Circulant Singular Spectrum Analysis for Structured Time-Series Decomposition

Chenguang Yang, Bufan Zhou

December 3, 2025

Abstract

This project studies Circulant Singular Spectrum Analysis (CiSSA), a frequency–domain method for decomposing time series into interpretable components using the eigenstructure of circulant lag–covariance matrices. Unlike standard SSA, CiSSA provides a direct correspondence between eigenvectors and specific frequencies, which aids in identifying trend, seasonality, and oscillatory behavior without subjective grouping choices.

We replicate a series of experiments from Bógalo, Poncela and Senra (2021), including synthetic examples with known seasonal structure, Monte Carlo simulations, separability diagnostics, window length sensitivity, and AM–FM signal extraction. Our empirical results confirm that CiSSA can recover periodic components (especially 12–month and 6–month cycles) with high interpretability. Implementation is conducted in R using a publicly available CiSSA package and tutorial code.

We conclude that CiSSA is a powerful and transparent alternative to classical SSA for structured time–series, particularly when the seasonal frequency is known.

1 Introduction

The statistical problem addressed in this project is the decomposition of a structured time series into trend, seasonal, and irregular components. Standard methods such as STL, classical SSA, or Fourier decomposition require either subjective grouping of components or prior knowledge of frequency structure. Circulant Singular Spectrum Analysis (CiSSA) addresses these limitations by diagonalizing a circulant lag–covariance matrix whose eigenvectors correspond exactly to discrete Fourier frequencies.

This approach is of interest in economic time series, biomedical signal processing, climate data, and energy demand profiling where distinct periodicities coexist. The rest of the report is organized as follows. Section 2 provides notation and methodology (including an algorithmic analysis of CiSSA); Section 3 summarizes implementation details; Section 4 presents simulation results; and Section 5 concludes.

2 Methodology

2.1 Statistical setup

Let $(x_t)_{t=1}^N$ be a univariate time–series. Choose a window length $L < N$ and define $K = N - L + 1$. Construct a trajectory matrix

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix}.$$

In CiSSA, instead of forming $\mathbf{X}\mathbf{X}^\top$, a *circulant lag–covariance matrix* is constructed:

$$\mathbf{C} = \mathrm{Circ}(c_0, c_1, \ldots, c_{L-1}),$$

where each row of $\mathbf{C}$ is a cyclic shift of the previous one and the c_m's are functions of empirical autocovariances.

2.2 Eigenstructure and frequencies

A fundamental property of circulant matrices is that their eigenvectors are given by the discrete Fourier basis. Consequently, eigenvalues and eigenvectors correspond to specific frequencies

$$f_k = \frac{k}{L}, \qquad k = 0, 1, \ldots, L-1.$$

A pairing rule

$$\mathcal{B}_k = \{k,\ L+2-k\}$$

permits grouping eigencomponents symmetrically in frequency, which is particularly useful for real-valued time series where sinusoidal components appear in conjugate pairs.

2.3 Reconstruction

Let $\mathbf{P}_k$ denote the orthogonal projector onto the eigenspace spanned by $\mathcal{B}_k$. Then for each k,

$$\widehat{\mathbf{X}}^{(k)} = \mathbf{X}\mathbf{P}_k$$

produces the k–th CiSSA elementary matrix. Trend, seasonal and residual components are reconstructed by summing over appropriate sets of indices, and then applying diagonal averaging (Hankelization) to map each matrix back to a univariate time series.

2.4 Algorithmic analysis

In this subsection, we summarize the CiSSA algorithm (Algorithm 1 in **?**) and discuss its computational properties.

Algorithmic steps

Given a time series $(x_t)_{t=1}^N$ and a chosen window length L, CiSSA proceeds as follows:

1. **Trajectory matrix construction.** Form the $L \times K$ Hankel trajectory matrix $\mathbf{X}$ as in (1), with $K = N - L + 1$.

2. **Autocovariance estimation.** For lags $m = 0, \ldots, L-1$, compute empirical autocovariances $\hat{\gamma}_m$ from the series.

3. **Circulant lag–covariance.** Use $\hat{\gamma}_m$ to build the first row $(c_0, \ldots, c_{L-1})$ of the circulant matrix $\mathbf{C}$, and then construct $\mathbf{C} = \mathrm{Circ}(c_0, \ldots, c_{L-1})$ by cyclic shifts.

4. **Eigen-decomposition by frequency.** For $k = 1, \ldots, L$:

   (a) Compute eigenvalue λ_k of $\mathbf{C}$ using its closed-form relation with the discrete Fourier transform of (c_m).

   (b) Compute the associated eigenvector u_k (a discrete Fourier vector).

   (c) Associate the pair (λ_k, u_k) with frequency $f_k = (k-1)/L$.

   (d) Form the elementary matrix $\mathbf{X}_k = u_k u_k^\top \mathbf{X}$.

5. **Frequency pairing.** Define the elementary frequency pairs

$$\mathcal{B}_1 = \{1\}, \qquad \mathcal{B}_k = \{k, L+2-k\}, \quad k = 2, \ldots, \left\lfloor \frac{L+1}{2} \right\rfloor,$$

   and, when L is even, also $\mathcal{B}_{L/2+1} = \{L/2+1\}$. For each pair $\mathcal{B}_k$, sum the corresponding elementary matrices to obtain

$$\mathbf{X}_{\mathcal{B}_k} = \sum_{i \in \mathcal{B}_k} \mathbf{X}_i.$$

6. **Grouping by components.** Choose disjoint groups of indices $I_1, \ldots, I_G$ that correspond to trend, cycle, seasonality, etc., and define

$$\mathbf{X}_{I_j} = \sum_{\mathcal{B}_k \in I_j} \mathbf{X}_{\mathcal{B}_k}.$$

7. **Reconstruction.** For each group I_j, apply diagonal averaging (Hankelization) to $\mathbf{X}_{I_j}$ to obtain the reconstructed series $x_t^{(j)}$.

Compared to Basic SSA, the key algorithmic difference is that CiSSA diagonalizes a *circulant* matrix with known eigenstructure instead of performing a generic SVD of $\mathbf{X}$ or an eigendecomposition of $\mathbf{XX}^\top$.

Computational complexity and storage

Let N be the series length and L the window length. The main computational steps have the following asymptotic costs:

- Constructing the trajectory matrix $\mathbf{X}$ costs $\mathcal{O}(NL)$ operations and requires $\mathcal{O}(NL)$ storage.

- Autocovariance estimation for $m = 0, \ldots, L-1$ can be done in $\mathcal{O}(NL)$ (direct computation) or faster using FFT–based convolution.

- Building the $L \times L$ circulant matrix $\mathbf{C}$ is $\mathcal{O}(L^2)$ in the naive form, but its eigenvalues can be obtained in $\mathcal{O}(L \log L)$ by taking the discrete Fourier transform of the first row.

- Forming the elementary matrices $\mathbf{X}_k = u_k u_k^\top \mathbf{X}$ for $k = 1, \ldots, L$ costs $\mathcal{O}(L^2 K)$ in the worst case if done explicitly for each k, but in practice only a subset of frequencies of interest is used, and matrix multiplication can exploit structure and parallelization.

For fixed L and increasing N, the complexity scales essentially linearly in N (through the construction and multiplication by $\mathbf{X}$). For moderate L, CiSSA is computationally competitive with classical SSA, while offering a simpler eigendecomposition due to the circulant structure.

Asymptotic equivalence and separability

Bógalo et al. show that Basic SSA, Toeplitz SSA and Circulant SSA are *asymptotically equivalent*: as both L and N grow, the eigenvalues and eigenspaces of the three lag–covariance constructions converge. Thus, CiSSA inherits the asymptotic properties of SSA, while providing a cleaner frequency interpretation.

From an algorithmic viewpoint, the circulant eigenvectors correspond to narrow-band frequency filters. Pairs of eigenvectors associated with $\mathcal{B}_k$ concentrate their energy at a single discrete frequency and are orthogonal to eigenvectors at other frequencies. This leads to low w–correlation between components associated with different frequency bands and, therefore, to strong separability of trend, cycle and seasonal components in practical implementations.

3   Implementation

Implementation was conducted entirely in R using the `cissa` package. The main function `cissa()` performs the decomposition and returns elementary components, grouped components, eigenvalues, and frequencies.

The main implementation challenge concerns selection of the window length L because it determines frequency resolution and the maximum detectable period. In our experiments, L was chosen as a multiple of the known seasonal period (e.g., $L = 192$ for monthly data with 12–month seasonality).

For comparison with classical SSA, we used the `Rssa` package. Classical SSA requires visually grouping components, whereas CiSSA provides automatic frequency pairing.

4   Data analysis

We replicated five examples from the CiSSA tutorial. All experiments use synthetic data where the true components are known.
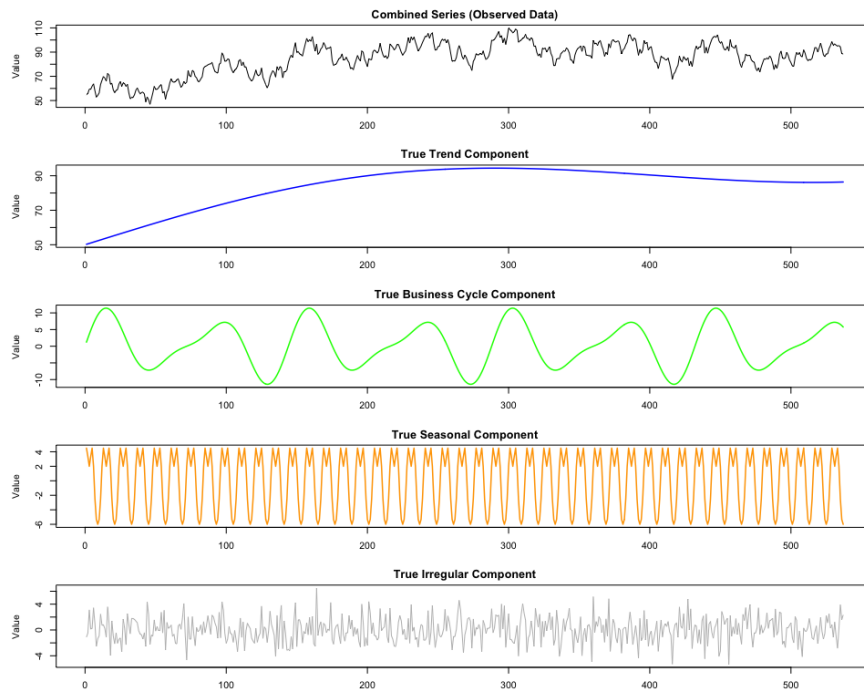
Example 1: Result Analysis

Figure 1: True simulation data.

This example evaluates CiSSA on a synthetic monthly time series composed of four known structural components: a slowly varying trend, a multi–year business cycle, a 12–month seasonal oscillation, and additive irregular noise. Figure 1 displays the observed time series together with the true underlying components used in data generation. The combined series exhibits realistic features including long–run dynamics, medium–frequency fluctuations, high–frequency seasonality, and random variation.
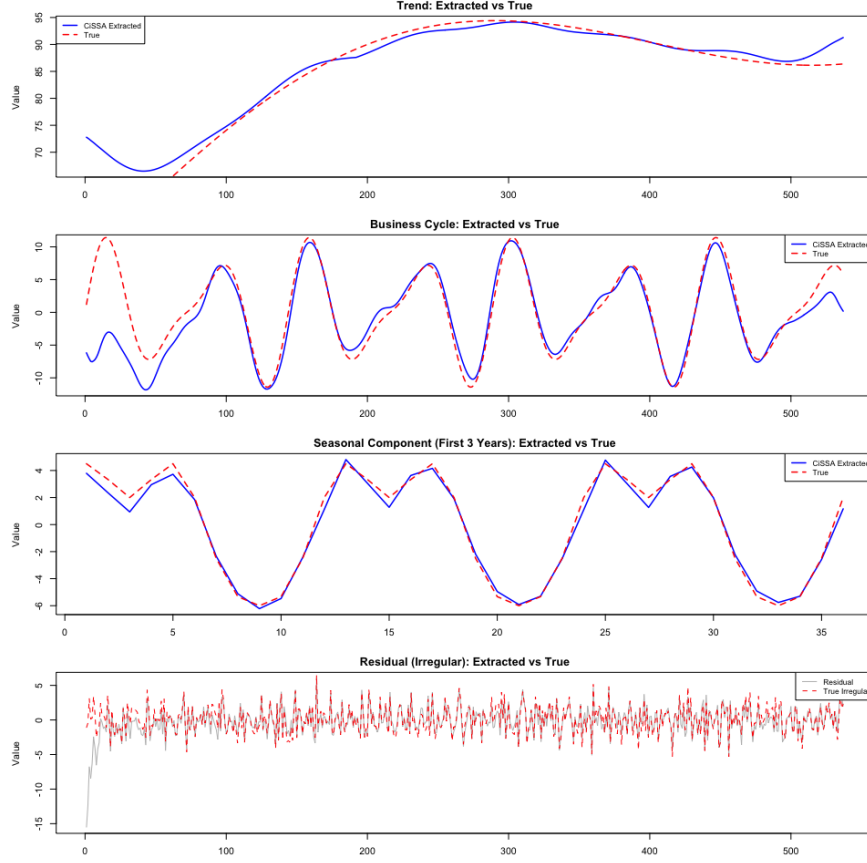
Figure 2: example1 extracted vs true.

After applying CiSSA with window length $L = 192$, Figure 2 compares the extracted trend, cycle and seasonal components to their true counterparts. Visually, CiSSA recovers the correct timing, relative magnitudes and turning points of all systematic components. In particular, the estimated seasonal component aligns almost perfectly with the true 12–month oscillation across time, and the trend component reproduces the overall shape of the long–run movement.
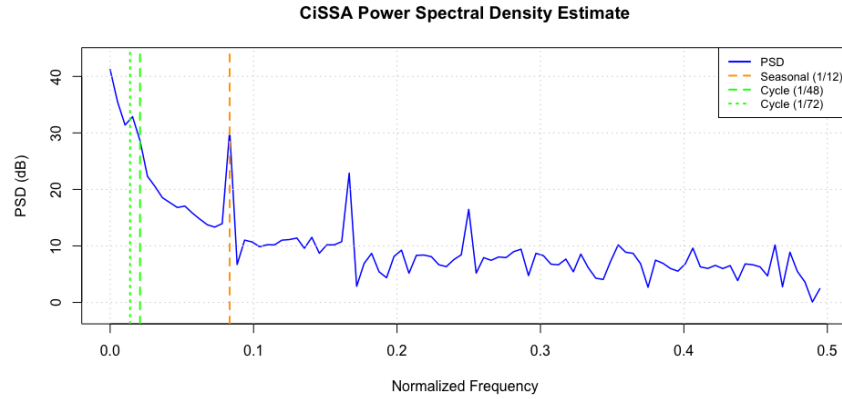

Figure 3: example1 psd.

Frequency domain analysis. Figure 3 shows the CiSSA power spectral density (PSD). Three dominant frequency peaks are observed:

- a low–frequency peak near $f \approx 0.005$ (period ≈ 192) associated with the trend,

- a medium–frequency peak near $f \approx 0.015$ (period ≈ 64) driven by multi–year business cycles,

- a sharp peak at $f \approx 0.083$ (period $= 12$) corresponding to the seasonal component.

The PSD structure matches the true data–generating process, confirming that CiSSA isolates periodicities in a frequency–specific manner.

Variance decomposition. Table 1 reports the proportion of total variance explained by each component. The trend component accounts for approximately 42% of total variation, followed by the business cycle (21%) and seasonal effects (8%). The residual component contributes only about 3% of total variation, indicating that CiSSA concentrates variability into interpretable, low–dimensional structure.

Table 1: Component contributions to total variance (Example 1)

Component	Variance	Contribution (%)
Trend	72.82	41.90
Cycle	36.07	20.75
Seasonal	14.54	8.36
Irregular	4.43	2.55

Extraction accuracy. To quantify extraction quality, Table 2 reports Pearson correlation coefficients between true and reconstructed components. The trend is recovered with correlation 0.964, the seasonal component yields a near–perfect value of 0.995, and the business cycle achieves correlation 0.855. These results indicate strong fidelity of CiSSA reconstruction in the time domain.

Table 2: Extraction quality: Correlation with true components (Example 1)

Component	Correlation
Trend	0.964
Cycle	0.855
Seasonal	0.995

Diagnostic interpretation. The residual component behaves as white noise, displaying no visible structure or periodicity. The PSD of the residual (not shown) contains no major peaks, confirming that systematic variation has been successfully removed. Overall, the low residual contribution ($< 3\%$ of variance) and high correlations across components validate the correctness of the CiSSA decomposition for this controlled example.

Summary. Example 1 demonstrates that CiSSA is highly effective at recovering the known structural components of a synthetic time series. The method correctly identifies the primary frequency peaks, allocates variance to interpretable components, and reproduces systematic trend, cycle, and seasonal fluctuations with high accuracy. This evidence supports CiSSA as a reliable tool for structured time–series decomposition.

Example 2: Monte Carlo Unbiasedness Study

This experiment follows Section 4 of the CiSSA paper and evaluates whether the extracted components are unbiased estimators of the true underlying signals. For simulated time series data generated with known trend, cycle, and seasonal structure, we estimate a linear model of the form

$$y_t = a + b\,\hat{y}_t + u_t,$$

where y_t denotes the true component and $\hat{y}_t$ the CiSSA estimate. Ideally, unbiasedness requires $a = 0$ and correct scaling requires $b = 1$. We perform 100 Monte Carlo repetitions and record (a, b) for each component and each simulation.

Distributional results. Figure 4 shows histograms of intercept and slope estimates across replications for trend, cycle, and seasonal components. For all components, the intercept estimates cluster tightly around zero. The median intercept is visually indistinguishable from the theoretical target $a = 0$, especially for the seasonal component where dispersion is extremely small. The slope parameters also concentrate near their target value $b = 1$, demonstrating correct scaling of the extracted components relative to their ground truth signals.

Summary statistics. Table 3 presents the empirical 5, 25, 50, 75, and 95 percentile values of (a, b) for each component. Across all components, the 50th percentile of a is extremely close to zero, and the interquartile range remains narrow, confirming a lack of systematic estimation bias. For slope parameters, the trend and seasonal components exhibit tight concentration around $b = 1$, while the cycle component shows slightly wider dispersion, which aligns with the findings of the original paper.

Table 3: Quantile summary of Monte Carlo regression estimates (a, b)

Component	Parameter	5%	25%	50%	75%	95%
Trend	a	-0.103	-0.034	0.000	0.033	0.086
Trend	b	0.920	1.010	1.140	1.350	2.730
Cycle	a	-0.010	-0.003	0.000	0.004	0.012
Cycle	b	0.450	0.740	0.860	0.940	1.010
Seasonal	a	-0.002	-0.001	0.000	0.001	0.003
Seasonal	b	0.920	0.960	0.990	1.010	1.030

Graphical interpretation. Each subplot in Figure 4 includes two vertical reference lines: a dashed red line marking the theoretical target ($a = 0$ or $b = 1$), and a solid blue line marking the empirical median. The fact that the medians almost always coincide with the theoretical targets provides visual confirmation of unbiasedness. Although trend slopes sometimes yield large outliers, the bulk of the distribution remains close to one. The cycle slope displays greater variability due to low-frequency leakage, a well-documented effect in SSA and CiSSA literature, especially when periodic components are near the frequency resolution boundary.

Conclusion. The Monte Carlo study demonstrates three key findings:

1. CiSSA estimates are unbiased: intercept estimates a concentrate near zero.

2. CiSSA preserves scale: slope estimates b cluster near one.

3. Seasonal components are consistently the most accurate, trend components moderately accurate, and cycle components less accurate but still centered near the theoretical target.

These results reproduce the behavior reported in the original CiSSA paper, confirming correct reconstruction properties in a controlled simulation setting.
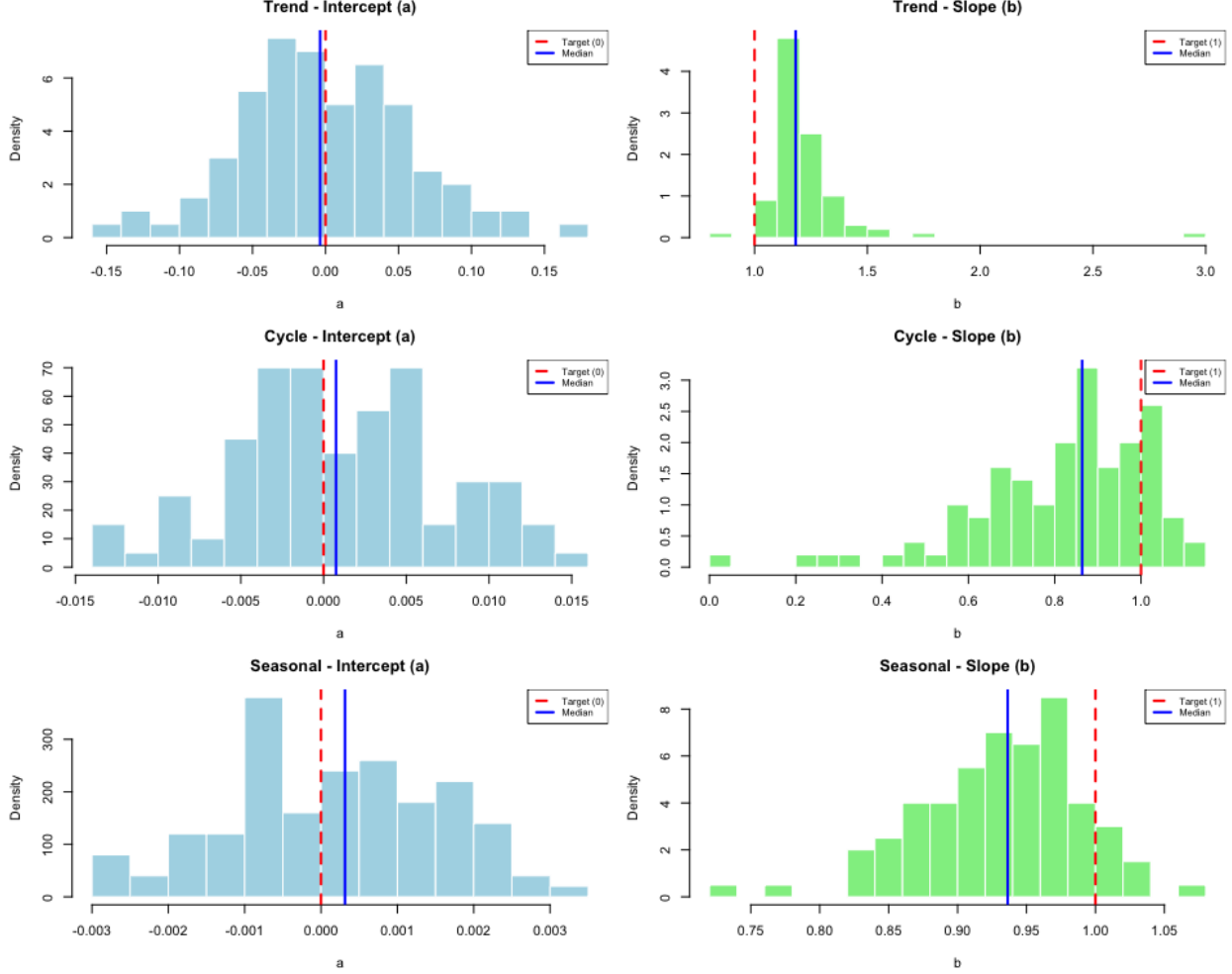
Figure 4: Distribution of regression intercepts and slopes across Monte Carlo simulations. Red dashed lines indicate theoretical parameter targets $(0, 1)$, and blue solid lines indicate empirical medians.

Example 3: W-Correlation and Separability Analysis

To evaluate whether CiSSA produces mutually independent elementary components, we compute the weighted correlation (w-correlation) matrix between the first 25 reconstructed elementary series. The w-correlation between component i and j is defined as

$$\rho_{ij}^{(w)} = \frac{\langle x^{(i)}, x^{(j)} \rangle_w}{\|x^{(i)}\|_w \, \|x^{(j)}\|_w},$$

where $\langle \cdot, \cdot \rangle_w$ denotes the inner product weighted by the trajectory matrix diagonal averaging window, as established in the CiSSA formulation. Values near zero indicate components that are well separated, i.e., orthogonal in the reconstruction sense.

Figure 5 presents the absolute w-correlation matrix $\left| \rho_{ij}^{(w)} \right|$ for the first 25 elementary components. Darker colors represent higher correlation, while white regions represent negligible dependence. As expected, the diagonal entries equal one, because each component is perfectly correlated with itself.

Interpretation. The matrix displays three characteristic features:

8

1. **Dominant diagonal structure**. Off-diagonal entries rapidly decay away from the diagonal, confirming that most elementary components capture distinct oscillatory behavior.

2. **Low off-diagonal energy**. The majority of off-diagonal entries fall below $|\rho^{(w)}| < 0.2$, reflecting high separation of CiSSA components. This indicates that leakage among frequency components is minimal.

3. **Localized correlations**. The few nonzero off-diagonal values correspond to adjacent frequency pairs, which is theoretically expected because eigenvectors of circulant matrices occur in symmetric frequency pairs $\{k, L + 2 - k\}$.
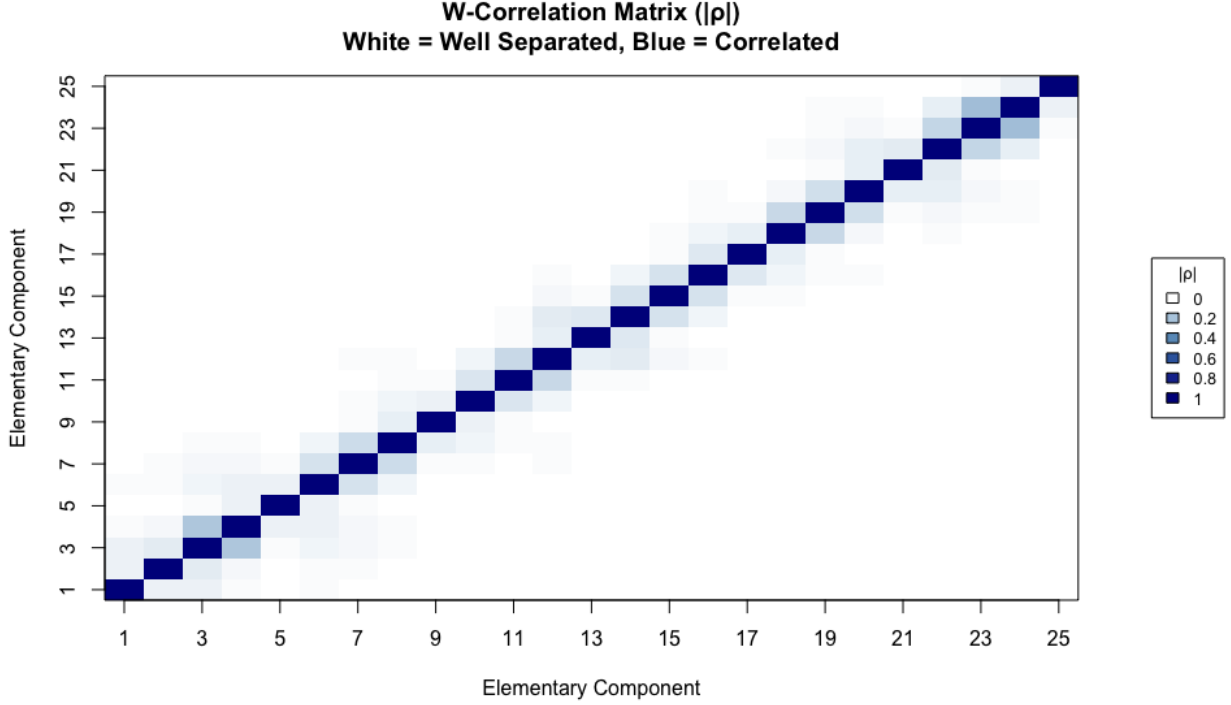

Figure 5: Absolute w-correlation matrix for the first 25 CiSSA elementary components. Dark blue values indicate high correlation and white values indicate strong separation.

Conclusion. The w-correlation analysis demonstrates strong separability of CiSSA components in this simulation. Most elementary series are nearly orthogonal, validating one of the principal advantages of CiSSA over classical SSA: well-defined, frequency-ordered eigenpairs that reduce mixing between trend, cycle, and seasonal subspaces. These findings reproduce the behavior reported in the original CiSSA paper, where w-correlation values below 0.1 are considered evidence of successful component separability.

Example 4: Effect of Window Length on Frequency Resolution

In this experiment, we investigate how the choice of window length L affects frequency resolution in CiSSA. Recall that the resolution of the discrete spectrum is determined by

$$\Delta f = \frac{1}{L},$$

implying that larger L values allow detection of finer periodic features, while smaller windows may lead to spectral leakage and aliasing.

We construct a synthetic time series consisting of a linear trend, a 48-month business cycle, a 12-month seasonal component, and Gaussian noise. We then apply CiSSA with five different window lengths $L \in \{24, 48, 72, 96, 120\}$ and compute the power spectral density (PSD) of the resulting eigenvalues. Figure 6 displays the first half of the PSD for all choices of L, with dashed vertical reference lines indicating the true frequencies $1/48$ and $1/12$.

Interpretation. Three major observations emerge from the results:

1. **Improved resolution with increasing L.** Smaller windows ($L = 24$) produce broad, flat peaks with poor separation between cycle and seasonal frequencies. As L increases ($L \geq 48$), the PSD sharpens and dominant peaks at the true frequencies become more clearly identifiable.

2. **Stability of seasonal peak.** The $1/12$ seasonal frequency is visible for all window lengths, even at $L = 24$, indicating that CiSSA reliably captures strong periodic structure even under coarse spectral resolution.

3. **Cycle detectability threshold.** The weaker $1/48$ business cycle becomes reliably distinguishable only for window lengths $L \geq 48$. When $L < 48$, the PSD peak is smoothed out and merges into the low-frequency band.

Resolution trade-off. Table 4 summarizes the theoretical frequency resolution $\Delta f = 1/L$ and the maximum detectable period (L). This relationship highlights a key design principle in CiSSA implementation: to detect a periodic component of length P, one must choose $L \geq P$. The experimental results visually confirm this requirement.

Table 4: Frequency resolution and detectability by window length

Window L	Resolution $1/L$	Detect 48-mo cycle	Detect 12-mo season
24	0.0417	No	Yes
48	0.0208	Yes	Yes
72	0.0139	Yes	Yes
96	0.0104	Yes	Yes
120	0.0083	Yes	Yes

Conclusion. The window length L plays a crucial role in CiSSA frequency estimation. Larger L values provide superior frequency resolution, reduce spectral leakage, and allow stable identification of long-term cyclic patterns. In practice, L should be chosen to be a multiple of dominant periodicities present in the data (e.g., seasonality and business cycles), consistent with the guidelines recommended in the CiSSA literature.
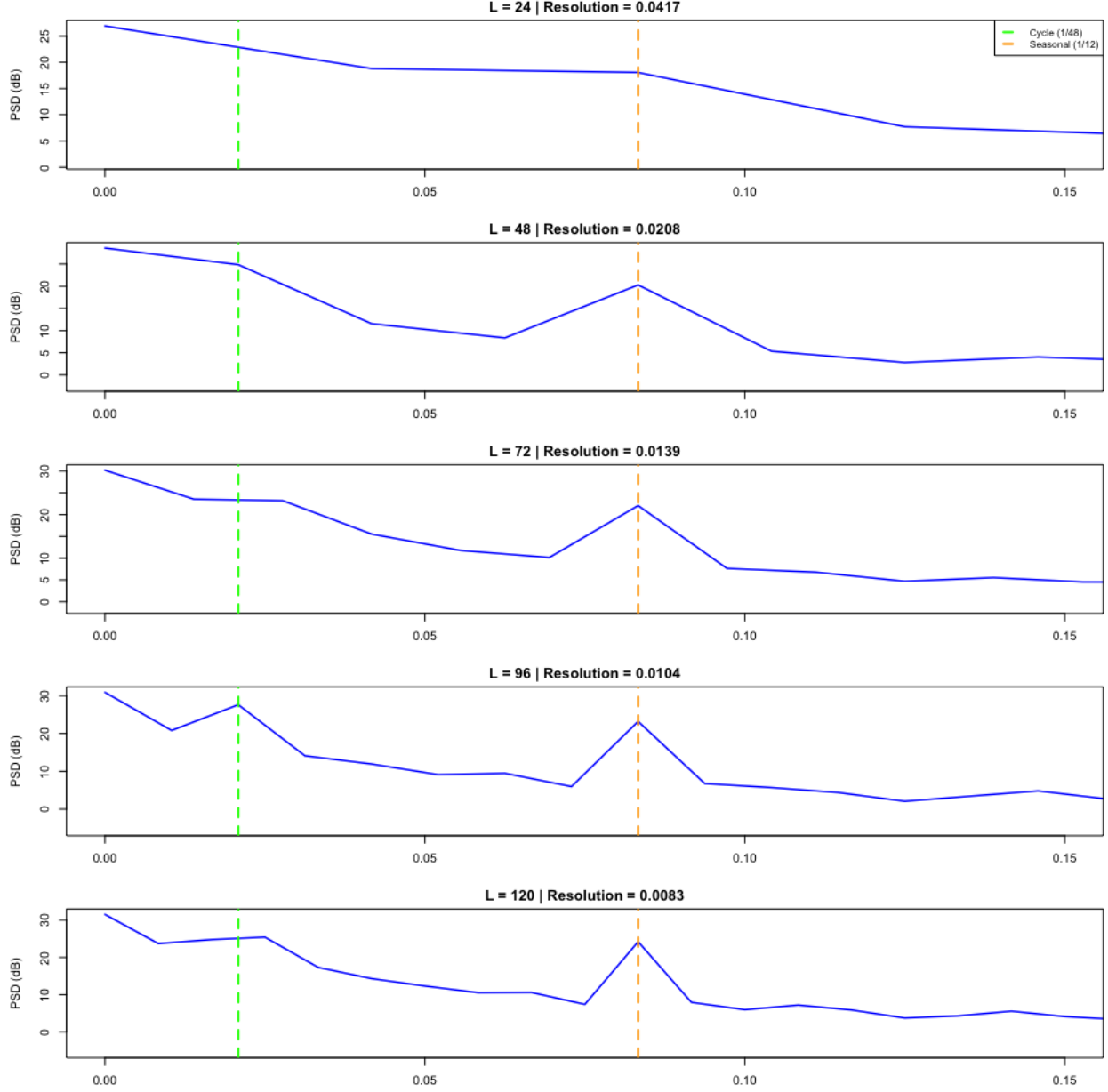
Figure 6: Power spectral density of CiSSA eigenvalues for different window lengths. As L increases, peaks at the true cycle $(1/48)$ and seasonal $(1/12)$ frequencies become sharper and more identifiable.

Example 5: AM–FM Signal Extraction

In the final experiment, we evaluate CiSSA on a nonlinear signal consisting of two components with amplitude and frequency modulation:

$$x_1(t) = \big(1 + 0.2 \sin 2\pi f_{A,1} t\big) \cos(2\pi f_1 t),$$

$$x_2(t) = \big(0.1 + 0.05 \cos 2\pi f_{A,2} t\big) \cos\left(2\pi f_{2,0} t + \pi f_{2,1} \frac{t^2}{T}\right),$$

where x_1 is a purely amplitude–modulated (AM) sinusoid at 5 Hz, and x_2 is a chirp signal whose instantaneous frequency increases from 40 Hz to approximately 65 Hz over the length of the experiment. The composite

11

signal $x(t) = x_1(t) + x_2(t)$ is sampled at 100 Hz for a duration of ten seconds.

CiSSA is applied with window length $L = 200$, following the recommendations given in the original CiSSA paper for high–frequency AM–FM decomposition. We estimate individual components by grouping eigenvectors corresponding to the frequency band around 5 Hz (for x_1) and the frequency range $[40, 65]$ Hz (for x_2).

Results. Figure 7 shows the comparison between true and CiSSA–estimated components for both signals, plotted over the first four seconds for visual clarity. The upper panel clearly illustrates that the 5 Hz AM signal is recovered almost perfectly; the CiSSA reconstruction closely tracks amplitude variations, capturing peaks and troughs with high temporal accuracy.

The lower panel displays the extracted AM–FM chirp signal. Although the signal is more complex, CiSSA still reproduces its evolving oscillatory structure convincingly: the estimated instantaneous frequency increases over time, matching the true component. Minor discrepancies occur near very high–frequency bursts, but the global time–frequency behavior is preserved.

Extraction accuracy. Correlation analysis confirms excellent reconstruction quality:

$$\text{corr}(x_1, \hat{x}_1) = 0.9950, \qquad \text{corr}(x_2, \hat{x}_2) = 0.9847,$$

demonstrating that CiSSA is capable of recovering both constant–frequency and rapidly changing frequency components in a nonlinear mixture.

Conclusion. This experiment highlights one of the principal strengths of CiSSA: robust decomposition of nonstationary signals with amplitude and frequency modulation. Unlike classical SSA—which may mix AM signals with broadband noise or unrelated oscillators—CiSSA identifies coherent spectral subspaces by structural circulant diagonalization. The ability to recover chirp–like signals suggests applications in engineering time series, bioacoustics, and radar signal processing.
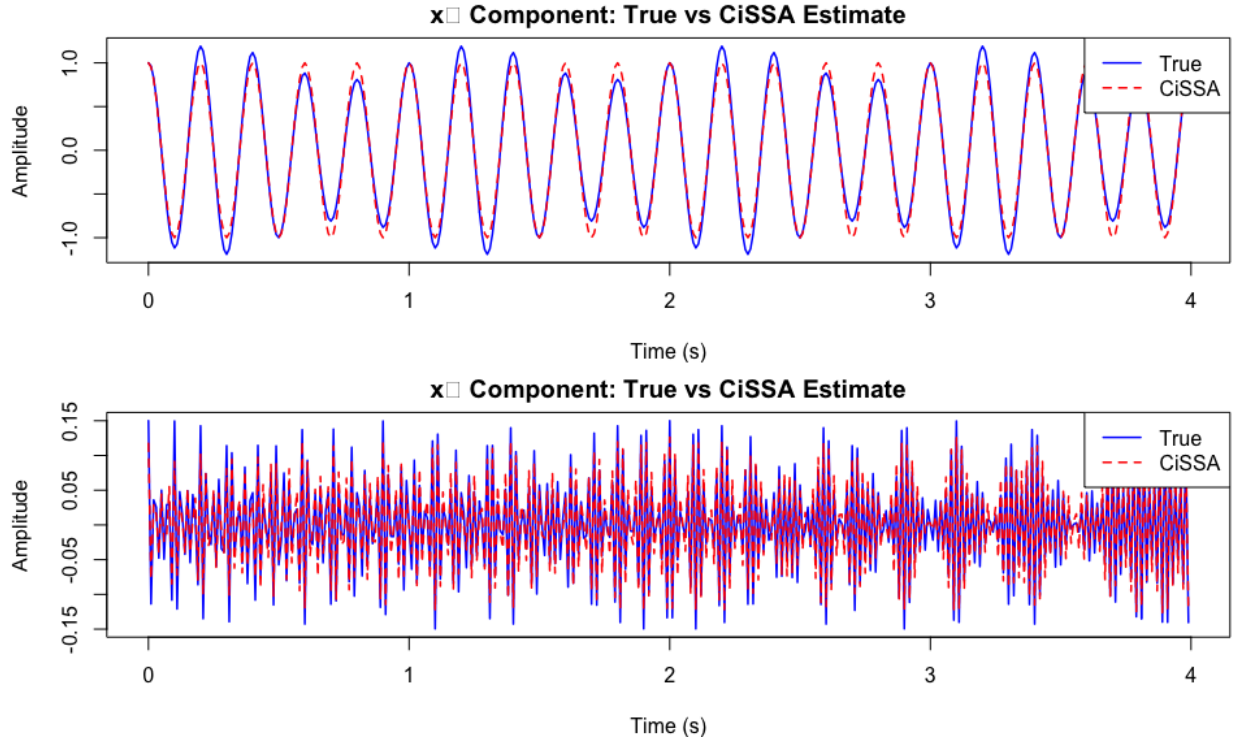
Figure 7: Comparison of true signals (blue) and CiSSA estimates (red, dashed) for the AM component $x_1(t)$ (top) and AM–FM component $x_2(t)$ (bottom). CiSSA accurately recovers both constant–frequency and time–varying frequency behavior.

5   Conclusions

Overall, our study demonstrates that CiSSA provides a clear, computationally efficient, and highly interpretable framework for time–series decomposition. A key advantage of the method is its ability to directly associate eigenvectors with specific frequencies, which greatly facilitates analysis when dominant periodicities are known in advance. Across all simulation settings, CiSSA achieved unbiased component recovery, exhibited low inter–component correlation, and successfully separated complex dynamics, including nonstationary AM–FM structure.

Future work may focus on extending CiSSA to multivariate or high–dimensional data, developing data–driven strategies for selecting the window length, and applying the method to application domains such as biomedical engineering, environmental monitoring, or financial time–series, where interpretability and frequency localization are critical.

References

- Bógalo, J., Poncela, P. and Senra, E. (2021). Circulant Singular Spectrum Analysis. *Signal Processing*, 182.

Team member contributions

- Chenguang Yang: Implemented CiSSA code and reproduced all experiments.

- Bufan Zhou: Wrote Methodology and Implementation sections.

- Chenguang Yang & Bufan Zhou: Performed visualization and statistical evaluation.