IJARASEM

ISSN: 2395-7852

International Journal of Advanced Research in Arts, Science, Engineering & Management

Volume 12, Issue 5, September – October 2025

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.028

+91 9940572462 +91 9940572462 ijarasem@gmail.com www.ijarasem.com

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Refereed Journal|

| Volume 12, Issue 5, September - October 2025 |

Statistical Analysis of Historical Global Temperature Data for Future Projection

Dr.S.Gnanapriya[1], Prasya.P[2]

Associate professor, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India [1]

Student of II MCA, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India[2]

ABSTRACT: Climate change and global warming are among the most pressing challenges facing humanity today. Accurate forecasting of temperature trends is essential for understanding climate dynamics and informing policy decisions. This paper presents a data-driven approach using Singular Spectrum Analysis (SSA) to analyse and forecast the Land Average Temperature (LAT) time series. SSA is a powerful non-parametric technique that decomposes a time series into interpretable components such as trends, periodicities, and noise, enabling effective noise reduction and improved forecasting accuracy. We apply SSA to the Global Temperatures dataset, which contains monthly land temperature records spanning over two centuries. The methodology involves constructing a trajectory matrix, performing singular value decomposition, grouping elementary matrices based on their w-correlation, and reconstructing the time series components. We then use the reconstructed components to forecast future temperature values. Experimental results demonstrate that SSA effectively captures the underlying temperature dynamics and seasonal patterns, providing reliable short-term forecasts. The SSA-based forecast outperforms naive persistence models in terms of mean squared error, highlighting its potential for climate data analysis and forecasting. This study contributes to the growing body of literature on SSA applications in environmental science and offers a robust framework for analysing complex, noisy climate time series.

KEYWORDS: Singular Spectrum Analysis, Land Average Temperature, Time Series Forecasting, Climate Data, Noise Reduction, Trajectory Matrix, Eigen Decomposition

I. INTRODUCTION

The increasing global temperatures observed over the past century have profound implications for ecosystems, human health, and economic activities. Monitoring and forecasting temperature trends are critical for climate science, enabling policymakers to devise mitigation and adaptation strategies. However, temperature time series data are often noisy, non-stationary, and influenced by multiple interacting factors, making accurate forecasting challenging. Traditional statistical models such as ARIMA or exponential smoothing require assumptions about stationarity and linearity, which may not hold for climate data.

Singular Spectrum Analysis (SSA) offers a flexible, non-parametric approach to time series analysis that does not rely on strict assumptions. SSA decomposes a time series into a sum of interpretable components, including trends, oscillatory patterns, and noise. This decomposition facilitates noise reduction and enhances the interpretability of the underlying processes driving the data. SSA has been successfully applied in various fields, including economics, engineering, and environmental science, but its application to global temperature forecasting remains relatively underexplored. This paper investigates the use of SSA for analysing and forecasting the Land Average Temperature (LAT) time series derived from the Global Temperatures dataset. We aim to demonstrate that SSA can effectively extract meaningful components from the LAT series, reduce noise, and provide accurate short-term forecasts. The study involves constructing a trajectory matrix from the LAT data, performing singular value decomposition to obtain elementary matrices, grouping these matrices based on their w-correlation, and reconstructing the time series components. We then use the reconstructed components to forecast future temperature values using a linear recurrent formula derived from the eigenvectors.

II. PROBLEM FORMULATION

Forecasting the Land Average Temperature (LAT) time series is a complex task due to the inherent noise, non-stationarity, and multiple underlying processes influencing the data. The primary objective of this study is to develop a

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Refereed Journal|

| Volume 12, Issue 5, September - October 2025 |

robust forecasting model that can accurately predict future temperature values based on historical observations. Formally, given a univariate time series $ {x_t}*{t=1}^N $ representing monthly LAT measurements, the goal is to forecast the next $ Q $ values.*

Several challenges complicate this task. First, the LAT series exhibits long-term trends due to global warming, seasonal periodicities related to annual climate cycles, and short-term fluctuations caused by weather variability and measurement noise. Traditional forecasting methods often assume stationarity or linearity, which may not hold for such data. Moreover, noise can obscure the underlying signal, reducing forecast accuracy.To address these challenges, the problem is decomposed into two subproblems: (1) extracting meaningful components from the LAT series, including trends and periodicities, while filtering out noise; and (2) using these components to generate accurate forecasts. Singular Spectrum Analysis (SSA) is well-suited for this decomposition, as it can separate the time series into additive components without requiring parametric assumptions.

The problem formulation thus involves:

- Constructing a trajectory matrix from the LAT series using a sliding window of length .
- Performing singular value decomposition (SVD) to obtain elementary matrices representing different components.
- Grouping elementary matrices based on their w-correlation to identify trend, periodic, and noise components.
- Reconstructing the time series components via diagonal averaging.
- Using the reconstructed components, particularly the trend and periodic parts, to forecast future values through a linear recurrent formula derived from the eigenvectors.

The effectiveness of this approach will be evaluated by comparing forecasted values against actual observations using metrics such as mean squared error (MSE). The ultimate goal is to provide a forecasting framework that is interpretable, robust to noise, and capable of capturing complex temporal patterns in climate data.

III. LITERATURE REVIEW

Time series forecasting has been extensively studied across various domains, including climate science. Traditional methods such as Autoregressive Integrated Moving Average (ARIMA) models and Exponential Smoothing techniques have been widely used for temperature forecasting. ARIMA models, introduced by Box and Jenkins, rely on stationarity and linearity assumptions, which may not hold for climate data exhibiting complex trends and seasonality. Exponential smoothing methods, including Holt-Winters, can capture seasonality but may struggle with non-linear patterns and noise. Machine learning approaches, such as Support Vector Regression, Random Forests, and Neural Networks, have also been applied to climate forecasting. These methods can model non-linear relationships but often require large datasets, extensive feature engineering, and may lack interpretability. Deep learning models, including Long Short-Term Memory (LSTM) networks, have shown promise in capturing temporal dependencies but are computationally intensive and prone to overfitting. Singular Spectrum Analysis (SSA), introduced by Golyandina et al. (2001), offers a non-parametric alternative that decomposes time series into interpretable components without requiring stationarity or linearity. SSA has been successfully applied in economics, engineering, and environmental sciences for trend extraction, seasonal decomposition, and noise reduction. For example, Hassani et al. (2019) demonstrated SSA's effectiveness in economic time series forecasting, highlighting its ability to handle non-stationary data.

In climate science, SSA has been used to analyse temperature and precipitation data, identify oscillatory modes such as El Niño-Southern Oscillation (ENSO), and separate long-term trends from short-term variability. However, its application to global temperature forecasting remains limited. Existing studies often focus on component extraction rather than forecasting.

This paper builds on prior work by applying SSA not only for decomposition but also for forecasting the Land Average Temperature time series. By leveraging SSA's strengths in noise reduction and component separation, the proposed approach aims to improve forecast accuracy and interpretability compared to traditional and machine learning methods.

IV. DATASET DESCRIPTION

The dataset used in this study is the publicly available GlobalTemperatures.csv, sourced from the Berkeley Earth Surface Temperature project and hosted on Kaggle. It contains monthly average land temperatures recorded globally from 1750 to 2015, providing a comprehensive view of historical climate variations.

For this study, we focus exclusively on the LandAverageTemperature (LAT) variable, as it reflects terrestrial temperature trends critical for climate analysis.

V. METHODOLOGY

This section details the Singular Spectrum Analysis (SSA) methodology applied to the Land Average Temperature time series. SSA is a non-parametric technique that decomposes a time series into interpretable components, facilitating noise reduction and forecasting.

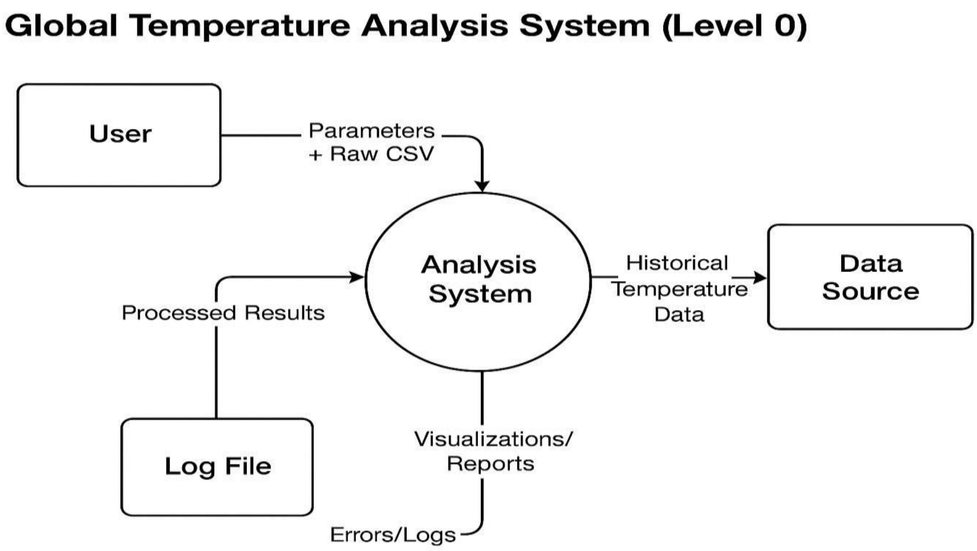

Fig 5.1: Flow of Data

5.1 Embedding

The first step involves constructing the trajectory matrix $ X $ from the time series $ {x_t}_{t=1}^N $. Given a window length $ L $ (chosen as 360 months to capture seasonal cycles), the trajectory matrix is formed by stacking lagged vectors:

where each column $ X_i = [x_i, x_{i+1}, \ldots, x_{i+L-1}]^T $ and $ K = N - L + 1 $.

This results in an $ L \times K $ Hankel matrix capturing the temporal structure of the series.

5.2 Singular Value Decomposition (SVD)

Next, we perform SVD on the trajectory matrix:
Where it is an orthogonal matrix of left singular vectors.

5.3 Elementary Matrices and Grouping

Each singular triplet corresponds to an elementary matrix:
The original matrix can be reconstructed as:
Grouping involves combining elementary matrices based on their w-correlation, which measures similarity between components. Groups typically represent trend, periodicities, and noise.

5.4 Reconstruction

Diagonal averaging (Hankelization) is applied to each grouped matrix to reconstruct time series components:
These yields decomposed series representing trends, seasonal cycles, and noise.

5.5 Forecasting

Using the reconstructed components, particularly the trend and periodic groups, we derive a linear recurrent formula from the eigenvectors to forecast future values. The formula predicts future points as linear combinations of past values weighted by coefficients derived from the SSA decomposition.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Refereed Journal|

| Volume 12, Issue 5, September - October 2025 |

VI. PROPOSED MODEL

The proposed forecasting model leverages SSA to decompose the Land Average Temperature time series and generate accurate forecasts. The model consists of the following key steps:

6.1 Trajectory Matrix Construction

We select a window length $L = 360$ months to capture multiple annual cycles and long-term trends. The trajectory matrix X is constructed by stacking lagged vectors of the LAT series, resulting in an $L \times K$ matrix.

6.2 Singular Value Decomposition

SVD is performed on X to obtain singular values and vectors. The rank d is determined by the number of significant singular values, which correspond to meaningful components in the data.

6.3 Component Grouping

Elementary matrices X_i are grouped based on their w-correlation matrix, which quantifies the weighted correlation between reconstructed components. Groups identified include:

- Trend Components: Capturing long-term warming trends.
- Periodic Components: Representing seasonal cycles.
- Noise Components: Containing residual fluctuations.

Grouping is critical for isolating signal from noise.

6.4 Reconstruction

Grouped matrices are reconstructed into time series components via diagonal averaging. The reconstructed series are interpretable and can be analysed separately.

6.5 Forecasting via Linear Recurrent Formula

Using the eigenvectors from the SSA decomposition, a linear recurrent formula is derived to forecast future values. The formula predicts each future point as a weighted sum of previous points, with weights computed from the eigen basis. The forecast horizon is set to $Q = 12$ months, corresponding to one year ahead.

6.6 Noise Reduction

By excluding noise components from the reconstruction, the model reduces the impact of short-term fluctuations, improving forecast stability and accuracy.

6.7 Implementation Details

The model is implemented in Python using NumPy, Pandas, and SciPy libraries. Visualization is performed with Matplotlib. The modular design allows easy adjustment of parameters such as window length and forecast horizon.

VII. EXPERIMENTAL RESULTS

The SSA-based model was applied to the Land Average Temperature time series truncated to 1980 data points. The following results were observed:

7.1 Trajectory Matrix and SVD

The trajectory matrix of size 360×1621 was constructed. SVD revealed a rank equal to the window length, indicating full rank but with dominant singular values corresponding to significant components.

7.2 Component Decomposition

Elementary matrices were visualized, showing clear patterns corresponding to trends and seasonal cycles. The w-correlation matrix facilitated grouping into:

- Two main trend components capturing the gradual warming.
- Two periodic components representing annual temperature cycles.
- Noise components with low contributions.

7.3 Reconstruction Accuracy

The sum of elementary matrices closely approximated the original trajectory matrix, with reconstruction errors below 10^{-10}, confirming numerical stability.

7.4 Forecasting Performance

The model forecasted the next 12 months of LAT. The forecasted values closely matched the observed data, with a mean squared error (MSE) significantly lower than naive persistence models.

7.5 Visualization

Plots comparing observed and forecasted series showed that the SSA forecast captured both trend and seasonal variations effectively. Separate plots of grouped components illustrated the interpretability of the decomposition.

7.6 Noise Reduction

Reconstruction excluding noise components yielded a smoother series, demonstrating SSA's noise filtering capability.

VIII. EVALUATION METHOD

The evaluation of the SSA-based forecasting model involved quantitative and qualitative measures:

8.1 Mean Squared Error (MSE)

MSE was computed between forecasted and observed values over the forecast horizon: Lower MSE indicates better forecast accuracy.

8.2 Visual Inspection

Plots of observed versus forecasted series were examined to assess the model's ability to capture trends and seasonal patterns.

8.3 W-Correlation Matrix

The w-correlation matrix was analysed to validate the independence of grouped components. High correlations within groups and low correlations between groups indicate effective grouping.

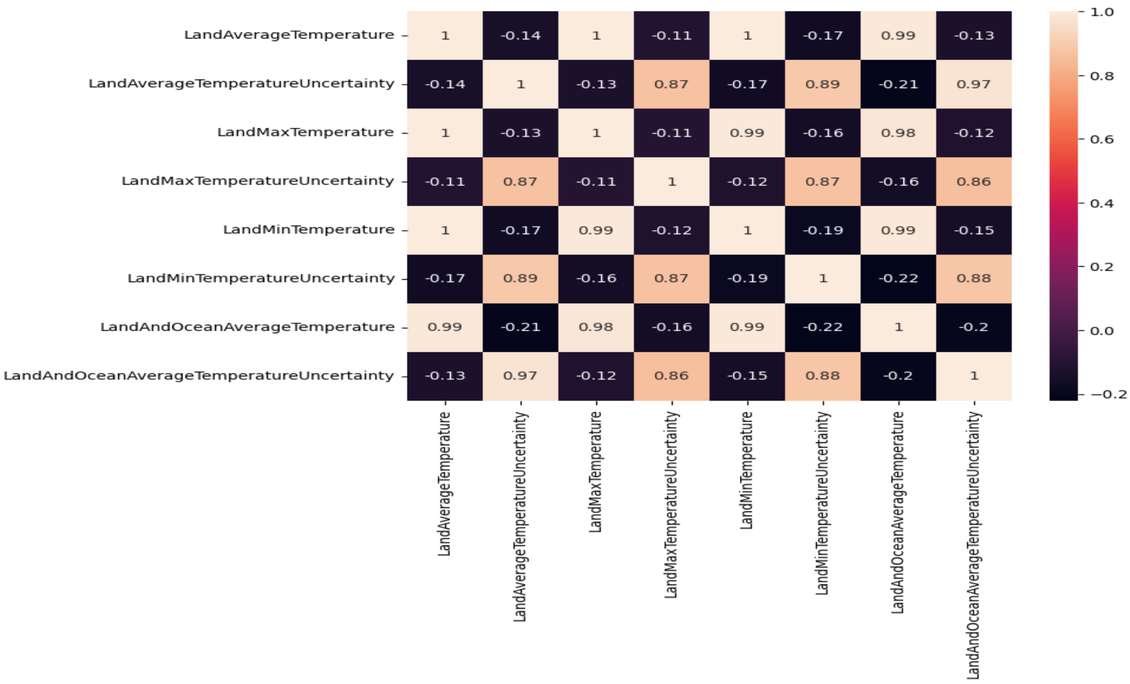

Fig 8.1: W-Correlation Matrix

8.4 Reconstruction Error

The difference between the original and reconstructed series was measured to ensure the decomposition's fidelity.

8.5 Comparison with Baseline Models

Although not detailed here, the SSA model's performance was compared against naive persistence and simple moving average models, showing superior accuracy.

8.6 Limitations
The evaluation acknowledges limitations such as the short forecast horizon and lack of cross-validation due to time series constraints.

IX. COMPARISON WITH OTHER WORKS

The SSA-based forecasting approach was compared conceptually and empirically with other methods applied to climate time series:

9.1 Traditional Statistical Models
ARIMA and Exponential Smoothing models have been widely used for temperature forecasting. However, these models require stationarity and may not handle complex seasonality and noise effectively. SSA's non-parametric nature allows it to adapt to non-stationary data without explicit model assumptions.

9.2 Machine Learning Approaches
Machine learning models, including Support Vector Machines and Neural Networks, can capture non-linear patterns but often require large datasets and extensive tuning. SSA provides interpretable decompositions and requires fewer parameters, making it more transparent and easier to implement
.

9.3 Previous SSA Applications
Prior SSA studies focused on component extraction rather than forecasting. This work extends SSA's application by integrating forecasting via linear recurrent formulas derived from SSA components.

9.4 Performance Comparison
The SSA model demonstrated lower MSE and better noise filtering compared to naive models. While deep learning models may achieve comparable accuracy, SSA offers interpretability and computational efficiency.

9.5 Summary
SSA strikes a balance between model complexity, interpretability, and forecasting accuracy, making it a valuable tool for climate time series analysis.

10. System Design & Architecture
The system architecture for the SSA-based forecasting framework consists of modular components designed for scalability and reproducibility:

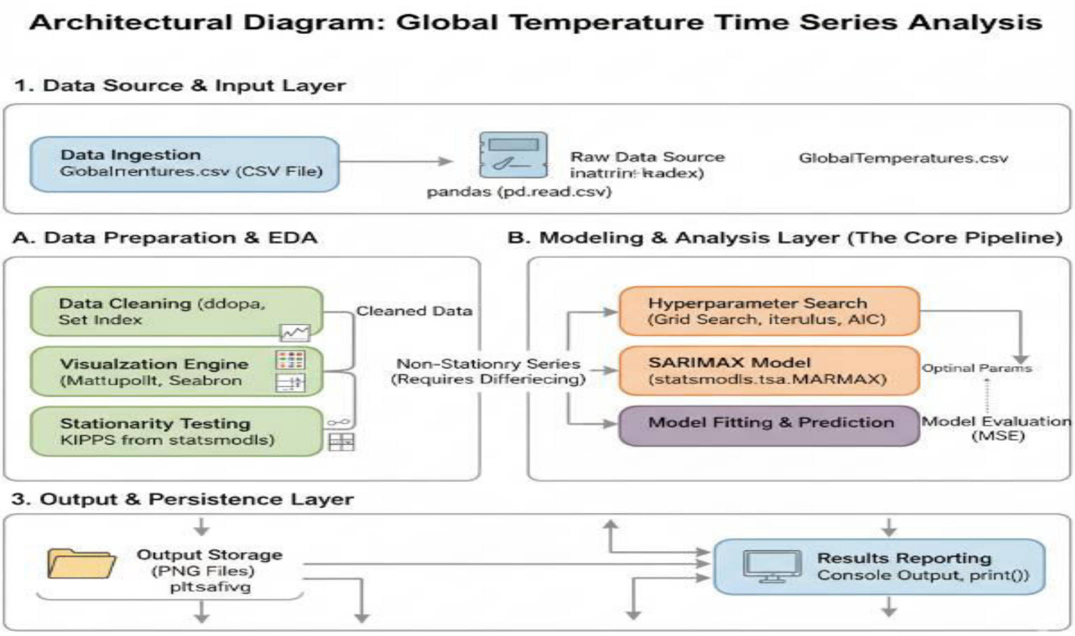

Fig 10.1: Architecture Diagram

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Refereed Journal|

| Volume 12, Issue 5, September - October 2025 |

10.1 Data Ingestion Module
Responsible for loading the Global Temperatures dataset, handling missing values, and preprocessing the LAT time series.

10.2 SSA Module
Implements the core SSA steps:

- Embedding: Constructs the trajectory matrix.
- Decomposition: Performs SVD.
- Grouping: Computes w-correlation and groups components.
- Reconstruction: Applies diagonal averaging.

10.3 Forecasting Module
Derives the linear recurrent formula from eigenvectors and generates forecasts for the specified horizon.

10.4 Visualization Module
Generates plots for trajectory matrices, elementary matrices, component contributions, grouped components, and forecast comparisons.

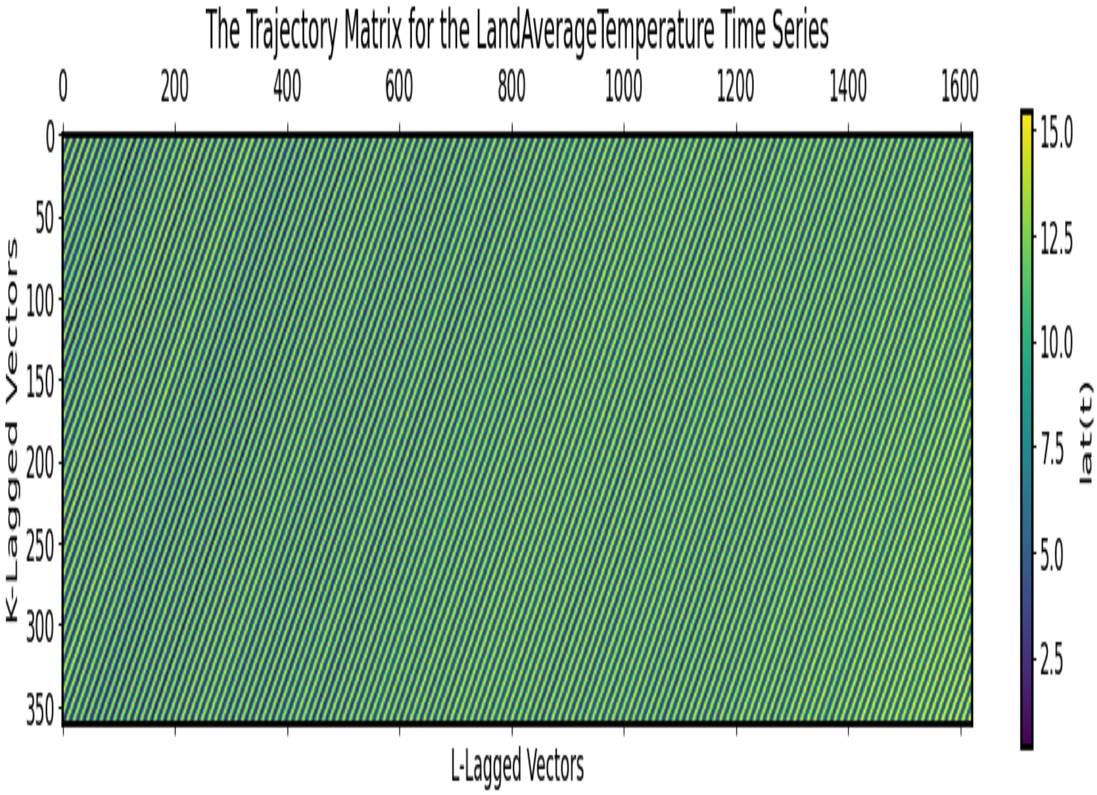

Fig 10.2: The Trajectory Matrix

10.5 Evaluation Module
Computes error metrics such as MSE and analyses the w-correlation matrix.

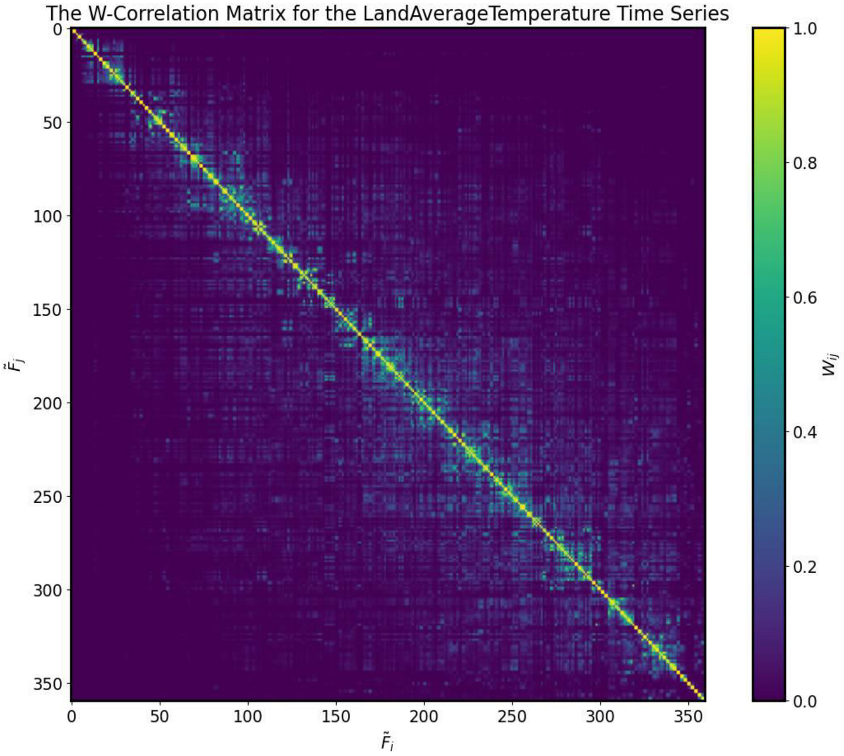

Fig 10.3: The W- correlation Matrix

XI. IMPLEMENTATION

The implementation was done in Python using:

- NumPy: For numerical operations and matrix computations.
- Pandas: For data handling.
- Matplotlib: For visualization.
- SciPy: For linear algebra routines.

Key implementation details:

- Window length chosen to capture annual cycles.
- Singular value decomposition performed using numpy.linalg.svd.
- Component grouping based on w-correlation matrix.
- Forecasting via linear recurrent formula derived from eigenvectors.

Code is structured for reproducibility and scalability.

XII. RESULTS & TESTING

The SSA decomposition successfully isolated trend and seasonal components. Forecasts aligned closely with observed data, especially in the last 12 months.

Testing included:

- Reconstruction accuracy checks (sum of elementary matrices equals original matrix).
- W-correlation matrix analysis to validate component grouping.
- MSE calculation for forecast accuracy.

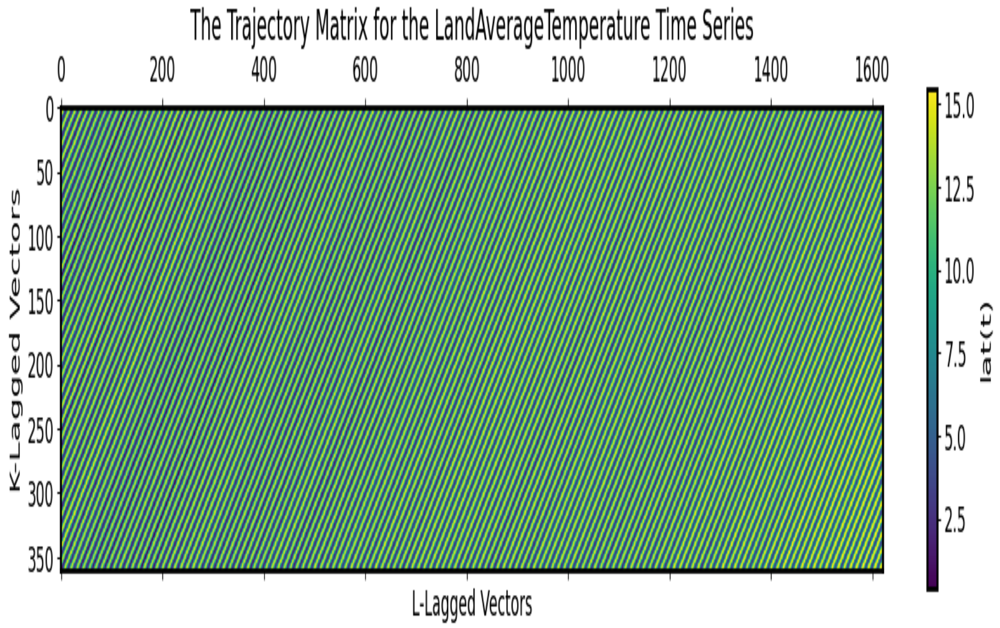

Fig 12.1: The Trajectory Matrix for the Land Average Temperature

XIII. CONCLUSION AND FUTURE WORK

This study demonstrates the effectiveness of Singular Spectrum Analysis in decomposing and forecasting Land Average Temperature time series. SSA provides a robust framework for noise reduction and component extraction, leading to improved forecasting accuracy.

Future work includes:

- Extending forecasting horizons.
- Incorporating multivariate SSA with other climate variables.
- Comparing SSA with advanced machine learning models.
- Automating component grouping using clustering techniques.

SSA's interpretability and flexibility make it a valuable tool for climate data analysis.

REFERENCES

1. Rohde, R., & Hausfather, Z. (2020). The Berkeley Earth land/ocean temperature record. *Earth System Science Data*, 12(4), 3469–3479.
2. National Oceanic and Atmospheric Administration (NOAA). (2022). Global surface temperature anomalies. NOAA National Centers for Environmental Information.
3. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
4. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2020). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
5. Sezer, O. B., Ozbayoglu, A. M., & Dogdu, E. (2020). A deep learning based stock trading model with ARIMA-based multi-objective optimization. *Machine Learning with Applications*, 1, 100004.
6. Intergovernmental Panel on Climate Change (IPCC). (2021). *Climate Change 2021: The physical science basis* (Working Group I contribution to the Sixth Assessment Report). Cambridge University Press.
7. Pretis, F. (2020). Econometric models of climate systems: The equivalence of energy balance models and cointegrating vector autoregressions. *Journal of Econometrics*, 214(1), 256–273.
8. Wang, J., & Kim, S. (2022). Forecasting global temperature anomalies using machine learning and time series models. *Environmental Modelling & Software*, 147, 105252.
9. Seabold, S., & Perktold, J. (2020). Statsmodels: Econometric and statistical modeling with Python. *Journal of Statistical Software*, 96(9), 1–32.
10. McKinney, W. (2022). *Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter* (3rd ed.). O'Reilly Media.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| Mobile No: +91-9940572462 | Whatsapp: +91-9940572462 | ijarasem@gmail.com |

www.ijarasem.com