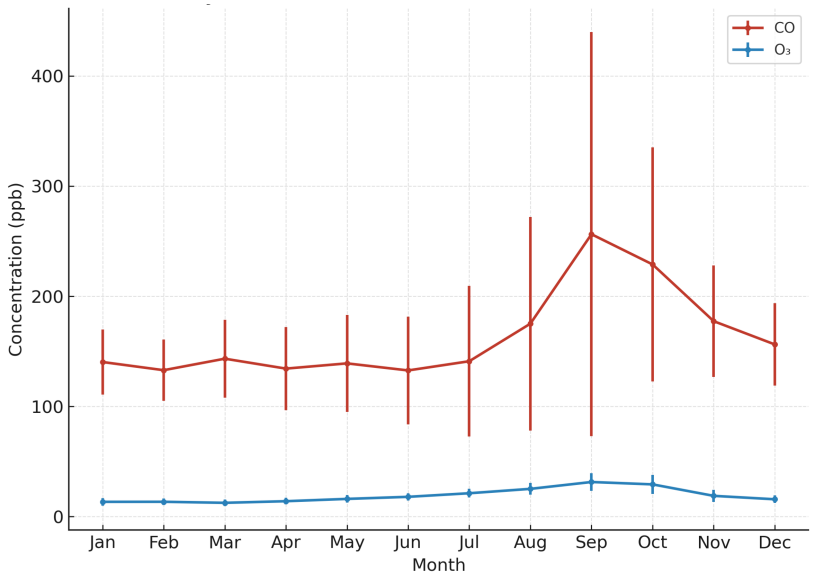


Figure 2. Monthly mean CO and O_3 concentrations (± 1 SD), illustrating seasonal contrasts between wet and dry periods.

CO shows a strong seasonal amplitude (≈ 82.6 ppb), while O_3 displays a moderate but consistent annual modulation (≈ 7.1 ppb). These seasonal signals are consistent with meteorological forcing and dry-season fire activity.

3.2. SSA Window-Length Assessment

Figure 3 compares the SSA eigenvalue spectra obtained using different window lengths (L = 6, 12, and 114), illustrating the strong dependence of component separability on the choice of L. For short windows (L = 6 and 12), the eigenvalues decay gradually, indicating substantial mode mixing and limited spectral resolution. In contrast, the longer window (L = 114) exhibits a clear separation between the leading components and the remaining eigenvalues, reflecting improved discrimination between trend, seasonal, and stochastic variability and supporting its selection for the subsequent SSA decomposition and forecasting analyses.

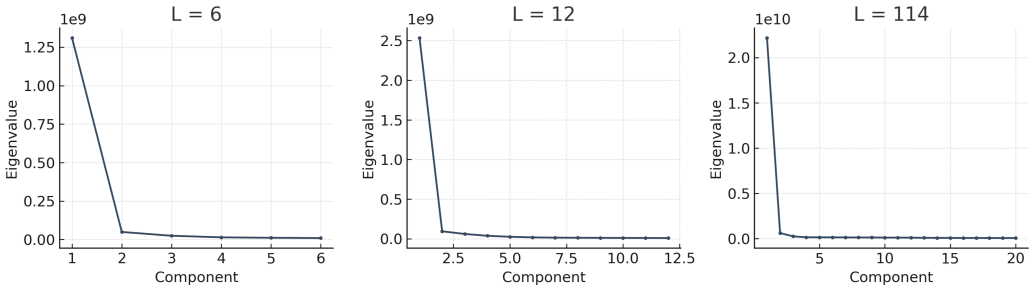


Figure 3. SSA eigenvalues (first 20 components) for L = 6, 12, and 114, showing improved separability with longer windows.

Short windows (L = 6, 12) capture basic seasonality but exhibit component mixing. The long window L = 114 (≈ N/2) achieves the best separability and lowest reconstruction error, confirming its suitability for extracting multiscale structure.

3.3. Spectral Structure of CO and O_3

Figure 4 presents the periodograms of the CO and O_3 time series for the 2000–2018 period, highlighting the dominant spectral features that govern their temporal variability. Both pollutants exhibit a pronounced annual peak, confirming the strong seasonal forcing associated with the wet–dry climatic cycle. Secondary spectral peaks at semiannual and intra-seasonal frequencies are also observed, particularly for CO, indicating the influence of synoptic-scale processes and episodic biomass-burning activity. The higher overall spectral power of CO reflects its greater variability compared to O_3, which is primarily controlled by smoother photochemical and meteorological processes.

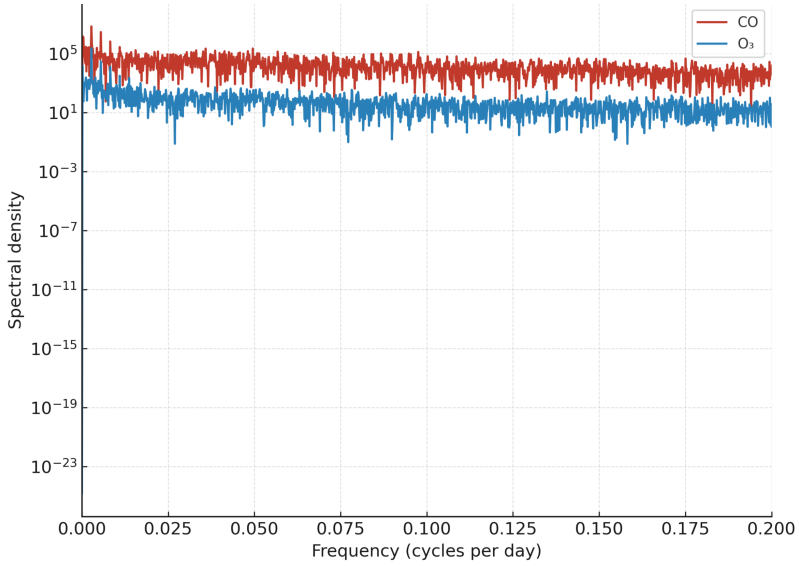


Figure 4. Periodogram of CO and O_3 (2000–2018) showing dominant annual and semiannual frequencies.

Both series show strong annual peaks near 0.083 cycles/month, with additional semiannual components. CO exhibits higher overall power, consistent with greater variability from fires and meteorology.

3.4. SSA Eigenvalues and Component Interpretation

Figure 5 displays the ordered SSA eigenvalues obtained with the selected window length (L = 114), illustrating the concentration of variance in the leading components of both pollutant time series. The sharp decay after the first few eigenvalues indicates that most of the signal energy is captured by a small number of components, primarily associated with the dominant annual cycle and low-frequency variability. Subsequent eigenvalues exhibit a flatter distribution, suggesting noise-dominated or irregular components linked to short-term fluctuations and episodic emission events.

PC1 and PC2 explain most of the variance and correspond to the annual cycle and intra-seasonal variability. Higher PCs represent irregular, noise-like fluctuations.

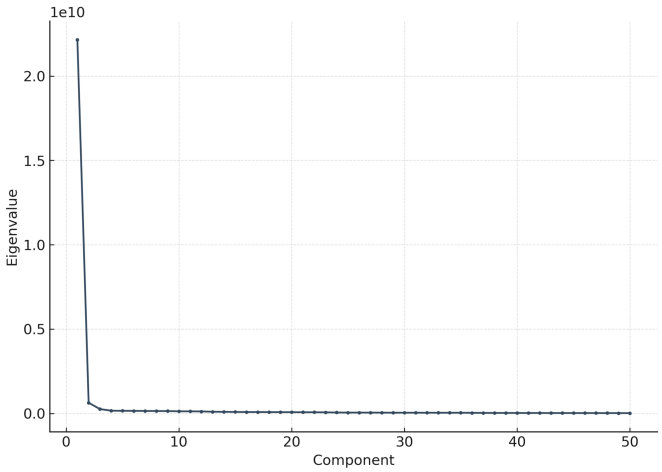


Figure 5. Ordered SSA eigenvalues (L = 114), showing concentration of variance in the first components.

3.5. Reconstruction of Dominant Components

Figure 6 compares the original CO and O_3 time series with their SSA-reconstructed counterparts obtained using the leading components (PC1–PC2). The reconstruction effectively preserves the dominant seasonal structure of both pollutants while attenuating high-frequency variability. For CO, extreme short-term peaks associated with biomass-burning events are partially smoothed, although their seasonal timing remains clearly identifiable. In contrast, the reconstructed O_3 series closely follows the observed signal, reflecting the stronger influence of regular photochemical and meteorological forcing on its temporal variability.

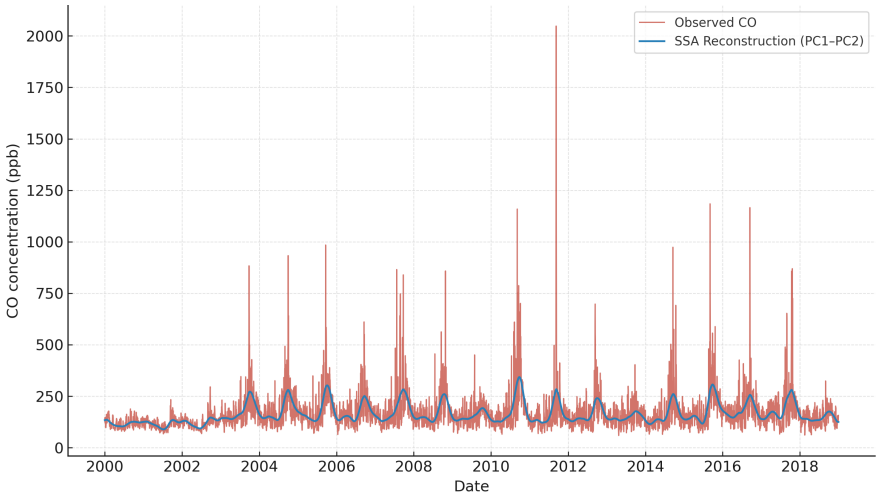


Figure 6. Original and SSA-reconstructed CO and O_3 series using PC1–PC2.

Reconstruction preserves the main seasonal cycle while smoothing high-frequency variability. For CO, extreme fire-driven peaks are attenuated but their seasonal timing remains. For O_3, the reconstruction closely follows observations.

3.6. Daily Variability and Multiscale Patterns

Figure 7 illustrates the daily CO and O_3 concentrations over the 2000–2018 period, emphasizing the contrasting temporal behavior of the two pollutants at short time scales. CO exhibits pronounced day-to-day variability, with abrupt and sporadic peaks associated with biomass-burning episodes, atmospheric stagnation, and regional transport. In contrast, O_3 displays a comparatively smoother temporal evolution, with gradual fluctuations superimposed on its seasonal cycle, reflecting its stronger dependence on photochemical processes and synoptic meteorological conditions.

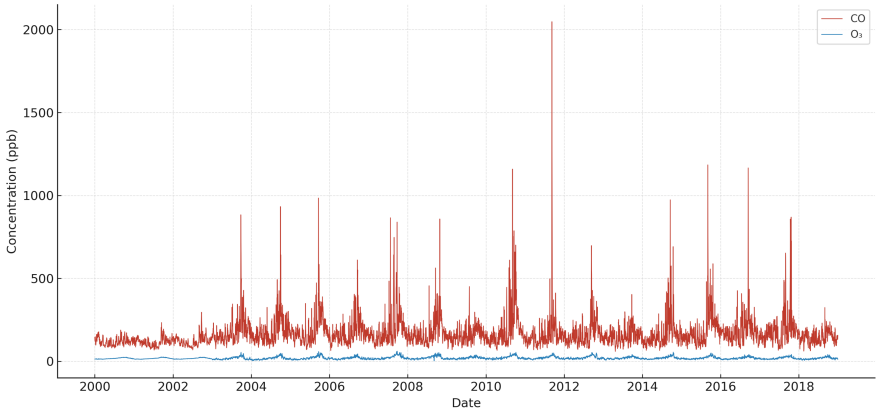


Figure 7. Daily CO and O_3 concentrations (2000–2018), showing episodic CO peaks and smoother O_3 behavior.

CO presents abrupt peaks often exceeding 2000 ppb, reflecting fire events and stagnation episodes. O_3 varies more smoothly but still increases in the dry season due to enhanced photochemistry.

3.7. Principal Components and Intra-Seasonal Modes

Figure 8 shows the explained variance associated with the first two principal components derived from the joint analysis of CO and O_3, highlighting the dominant modes of shared variability between the two pollutants. The first principal component (PC1) accounts for most of the total variance and represents the common annual seasonal cycle, driven by regional climatic forcing and biomass-burning activity. The second principal component (PC2) captures sub-seasonal variability, reflecting shorter-term atmospheric processes that modulate pollutant concentrations beyond the dominant seasonal signal.

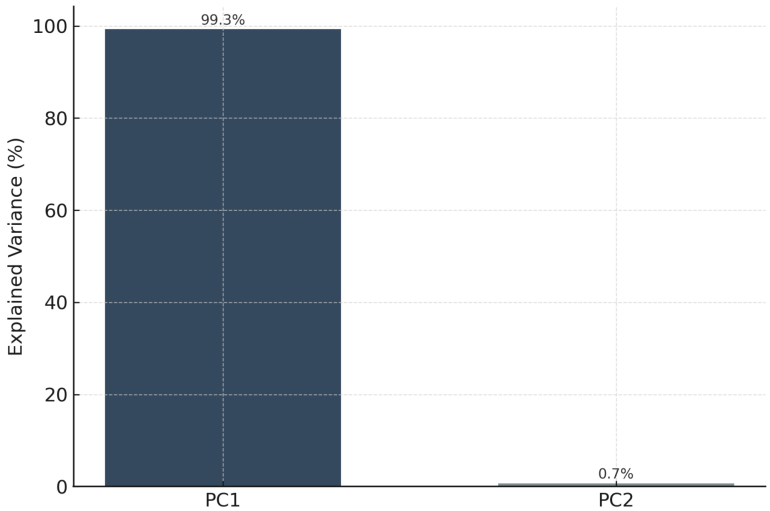


Figure 8. Explained variance of PC1 and PC2 from joint PCA of CO and O_3.

PC1 captures shared annual seasonality; PC2 captures sub-seasonal variability.

Figure 9 illustrates the first SSA principal component (PC1) associated with CO, which represents the dominant low-frequency mode of variability in the time series. This component clearly captures the annual seasonal cycle, characterized by enhanced values during the dry season and lower levels during the wet period. The temporal structure of PC1 reflects the combined influence of regional climatic seasonality and recurrent biomass-burning activity, confirming its role as the primary driver of long-term CO variability.

PC1 aligns with dry-season maxima and wet-season minima.

Figures 10a and **10b** present the higher-order SSA principal components associated with CO variability, highlighting distinct modes of intra-seasonal and irregular behavior. **Figure 10a** shows PC2, which exhibits an oscillatory pattern with a characteristic period of approximately 60 days, indicative of intra-seasonal variability driven by synoptic-scale atmospheric processes and modulation of biomass-burning activity. In contrast, **Figure 10b** displays PC3, characterized by more irregular and intermittent fluctuations, reflecting medium-scale variability linked to episodic pollution events and short-term atmospheric dynamics.

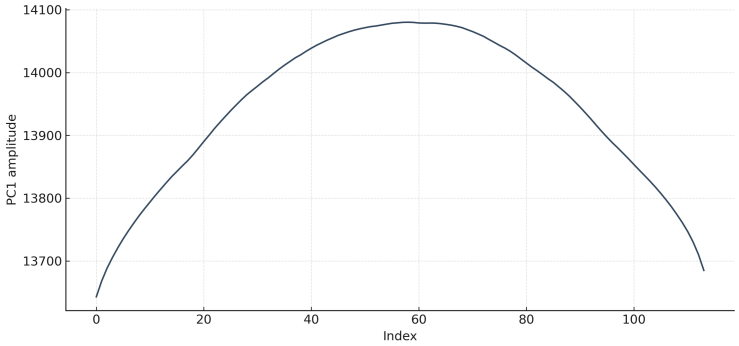
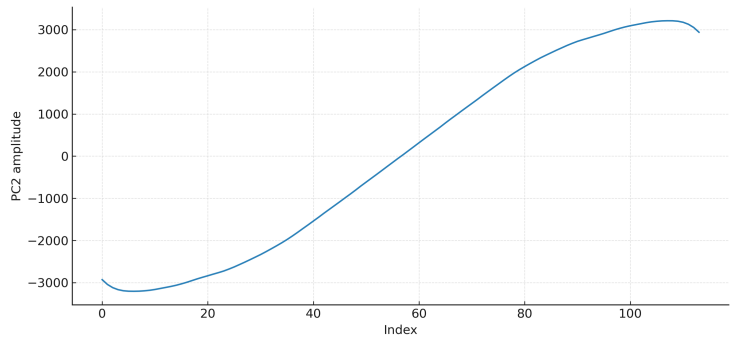
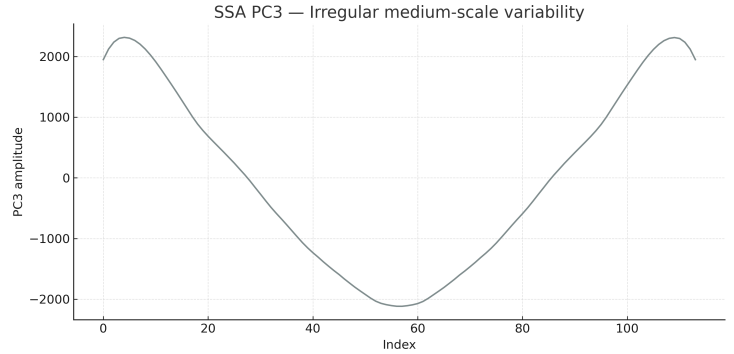


Figure 9. SSA PC1 representing the dominant annual mode of CO variability.



(**a**)



(**b**)

Figure 10. (**a**) SSA PC2 showing ≈ 60-day intra-seasonal oscillation driven by synoptic and fire-related variability. (**b**) SSA PC3 capturing irregular medium-scale fluctuations linked to episodic pollution.

Figure 11 presents an idealized 60-day sinusoidal signal used as a reference to support the interpretation of the intra-seasonal oscillations identified in the SSA decomposition. The similarity between this idealized pattern and the temporal structure of PC2 reinforces the identification of a quasi-60-day mode, suggesting that the extracted component represents a physically meaningful intra-seasonal process rather than random noise. This comparison aids in distinguishing coherent oscillatory behavior from irregular variability in the higher-order SSA components.

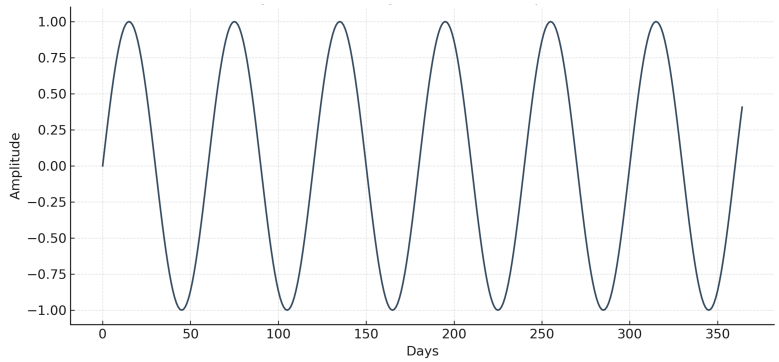


Figure 11. Idealized 60-day sinusoid used to assist interpretation of intra-seasonal SSA modes.

3.8. SSA Reconstruction of Low-Frequency Variability

Figure 12 shows the SSA-reconstructed CO and O_3 time series obtained by combining the leading low-frequency components (PC1–PC2), emphasizing the dominant seasonal variability of both pollutants. The reconstruction highlights the persistence and timing of the annual cycle, with enhanced concentrations during the dry season and reduced levels during the wet period. While the reconstructed O_3 series closely follows the observed signal, the CO reconstruction smooths high-frequency fire-related peaks, underscoring the capability of SSA to isolate coherent low-frequency behavior while filtering irregular short-term fluctuations.

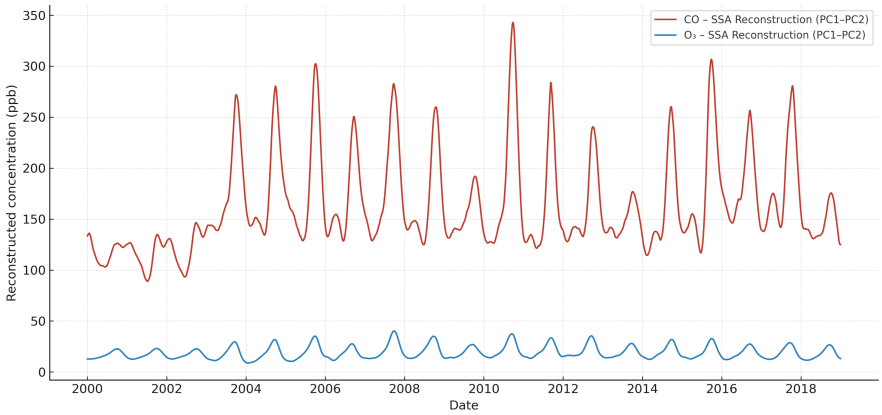


Figure 12. SSA-reconstructed CO and O_3 series (PC1–PC2), highlighting dominant seasonal variability.

O_3 reconstruction remains highly consistent with observations; CO reconstruction smooths irregular fire peaks but retains their seasonal timing.

3.9. Multiscale Wavelet Patterns

Figure 13 presents the wavelet power spectra of CO and O_3 at daily and monthly scales, highlighting the multiscale temporal variability of both pollutants. CO exhibits strong power at high-frequency bands, particularly in the 2–15-day range during the dry season, as well as persistent intra-seasonal variability at periods of approximately 30–60 days, reflecting the influence of biomass-burning activity and synoptic-scale processes. In contrast, O_3 is dominated by low-frequency variability, with a pronounced annual cycle and comparatively weaker high-frequency power, consistent with its smoother photochemical and meteorological control.

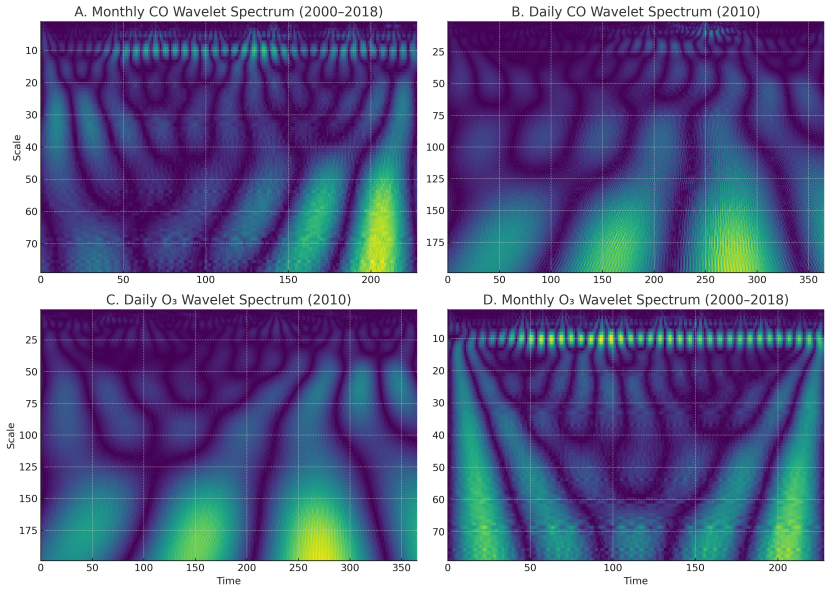


Figure 13. Wavelet spectra of CO and O_3 at monthly and daily scales.

CO shows strong high-frequency power (2–15 days) during burning seasons and persistent 30–60-day intraseasonal modes. O_3 is dominated by annual cycles and moderate short-term variability associated with meteorological forcing.

3.10. Forecast Evaluation

Figure 14 compares the observed and SSA-based predicted daily concentrations of CO and O_3 during the validation period (2016–2018), illustrating the forecasting capability of the univariate SSA–LRF approach. The results show that the model successfully reproduces the low-frequency seasonal structure of both pollutants but fails to capture short-term variability and abrupt concentration changes. This limitation is particularly evident for CO, whose fire-related peaks are largely underestimated, while O_3 forecasts reflect the predictable seasonal cycle but still miss high-frequency meteorological fluctuations.

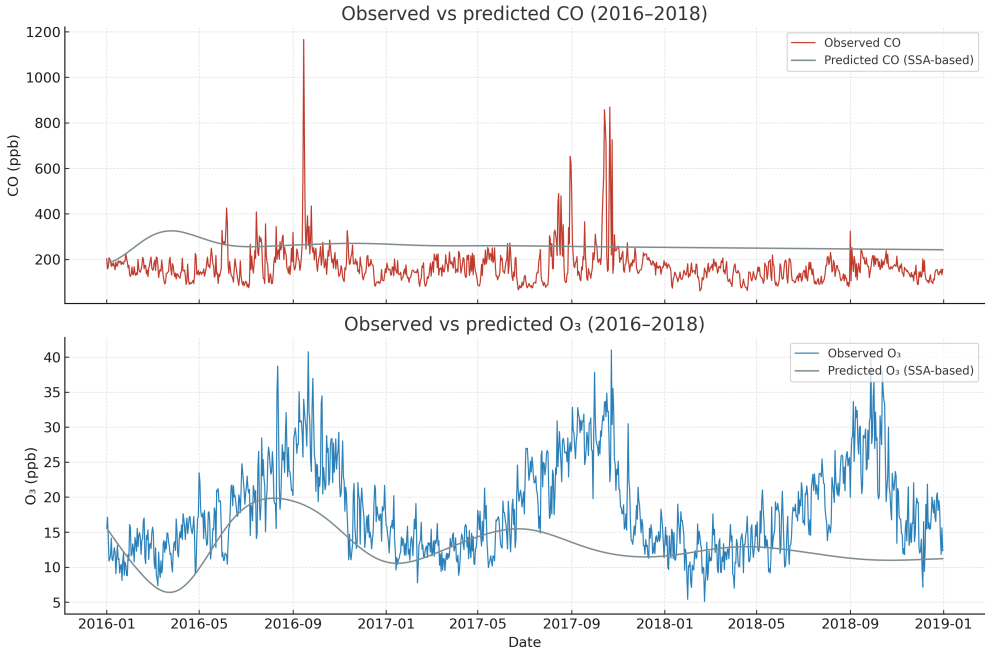


Figure 14. Observed versus SSA-based predicted CO (top) and O_3 (bottom) during 2016–2018.

The SSA-LRF model captures low-frequency structure but not short-term variability. CO forecasts are limited by abrupt fire-driven spikes; O_3 forecasts reflect predictable seasonality but still miss high-frequency meteorological variations.

Table 1 shows performance indicators of the univariate SSA–LRF daily forecasts for CO and O_3 during the validation period (2016–2018). Negative R^2 values indicate limited skill in reproducing day-to-day variability, despite adequate representation of low-frequency seasonal components.

Table 1. SSA-LRF forecasting metrics (2016–2018) for CO and O_3.

Pollutant	MSE	RMSE	MAE	MAPE (%)	R^2
CO	8497.26	92.18	64.41	45.17	−0.18
O_3	60.68	7.79	6.45	43.68	−0.44

Negative R^2 values confirm that univariate SSA is insufficient for short-term prediction of pollutants influenced by episodic events.

4. Discussion

Singular Spectrum Analysis has been increasingly applied to environmental and atmospheric time series due to its ability to extract structured temporal components from noisy and nonstationary data. Previous studies in climatology and hydrology have consistently demonstrated that SSA performs best when the dominant variability is governed by smooth, low-frequency oscillations, such as seasonal and interannual cycles [5,7,15,16]. In the context of air-pollution studies, however, the literature remains relatively limited and often focused on hybrid implementations rather than on the intrinsic behavior of SSA itself.

Several recent studies have combined SSA with statistical or machine-learning models to enhance air-quality forecasting. For instance, hybrid SSA–ARIMA and SSA–SARIMA frameworks have shown improved performance for pollutants exhibiting strong seasonality and moderate variability, particularly when forecasting monthly or seasonal averages [20–22]. These studies generally report satisfactory predictive skill when the target variable is dominated by recurrent cycles, but they also note substantial degradation in performance at daily time scales, especially under the influence of episodic emission events.

The findings of the present study are fully consistent with this body of literature. For O_3, whose temporal behavior is largely controlled by photochemical processes and seasonal meteorological forcing, SSA successfully isolated the dominant annual mode and reproduced its temporal evolution. Similar results were reported by Ferreira et al. [17] and Palacios et al. [18], who applied SSA to atmospheric variables in tropical Brazil and observed that seasonal reconstruction was robust, while short-term variability remained poorly captured.

In contrast, the limited daily forecasting skill obtained for CO aligns with previous observations that SSA tends to smooth extreme events associated with biomass burning, urban plumes, and synoptic-scale transport [21,22]. Studies focusing on fire-affected regions emphasize that pollutants with strong heteroscedasticity and intermittent emission sources require modeling approaches capable of representing nonlinear dynamics and abrupt regime changes. Purely linear SSA reconstruction, even when using optimal window lengths, is inherently constrained in this respect.

Importantly, many published SSA-based air-pollution studies report performance metrics aggregated over longer temporal scales, such as monthly means or seasonal indices, which can mask deficiencies in daily prediction. By explicitly evaluating daily forecasts and reporting negative R^2 values, the present study provides a more transparent and rigorous assessment of SSA limitations, avoiding overestimation of predictive skill. This distinction represents a methodological contribution, as it clarifies that SSA's primary value lies in decomposition, diagnosis, and low-frequency forecasting rather than in operational short-term prediction.

Overall, when compared with existing literature, this study reinforces the view that SSA should be regarded as a foundational tool within a broader modeling framework. Its ability to extract physically interpretable components makes it particularly valuable for preprocessing and feature extraction, which can then be coupled with stochastic or machine-learning models to address high-frequency, event-driven variability. This perspective is increasingly adopted in recent environmental forecasting research and is especially

relevant for tropical regions influenced by biomass burning and strong seasonal atmospheric dynamics.

From a broader methodological perspective, the results obtained in this study highlight important distinctions between decomposition-based approaches and purely predictive models frequently employed in air-quality research. While machine-learning techniques such as neural networks and deep-learning architectures often achieve superior short-term predictive accuracy, their performance is commonly optimized at the expense of physical interpretability. In contrast, SSA provides an explicit representation of the underlying temporal structure of pollutant time series, allowing the identification of dominant seasonal, intra-seasonal, and low-frequency modes that are directly linked to atmospheric processes. This distinction is particularly relevant in tropical regions, where air-pollution variability arises from the interaction between deterministic seasonal forcing and stochastic, event-driven emissions associated with biomass burning. By clarifying the scale-dependent strengths and limitations of SSA, the present study contributes to a more balanced understanding of its role within the spectrum of available modeling tools. Rather than competing with machine-learning approaches, SSA should be viewed as complementary, offering a transparent framework for signal decomposition, noise reduction, and feature extraction. When used as a preprocessing step, SSA can enhance hybrid forecasting systems by providing physically meaningful inputs that improve model stability and interpretability. This perspective aligns with recent trends in environmental data science that emphasize the integration of statistical rigor, physical insight, and predictive performance. Consequently, the findings presented here support the adoption of SSA not only as a forecasting component but also as a diagnostic tool capable of informing model selection, scale-aware analysis, and decision-making in air-quality studies conducted under complex climatic and emission regimes.

5. Conclusions

This study applied Singular Spectrum Analysis (SSA) to long-term daily concentrations of carbon monoxide (CO) and ozone (O_3) in Campo Grande, Brazil, to investigate their multiscale temporal structure and assess the suitability of SSA for air-pollution analysis in a tropical urban environment influenced by biomass burning. By systematically evaluating multiple window lengths (L = 6, 12, 60, and 114), the study demonstrated that larger windows—particularly $L \approx N/2$—provide superior separability between trend, seasonal cycles, and stochastic variability, confirming theoretical recommendations and extending their practical relevance to air-quality time series SSA proved highly effective as an exploratory and decomposition tool, successfully isolating the dominant annual and intra-seasonal cycles of both pollutants. O_3 exhibited relatively smooth and regular seasonal behavior, largely controlled by photochemical production and meteorological conditions, whereas CO displayed stronger variability and pronounced episodic peaks associated with biomass-burning events. The reconstructed SSA components offered a clear and physically interpretable representation of long-term variability and seasonal modulation, supporting the application of SSA in climatological diagnostics and long-term air-quality assessment.

However, the forecasting analysis highlighted important scale-dependent limitations. Although SSA-LRF accurately reproduced low-frequency seasonal behavior, its skill in daily forecasting was limited for both pollutants, as indicated by negative R^2 values during the validation period. This limitation reflects the inherent smoothing nature of linear SSA and its reduced ability to capture short-term variability driven by meteorology, atmospheric transport, and episodic emission events. The effect was particularly evident for CO, whose abrupt fire-related peaks cannot be adequately represented using only a small number of low-frequency components.

These findings emphasize that univariate SSA should not be interpreted as a standalone solution for operational daily air-quality forecasting, especially in regions subject to strong emission intermittency. Instead, SSA should be viewed as a robust preprocessing and decomposition framework, capable of extracting physically meaningful low-frequency components that can be effectively integrated into hybrid modeling approaches. Combining SSA with statistical or machine-learning models—such as ARIMA/SARIMA, GARCH-type models, or deep-learning architectures—represents a promising strategy for improving the representation of nonlinear, event-driven variability in pollutant time series.

From an applied perspective, the identification of per-

sistent seasonal cycles has direct implications for air-quality management in fire-prone tropical regions. Anticipating periods of enhanced O_3 formation and recurrent CO increases during the dry season can support early-warning systems, public health planning, and targeted mitigation strategies. Moreover, the methodological framework developed here is transferable to other tropical and subtropical urban environments affected by biomass burning and strong seasonal atmospheric dynamics.

In summary, this study advances the application of Singular Spectrum Analysis in air-pollution research by providing a rigorous assessment of window-length sensitivity, clarifying the scale-dependent predictive capability of SSA, and demonstrating its value as a foundational tool for hybrid air-quality modeling. Future research should explore multivariate extensions incorporating meteorological drivers and fire-activity indicators, as well as hybrid SSA-based forecasting systems designed to jointly capture seasonal structure and short-term pollution extremes.

In addition, the approach presented here can be readily extended to other tropical and subtropical urban areas where air quality is influenced by seasonal climate variability and biomass burning. The integration of SSA-based decomposition with satellite-derived products and meteorological reanalysis data represents a promising direction for future research aimed at improving both spatial representativeness and predictive capability of air-pollution models.

In addition, the methodological framework proposed in this study can be readily applied to other tropical and subtropical urban regions affected by seasonal climate variability and biomass-burning activity, allowing consistent identification of dominant temporal modes across different environments. The integration of SSA-based decomposition with satellite-derived air-quality products and meteorological reanalysis datasets represents a promising avenue for future research aimed at enhancing spatial representativeness, interpretability, and predictive robustness of air-pollution modeling frameworks.

Author Contributions

Conceptualization: A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. Methodology A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. Validation: A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. Formal analysis writing: A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. Preparation of original draft: A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. Writing - proofreading and editing A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. Visualization: A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. Supervision: A.d.S., R.S.C.N., J.F.d.O.J., I.P., S.V.S.F., and K.R.A.C. All authors read and agreed with the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Air Quality Information System (SISAM), maintained by the National Institute for Space Research (INPE).

Conflicts of Interest

We declare that no conflicts of interest occurred.

References

[1] 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

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

[3] Rahman, M.M., Islam, M.T., Hasan, M.M., et al., 2025. Spatiotemporal Dynamics of Urban Air Pollution in Dhaka Using Sentinel-5P Satellite Data (2020–2024). Environmental Research. 12(8), 274. DOI: https://doi.org/10.3390/environments12080274

[4] Souza, A., Jimenez, J.R.Z., Oliveira Junior, J.F., et al., 2025. Statistical Modeling of PM2.5 Concentrations: Prediction of Extreme Events and Evaluation of Advanced Methods for Air Quality Management. Journal of Atmospheric Science Research. 8(3): 67–92. DOI: https://doi.org/10.30564/jasr.v8i3.10878

[5] Broomhead, D.S., King, G.P., 1986. Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena. 20(2–3), 217–236. DOI: https://doi.org/10.1016/0167-2789(86)90031-X

[6] Vautard, R., Ghil, M., 1989. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Physica D: Nonlinear Phenomena. 35(3), 395–424. DOI: https://doi.org/10.1016/0167-2789(89)90077-8

[7] Elsner, J.B., Tsonis, A.A., 1996. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Springer: New York, NY, USA.

[8] Golyandina, N., Nekrutkin, V., Zhigljavsky, A., 2001. Analysis of Time Series Structure: SSA and Related Techniques. Chapman and Hall/CRC: New York, NY, USA. DOI: https://doi.org/10.1201/9781420035841

[9] Golyandina, N., Zhigljavsky, A., 2020. Singular spectrum analysis for time series. Springer. DOI: https://doi.org/10.1007/978-3-662-62436-4

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

[11] Golyandina, N., Korobeynikov, A., Zhigljavsky, A., 2018. Singular spectrum analysis with R. Springer: Berlin, Germany. DOI: https://doi.org/10.1007/978-3-662-57380-8

[12] Leles, M.C., Sansão, J.P.H., Mozelli, L.A., et al., 2018. Improving reconstruction of time series based on singular spectrum analysis: A segmentation approach. Digital Signal Processing. 77, 63–76. DOI: https://doi.org/10.1016/j.dsp.2017.10.025

[13] Golyandina, N., Shapoval, E., 2023. Forecasting of signals by forecasting linear recurrence relations. Engineering Proceedings. 39(1), 12. DOI: https://doi.org/10.3390/engproc2023039012

[14] Ghil, M., Vautard, R., 1991. Interdecadal oscillations and the warming trend in global temperature time series. Nature. 350, 324–327.

[15] Mahmoudvand, R., 2025. Comparing methods for forecasting time series with multiple observations per period using singular spectrum analysis. Environmental and Ecological Statistics. 32, 175–193. DOI: https://doi.org/10.1007/s10651-025-00648-8

[16] 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.

[17] Ferreira, H.S., Paulo, S.R.d., Paulo, I.J.C.d. et al., 2016. Time series analysis of air temperature and relative humidity in Cuiabá using the Singular Spectrum Analysis method. Brazilian Journal of Climatology. 19, 191–202. Available from: https://revistas.ufpr.br/revistaabclima/article/view/44334/29389 (in Portuguese)

[18] Palacios, M.D., Lopes, F.J.S., Arbilla, G., et al., 2016. Singular spectrum analysis and wavelet analysis of aerosol optical depth time series in the Pantanal of Brazil. Brazilian Journal of Meteorology. 31(4). DOI: https://doi.org/10.1590/0102-778631231420150104 (in Portuguese)

[19] Yiou, P., Baert, E., Loutre, M.F., 1996. Spectral analysis of climate data. Surveys in Geophysics. 17(6), 619–663.

[20] Duan, J., Gong, Y., Luo, J., et al., 2023. Air-quality prediction based on the ARIMA–CNN–LSTM combination model optimized by dung beetle optimizer. Scientific Reports. 13, 12127. DOI: https://doi.org/10.1038/s41598-023-36620-4

[21] Song, C., Fu, X., 2020. Research on different weight combination in air quality forecasting models. Journal of Cleaner Production. 261, 121169. DOI: https://doi.org/10.1016/j.jclepro.2020.121169

[22] Gilik, A., Ogrenci, A.S., Ozmen, A., 2022. Air-quality prediction using CNN+LSTM-based hybrid deep learning architecture. Environmental Science and Pollution Research. 29, 11920–11938. DOI: https://doi.org/10.1007/s11356-021-16227-w

[23] Golyandina, N., Zvonarev, N., 2024. Information criteria for signal extraction using singular spectrum analysis: White and red noise. Algorithms. 17(9), 395. DOI: https://doi.org/10.3390/a17090395

[24] Sun, M., Li, X., 2017. Window length selection of singular spectrum analysis and application to precipitation time series. Global NEST Journal. 19(2), 306–317. DOI: https://doi.org/10.30955/gnj.002117

[25] Elsner, J.B., 2002. Analysis of time series structure: SSA and related techniques. Journal of the American Statistical Association. 97(460), 1207–1208. DOI: https://doi.org/10.1198/jasa.2002.s239

[26] Wang, M., Zhao, G., Liang, W., et al., 2023. A comparative study on the development of hybrid SSA–RF and PSO–RF models for predicting the uniaxial compressive strength of rocks. Case Studies in Construction Materials. 18, e02191. DOI: https://doi.org/10.1016/j.cscm.2023.e02191

[27] Zhao, C., Sun, Y., Zhong, Y., et al., 2021. Spatiotemporal analysis of urban air pollutants throughout China during 2014–2019. Air Quality, Atmosphere and Health. 14(10), 1619–1632. DOI: https://doi.org/10.1007/s11869-021-01043-5