BAREKENG: Journal of Mathematics and Its Applications
December 2026 Volume 20 Issue 4 Page 3573-3586
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss4pp3573-3586

FORECASTING OIL PRODUCTION USING SSA AND TREND REGRESSION IN THE WORLD'S TOP THREE OIL PRODUCERS

Ega Saherti [1], **Salwa Azzah Imtiyaz** [2*], **Gumgum Darmawan** [3], **Budi Nurani Ruchjana** [4]

[1,2,3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran*
[4]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran*
Jln. Raya Bandung Sumedang. KM.21, Jatinangor District, Sumedang City, 45363, Jawa Barat, Indonesia

*Corresponding author's e-mail: * **salwa.tiyas@gmail.com***

Article Info

Article History:
Received: 30th November 2025
Revised: 02nd May 2026
Accepted: 07th June 2026
Available online: 01st July 2026

Keywords:
Forescasting.
Global oil production;
Singular Spectrum Analysis (SSA);
Trend regression.

ABSTRACT

The world's primary energy source, which plays a significant role in various global sectors, is petroleum. Due to geological influences, energy policies, and geopolitical factors, oil production patterns are often fluctuating and non-stationary. This also applies to the world's of top three oil-producing countries, that is, the United States, Saudi Arabia, and Iraq. This situation poses a challenge in developing accurate forecasting models to support global energy planning. This study aims to forecast oil production trends in the United States, Saudi Arabia, and Iraq using the Singular Spectrum Analysis (SSA) method combined with Trend Regression to obtain a forecasting model capable of capturing long-term patterns and mitigating the influence of short-term fluctuations. Annual oil production data for the period 1936–2024 were taken from Our World in Data. The analysis stages include data decomposition using SSA to separate trends, noise, and cycles, followed by trend component modeling using trend regression. Model evaluation was carried out using the coefficient of determination (R^2). The results of the study indicate that the SSA and Trend Regression methods are able to produce stable and accurate projections, with the highest R^2 value in Saudi Arabia (0.98), followed by Iraq (0.75), and the United States (0.48). All three countries show an increasing production trend until 2034 with different patterns. The SSA and Trend Regression methods are effective in capturing the complex and non-stationary dynamics of oil production. This study provides both academic and practical contributions in the application of the SSA–Trend Regression hybrid method for global oil production forecasting as well as practical contributions for policymakers in projecting global oil production trends.

How to cite this article:

E. Saherti, S. A. Imtiyaz, G. Darmawan, dan B. N. Ruchjana., "FORECASTING OIL PRODUCTION USING SSA AND TREND REGRESSION IN THE WORLD'S TOP THREE OIL PRODUCERS", *BAREKENG: J. Math. & App.*, vol. 20, no. 4, pp. 3573-3586, Dec,

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Oil is essential for humans in various aspects of life. This natural resource is a primary raw material in the energy industry, transportation, and even in various everyday products [1]. As technology advances, the need for oil continues to increase. Therefore, countries worldwide are competing to discover and manage their oil reserves. Without oil, many economic sectors would be hampered, from transportation to the production of consumer goods. The use of oil in daily life is diverse and almost unavoidable in modern society. One of its most important uses is as a vehicle fuel, which supports human mobility around the world. Furthermore, oil is also used in the production of cooking gas and in the medical sector, such as in the manufacture of medicines and medical devices [2]. Without oil, many essential facilities, such as hospitals and factories, would not be able to operate efficiently. Therefore, a stable oil supply is crucial for the global economy.

However, modeling oil production remains a significant challenge. Previous studies utilizing pure ARIMA models often struggle with the non-stationary and highly volatile nature of energy data, as they rely heavily on linearity assumptions [3], [4]. While Singular Spectrum Analysis (SSA) has emerged as a powerful non-parametric tool for de-noising and signal extraction [5], [6], direct forecasting using pure SSA, such as recurrent or vector forecasting, can be overly sensitive to localized fluctuations or missing values, potentially leading to instability in long-term projections [7], [8].

The integration of SSA and Trend Regression proposed in this study is theoretically superior because it leverages the strength of SSA in decomposing complex, noisy signals into interpretable components [5], [9], while utilizing the parametric stability of trend regression to capture and project long-term structural directions [10]. This hybrid approach bridges the gap between non-parametric signal extraction and stable long-horizon forecasting. The novelty of this research lies in this specific methodological integration applied to a comparative context of the world's top three oil producers, utilizing an extended dataset up to 2024 to generate a decade-long strategic forecast (2025–2034). This provides a more nuanced evaluation of how different national production characteristics, from the technology-driven US shale dynamics to the policy-governed production of Saudi Arabia, respond to long-term hybrid modeling [1], [11].

Oil production drives national progress, particularly in the economic and infrastructure sectors. Countries with large oil reserves can generate significant revenue from oil exports, which in turn can be used for national development. Investments in energy infrastructure and technology are also heavily influenced by revenues generated from the oil sector. Some countries even rely entirely on the oil industry to support their economies. Therefore, success in oil production can be a determining factor in a country's progress on the international stage. The three countries with the largest oil production currently are the United States, Saudi Arabia, and Iraq [12].

Differences in geological characteristics, energy policies, and levels of technological advancement create highly diverse oil production dynamics across countries. The United States, for example, has experienced significant production increases thanks to the implementation of fracking technology, which enables the exploitation of shale oil. Conversely, Saudi Arabia and Iraq, as countries with the world's largest oil reserves, frequently experience production fluctuations due to OPEC policies, regional political situations, and dependence on crude oil exports [11]. These differing conditions make these three countries representative for analyzing the complex and volatile patterns of global oil production.

Therefore, fluctuations create data patterns that are difficult to analyze. Oil production data is often unstable over time. Changes in price, demand, and technical factors contribute to irregular data. As a result, data patterns are difficult to predict with simple methods [4]. Large fluctuations also disrupt long-term forecasting. This poses a significant challenge for researchers and policymakers. Not all analysis methods can address poor patterns. Some classical statistical methods are only effective for stable data. When data contains changing trends, these methods often fail. Analysis results can reveal significant errors, which can impact energy policy. Therefore, selecting the right method is crucial [1].

High uncertainty in production data often leads to biased or misleading analysis results. Uncertainty stems from both internal and external factors. Internal factors include production technology and resource quality. External factors include geopolitics, climate, and global market conditions. This combination of factors makes the data difficult to interpret. Careless or inappropriate analysis may result in incorrect conclusions, potentially leading to suboptimal or even harmful policy decisions [3]. Differences in production characteristics between countries also add complexity, making analysis increasingly difficult to conduct consistently. Each country has different geological conditions. These differences affect the number of reserves and production levels. Some countries have low fluctuations, while others have very high ones. These differences mean that one method is not always suitable for all countries. The Singular Spectrum Analysis (SSA) model can improve accuracy in addressing this issue and has been shown to effectively handle non-linear, non-stationary time series by decomposing data into trend, oscillatory, and noise components, thereby improving forecasting accuracy in the presence of high uncertainty [6], [13].

This research is motivated by the need for accurate modeling and forecasting of fluctuating, non-stationary, and complex oil production data, given that the dynamics of the global energy market are heavily influenced by economic, political, and environmental factors. To address this challenge, this study proposes a solution through the application of the SSA method combined with trend regression. This approach is particularly well suited for handling non-linear and non-stationary time series, as it decomposes the data into trend, seasonal, and noise components, thereby enhancing interpretability and forecasting accuracy [6], [13]. The main objective of this research is to predict the 2024 oil production trends of the world's three largest oil producers (the United States, Saudi Arabia, and Iraq) to obtain an accurate and stable predictive model that can effectively capture the complex dynamics governing global oil production.

This research builds upon previous studies that have demonstrated the effectiveness of Singular Spectrum Analysis (SSA) in handling non-linear and non-stationary time series, particularly in economic and energy-related applications [6], [13], [14]. Earlier research has mainly applied SSA as a standalone forecasting technique or combined it with short-term predictive models, with a primary focus on signal extraction and short-horizon accuracy. However, limited attention has been given to the integration of SSA with parametric trend modeling for long-term forecasting, especially in highly volatile and heterogeneous oil production data.

From a statistical science development perspective, lies in the formulation of a hybrid SSA–Trend Regression framework that explicitly links non-parametric SSA decomposition with parametric trend regression to enhance long-horizon forecast stability and interpretability. Unlike existing studies, this research systematically compares a single SSA model with the proposed hybrid model to quantify improvements in predictive accuracy using statistical performance measures. Moreover, the proposed framework is applied across the world’s three largest oil-producing countries is the United States, Saudi Arabia, and Iraq each exhibiting distinct production dynamics, volatility structures, and geological conditions. This cross-country application has not been comprehensively explored in previous SSA-based studies and provides new methodological insight into the robustness of hybrid statistical models under varying degrees of non-stationarity and uncertainty [15], [16]. Consequently, this study contributes to statistical science by extending the application of SSA beyond decomposition and short-term forecasting toward a generalizable hybrid modeling strategy for complex, long-term energy forecasting problems.

In light of these challenges, the primary objective of this research is to model and forecast annual oil production for the 2025–2034 horizon based on historical data from 1936–2024 using the hybrid SSA–Trend Regression framework. Specifically, this study aims to develop a robust predictive model that mitigates non-stationary noise, compares the long-term production trajectories across the United States, Saudi Arabia, and Iraq, and generates stable trend projections to assist policymakers in global energy planning and resource management [1], [17].

2. RESEARCH METHODS

This study uses a combined design of SSA and trend regression to forecast oil production in these three countries. This approach was chosen because it is effective in handling non-stationary and highly variable time series data. The study population includes annual oil production data from 1936-2024 sourced from the Our World in Data (Oil Production) database. The sample used is data from the 1936-2024 period for the United States, Saudi Arabia, and Iraq, sourced from the Our World in Data database. The primary instrument used is the SSA-Trend Regression method. The research stages include data collection and exploration to identify trend and fluctuation patterns, followed by data decomposition using SSA to separate key components such as trend, seasonality, and noise. Next, the resulting trend components are analyzed using trend regression models, both linear and nonlinear, to produce a stable and accurate forecasting model for non-stationary data. All statistical analyses, SSA decompositions, and regression modeling were performed using R programming language (version 4.x) with the 'Rssa' package. The research procedure is conducted in a step-by-step operational sequence:

1. Data collection and exploration,
2. SSA Decomposition to separate trend from noise,
3. SSA Reconstruction of the trend signal, and
4. Trend Regression modeling for long-term forecasting.

2.1 Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis or SSA is a non-parametric statistical method utilized to decompose time series data into its primary components, such as trend, seasonality, and noise [18]. Essentially, the SSA procedure comprises two fundamental stages, that is decomposition and reconstruction. The main advantage of SSA lies in its ability to handle nonlinear and non-stationary data and extract hidden trends from fluctuating data patterns.

2.1.1 Decomposition Stage

In the decomposition stage, an embedding process is performed, which involves forming a trajectory matrix from the original data. This is followed by Singular Value Decomposition (SVD) to extract singular values and identify dominant patterns. The following are the steps in the decomposition stage [18]:

1. Embedding

   In the embedding process, select a window length of $L(1 < L < N)$ and set $K = N - L + 1$. In this study, L = 36 was chosen to capture decadal cycles while maintaining a stable trajectory matrix. Select the eigentriple combination (e.g., r eigentriple) associated with the main trend/signal, and consider the remainder as oscillatory signals or noise. The principle is that eigentriples with large eigenvalues tend to contribute significant variance (trends or cycles). After that, form the following trajectory matrix:

$$\mathbf{X} = [X_1, X_2, \dots, X_K] = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_K \\ x_2 & x_3 & x_4 & \cdots & x_{K+1} \\ x_3 & x_4 & x_5 & \cdots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \cdots & x_N \end{pmatrix}, \quad (1)$$

   where each column $X_i = (x_i, x_{i+1}, \dots, x_{i+L-1})^T$. The matrix is a Hankel matrix (the “anti-diagonal” diagonals are equal).

2. Singular Value Decomposition (SVD)

   Calculate the matrix $\mathbf{XX}^{\mathrm{T}}$, then decompose it into eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$ and eigenvectors $U_1, U_2, \dots, U_L$ (EOFs). Then, form eigentriples (elementary matrices) for each $\sqrt{\lambda_i} \cdot U_i \cdot V_i^T$ with:

$$V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}. \quad (2)$$

2.1.2 Reconstruction Stage

The reconstruction stage includes grouping, which groups components according to structure (trend, seasonality, and noise) and diagonal averaging to transform the grouped matrices back into a one-dimensional time series, thereby reconstructing the filtered signal based on the relevant components (trend and seasonality) and eliminating the noise. The following are the steps in the reconstruction stage [18]:

1. Grouping

   Select a combination of eigentriples (r eigentriples) associated with the main trend/signal, and consider the rest as noise. Eigentriples with large eigenvalues tend to contribute significant variance (trend or cycle). Separation can be performed using w-correction (weighted correlation) between components to evaluate good separation. For the USA and Saudi Arabia, components 1–2 were selected to represent the main trend. For Iraq, components 1–4 were grouped to better capture the trend amidst higher volatility.

2. Diagonal Averaging

After selecting a group, reconstruct the trend series $\tilde{x}_n^{(trend)}$ using diagonal averaging or the diagonal averaging technique of the group matrix. This will yield the following components, oscillations, and noise:

$$x_n = \tilde{x}_n^{(trend)} + \tilde{x}_n^{(osc)} + \tilde{x}_n^{(noise)}. \quad (3)$$

2.2 Trend Regression in Time Series Analysis

Meanwhile, trend regression is used to model the long-term direction of a time series. In this study, simple linear regression was applied after extracting the trend component using the SSA method. Trend regression can be used to estimate the development of a variable over a specific period and measure the strength of the relationship between time and changes in data values. After obtaining the trend series $\tilde{x}_n^{(trend)}$ pa in SSA, the next step in trend regression is as follows:

1. Trend Regression Model Formulation

The form of the regression model is linear or non-linear (if pattern trend nonlinear) with $\tilde{x}_n^{(trend)}$ respect to time *t*. Simple linear regression model:

$$\tilde{x}_n^{(trend)} = \beta_0 + \beta_1 t + \varepsilon_t. \quad (4)$$

2. Parameter Estimation

Perform Estimate the parameters using the Ordinary Least Squares (OLS) method or nonlinear regression according to the model. Ensure that the classical regression assumptions (homoscedasticity, autocorrelation, normal residuals) are met. If the residuals show autocorrelation, corrections (such as GLS, Prais-Winsten) are necessary to ensure consistent estimates. This process determines the intercept (β_0) and the slope (β_1) that best fit the reconstructed trend. Based on the analysis, the final model equations for each country are:

United States:

$$\tilde{x}_{USA}^{(trend)} = \beta_0 + \beta_1 t. \quad (5)$$

Saudi Arabia:

$$\tilde{x}_{SAU}^{(trend)} = \beta_0 + \beta_1 t. \quad (6)$$

Iraq:

$$\tilde{x}_{IRQ}^{(trend)} = \beta_0 + \beta_1 t. \quad (7)$$

(Note: Replace β_0 and β_1 with the actual calculated coefficients from the Results section).

3. Model Evaluation

Model fit is evaluated using the coefficient of determination (R^2). The use of R^2 refers to [7], who assessed the results of the AAS reconstruction through a regression between the measured series and the reconstructed results using the squared correlation coefficient. Additionally, [15] also used adjusted R^2 to assess the fit of the BiLSTM regression model after component extraction with SSA and error metrics such as MAPE and RMSE are included to ensure the robustness of the trend projection.

4. Forecasting

Forecasting was performed using the selected regression models selected for projecting $\tilde{x}_n^{(trend)}$ in 2025-2034.

3. RESULTS AND DISCUSSION

3.1. Data Identification

In this research, the data used was annual oil production data in the USA, Saudi Arabia, and Iraq for the period 1936–2024. The initial stage was to identify data trends. Oil production in these three countries exhibits a recurring pattern of increases and decreases over time, reflecting significant fluctuations.

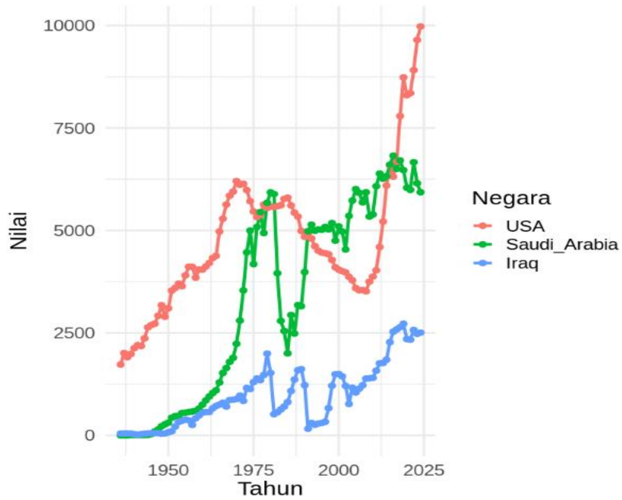


Figure 1. Time Series Plot of Oil Production (1936-2024) in the Top Three Oil-Producing Countries

From this fluctuation pattern, it can be identified that the oil production data exhibits a long-term trend with cyclical disturbances. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots show partial correlations at certain lags, indicating a relationship between production between previous periods.

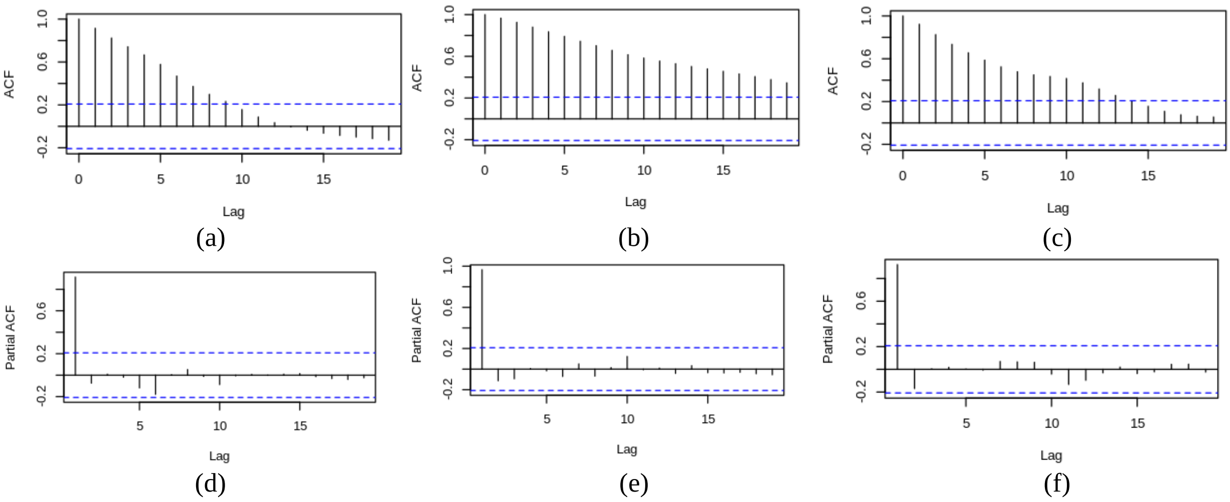


Figure 2. ACF and PACF Plot of Production Data for USA, Saudi Arabia, and Iraq Oil
(a) ACF-USA, (b) ACF-Saudi Arabia, (c) ACF-Iraq, (d) PACF-USA, (e) PACF-Saudi Arabia, (f) PACF-Iraq

The ACF plots show a slow decay across lags, indicating strong persistence and non-stationary behavior in oil production series for all three countries. The PACF plots exhibit a significant spike at lag 1 followed by insignificant lags, suggesting that an AR(1) structure is dominant in the data.

Next, an Augmented Dickey-Fuller (ADF) test was conducted to assess the data's stationarity. The findings from the ADF test are presented below:

Table 1. ADF Test Results

Country	ADF Statistic	Lag Order	p-value	Conclusion
USA	-1.7134	4	0.694	Non-stationary
Saudi Arabia	-3.0686	4	0.1364	Non-stationary
Iraq	-2.6404	4	0.3126	Non-stationary

ADF test results show that the three oil production data sets (USA, Saudi Arabia, Iraq) are non-stationary, meaning their means and variances change over time, and there are long-term trends. Therefore, an analysis method capable of handling non-stationary data is essential. In this research, Singular Spectrum Analysis (SSA) was chosen appropriately because SSA can extract long-term trends without having to make the data stationary through differencing, as in the ARIMA method.

3.2 Data Analysis

This study used the SSA and trend regression methods to analyze oil production patterns in the world's top three oil-producing countries, that is the USA, Saudi Arabia, and Iraq. SSA has two stages, that is the Decomposition stage and the Reconstruction stage.

3.2.1 Stage 1: Decomposition Stage

The decomposition stage involves two primary steps, that is embedding and Singular Value Decomposition (SVD).

1. Embedding

   In the SSA decomposition stage, the initial step is embedding, which organizes annual oil production data into a Hankel trajectory matrix. This study used n = 84 years of data, so the window length was chosen to be L = 36. The K value was obtained from the formula $K = n - L + 1 = 49$. This trajectory matrix will be used in the next stage, SVD, to extract the principal components (trends, cycles, and noise). The results of this process are as follows:

   The USA trajectory matrix shows a strong trend pattern. Relative production values increase with moderate fluctuations. The trajectory matrix shows a strong correlation between periods, indicating a stable long-term trend.

$$X_{USA} = \begin{bmatrix} 1728.29 & 2010.42 & 1908.56 & 1988.10 & 2126.81 & 2203.84 \\ 2010.42 & 1908.56 & 1988.10 & 2126.81 & 2203.84 & 2179.35 \\ 1908.56 & 1988.10 & 2126.81 & 2203.84 & 2179.35 & 2365.64 \\ 1988.10 & 2126.81 & 2203.84 & 2179.35 & 2365.64 & 2637.11 \\ 2126.81 & 2203.84 & 2179.35 & 2365.64 & 2637.11 & 2693.30 \\ 2203.84 & 2179.35 & 2365.64 & 2637.11 & 2693.30 & 2725.18 \end{bmatrix}.$$

   The SAU trajectory matrix displays a production pattern with large jumps in the initial periods leading to a peak in high production. This indicates a strong trend but with significant variations, likely related to OPEC policies and oil supply controls.

$$X_{SAU} = \begin{bmatrix} 0.02326 & 0.09304 & 0.77921 & 6.26857 & 8.14100 & 6.86170 \\ 0.09304 & 0.77921 & 6.26857 & 8.14100 & 6.86170 & 7.21060 \\ 0.77921 & 6.26857 & 8.14100 & 6.86170 & 7.21060 & 7.55950 \\ 6.26857 & 8.14100 & 6.86170 & 7.21060 & 7.55950 & 12.36269 \\ 8.14100 & 6.86170 & 7.21060 & 7.55950 & 12.36269 & 33.30136 \\ 6.86170 & 7.21060 & 7.55950 & 12.36269 & 33.30136 & 95.36600 \end{bmatrix}.$$

   The Iraq trajectory matrix shows more irregular fluctuations. The matrix columns contain sharply fluctuating production values, reflecting the impact of significant geopolitical and domestic security factors. As a result, the trend components are difficult to capture with just a few groupings, requiring a broader set of combinations (e.g., 1–4).

$$X_{IRQ} = \begin{bmatrix} 45.29 & 47.83 & 48.12 & 46.09 & 29.16 & 18.22 \\ 47.83 & 48.12 & 46.09 & 29.16 & 18.22 & 30.18 \\ 48.12 & 46.09 & 29.16 & 18.22 & 30.18 & 41.54 \\ 46.09 & 29.16 & 18.22 & 30.18 & 41.54 & 48.42 \\ 29.16 & 18.22 & 30.18 & 41.54 & 48.42 & 53.58 \\ 18.22 & 30.18 & 41.54 & 48.42 & 53.58 & 54.43 \end{bmatrix}.$$

2. Singular Value Decomposition

   The second step is singular value decomposition or SVD, which involves finding the eigentriple values of the path matrix $\mathbf{X}$(36×49). Before finding the eigentriple values, a symmetric matrix is first created using the formula $\boldsymbol{S} = \mathbf{X}\mathbf{X}^{\boldsymbol{t}}$. Thus, the symmetric matrix is obtained by multiplying the path matrix by its transpose, which is then used to extract the singular values and eigenvectors. This process aims to separate the principal components in the oil production data so that dominant trends, cyclical patterns, and noise can be identified separately.

Table 2. Singular Value Decomposition

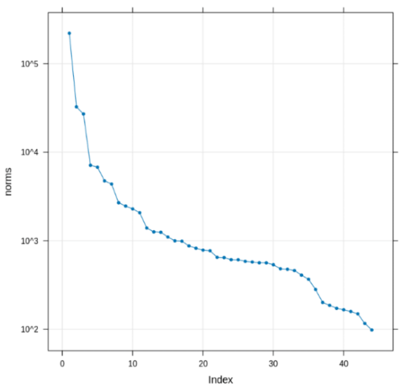
Component	Eigen Value			Proportion of Variance		
	USA	Saudi Arabia	Iraq	USA	Saudi Arabia	Iraq
1	214574.978	168314.345	43377.688	67.88	47.27	35.46
2	33047.363	29531.091	10238.984	10.45	8.29	8.37
3	22958.933	22086.817	7342.764	7.26	6.20	6.00
4	6958.552	21273.816	7115.320	2.20	5.97	5.82
5	6531.624	17398.841	6967.747	2.07	4.89	5.70
6	4699.021	13269.809	5373.263	1.49	3.73	4.39
7	3221.235	9512.457	4524.465	1.02	2.67	3.70
8	2364.787	5946.096	2722.803	0.75	1.67	2.23
9	2160.770	5236.683	2701.462	0.68	1.47	2.21
10	1766.055	5162.830	2591.743	0.56	1.45	2.12

Based on Table 2, it can be seen that:

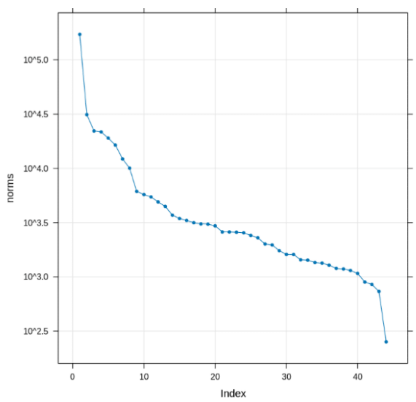
In the USA, Component 1 obtained an Eigenvalue of 214.574; Proportion of Variance of 67.88%. For Component 2, the Proportion of Variance of 10.45% was obtained. For Component 3, the Proportion of Variance of 7.26%. The first three components explained nearly 85.6% of the data variation. After component 3, the contribution was very small, less than 2%. This means that US oil production is dominated by the long-term trend (component 1) with little additional variation in components 2 and 3. The data is relatively stable, so it is sufficient to analyze it by grouping components 1–2.

In Saudi Arabia, Component 1 obtained an Eigenvalue of 168.314; Proportion of Variance of 47.27%. Components 2–4 explained 8.29%, 6.20%, and 5.97%, respectively. The first five components explained approximately 72.6% of the total variation. Contributions were more dispersed than in the USA. This means that Saudi Arabian oil production has a dominant trend (component 1), but several additional cycles (components 2–5) are quite significant, which could reflect the influence of OPEC policies or global market fluctuations.

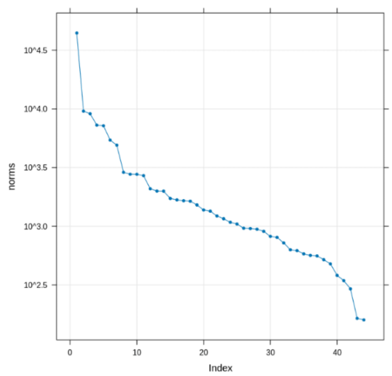
For Iraq, Component 1 obtained an Eigenvalue of 43.377; Proportion of Variance = 35.46%. Components 2–5 contributed 8.37%, 6.00%, 5.82%, and 5.70%, respectively. The total of the first five components explained approximately 61.4% of the variation, much smaller than for the USA and Saudi Arabia. The contributions between components were more evenly distributed. This means that Iraqi oil production is highly volatile. No single component dominates. Many components are significant (components 1–5), resulting in a mixed trend with large cyclical variations. This aligns with Iraq's frequent political and geopolitical instability.



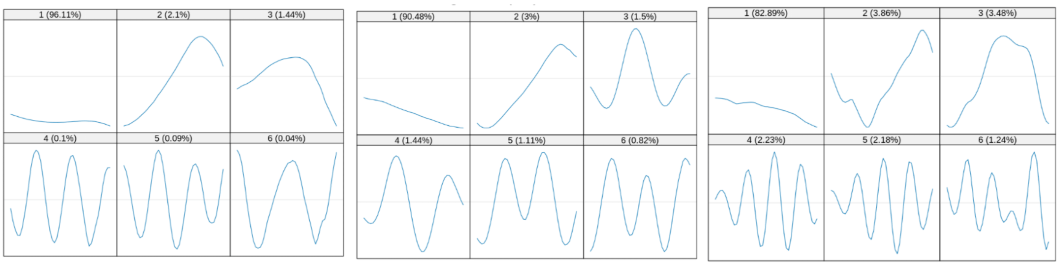
(a)



(b)



(c)

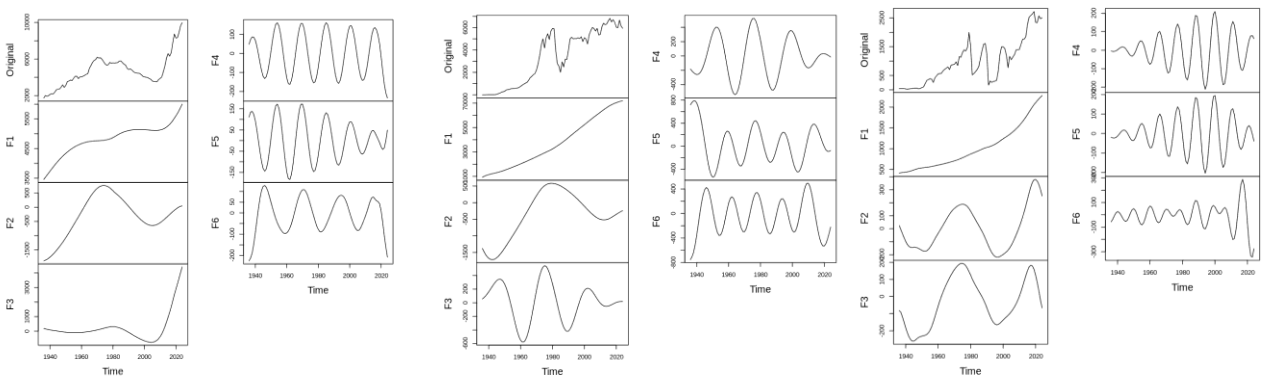


(d) (e) (f)

Figure 3. Singular Value Pattern in USA, SAU and Iraq
(a) Singular Values Spectrum-USA, (b) Singular Values Spectrum-Saudi Arabia, (c) Singular Values Spectrum -Iraq, (d) Eigenvalue (EOF)-USA, (e) Eigenvalue (EOF)-Saudi Arabia, (f) Eigenvalue (EOF)-Iraq

The first singular value dominates, indicating that the main variation in oil production is explained by a single dominant trend component. The leading EOF captures long-term trends, while higher-order EOFs represent cyclical and short-term fluctuations, reflecting production dynamics and structural changes over time.

Next, the purpose of SSA component reconstruction is to filter, smooth, and redisplay the main signals from the data, so that important trends and patterns in oil production can be more clearly understood and serve as a basis for more accurate forecasting.



(a) (b) (c)

Figure 4. Components Reconstruction (1–6) in USA, SAU and Iraq
(a) Components Reconstruction SSA-USA, (b) Components Reconstruction SSA-Saudi Arabia, (c) Components Reconstruction SSA-Iraq

The results of component reconstruction (1–6) on oil production data show that the USA is dominated by a smooth and stable long-term trend, Saudi Arabia displays a major trend with additional policy cycles, while Iraq continues to exhibit an unstable trend pattern with large fluctuations.

3.2.2 Stage 2: Reconstruction Stage

The second phase of the procedure is the reconstruction stage, consisting of two subsequent steps, that is grouping and diagonal averaging.

1. Grouping

   The grouping stage is carried out by grouping the eigenvectors from the decomposition based on the similarity of the pattern characteristics they exhibit. In the oil production data analysis, the eigenvector plot shows that the first component represents the main long-term trend, while components 2–4 display patterns resembling cycles or seasonal variations. Components 5 and 6 again reflect additional, smoother trends, while components 7–9 tend to act as noise or random fluctuations with no clear pattern. The grouping results are as follows.

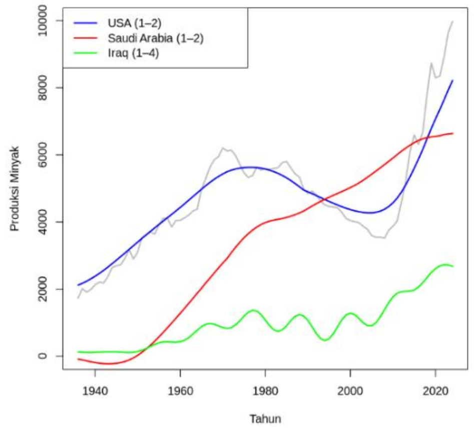


Figure 5. Interstate Oil Production Grouping Results

In the grouping stage of oil production analysis, the SSA decomposition results indicate that the USA trend can be reconstructed with just 1–2 components due to its stable long-term pattern. Saudi Arabia also requires 1–2 groups, which represent the main trend, although still influenced by cyclical variations in OPEC policy. While Iraq requires a broader grouping of 1–4 to clearly discern its long-term trend, as it is mixed with sharp fluctuations due to geopolitical factors.

2. Diagonal Averaging

The diagonal averaging process for the entire matrix reorganizes the time series, resulting in a smoother reconstruction of the oil production signal. From this diagonal averaging, the main trend, cyclical patterns, and random fluctuations in oil production data can be separated and analyzed more clearly.

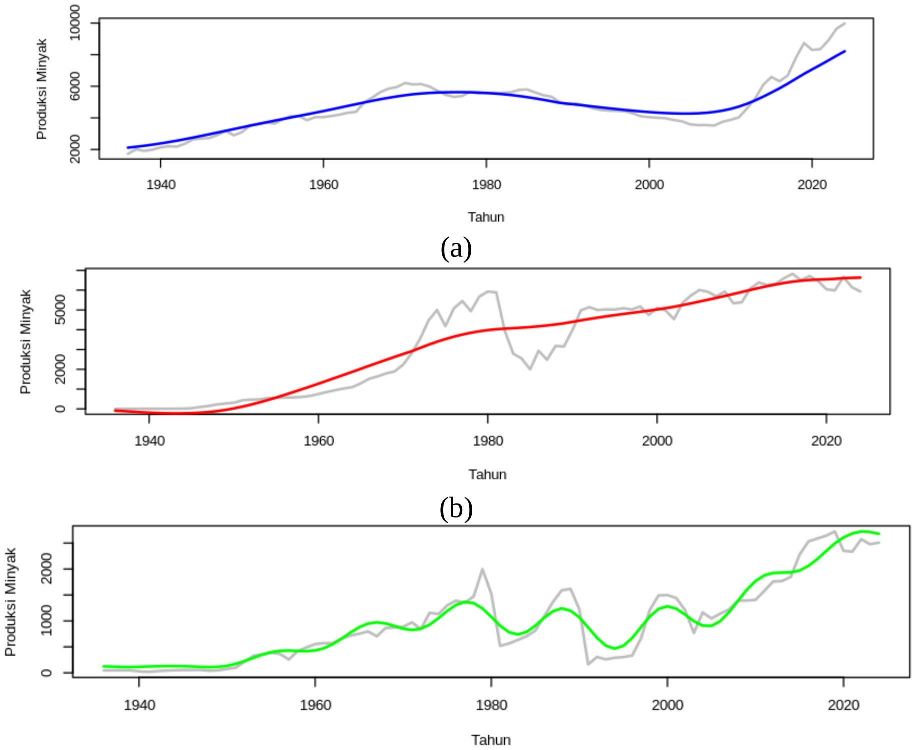


Figure 6. Diagonal Averaging Results for Interstate Oil Production
(a) Diagonal Averaging-Oil Production USA, (b) Diagonal Averaging-Oil Production Saudi Arabia, (c) Diagonal Averaging-Oil Production Iraq

The diagonal averaging results show differences in oil production characteristics between countries. The USA is more stable with a clear dominant trend, Saudi Arabia is strong with a main trend but is still affected by policy cycles, while Iraq shows a mixed trend with large fluctuations reflecting high uncertainty in its production.

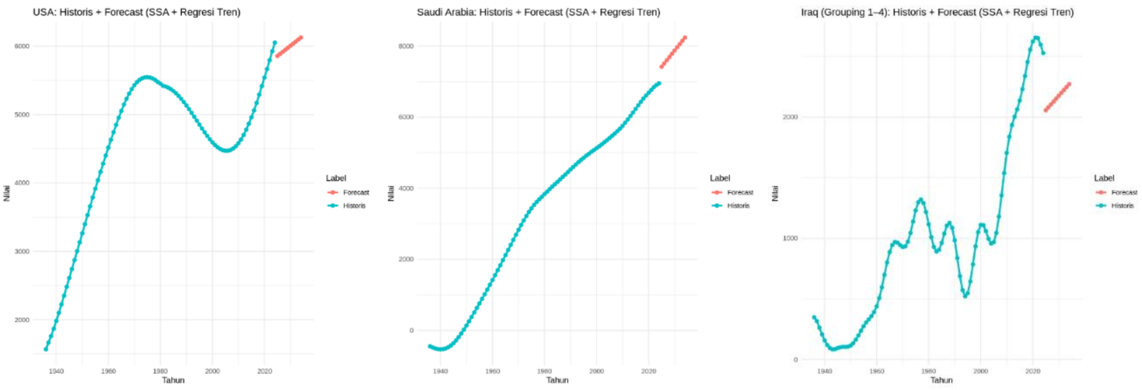
3.3 Forecasting Stage

The next step in this research is to conduct a forecast to estimate future oil production by utilizing key trends and patterns that have been separated from noise. This forecasting method uses the SSA + Trend Regression method. The SSA method was chosen because it is able to unravel trends, cycles, and noise in complex oil production data, while trend regression is used to capture long-term trends and provide more interpretable forecasts. The combination of the two makes the analysis more robust, flexible, and accurate in describing oil production dynamics in the USA, Saudi Arabia, and Iraq. The forecasting results are as follows:

Table 3. Forecasting Results

Year	USA	Saudi Arabia	Iraq
2025	5857.75	7415.64	2056.09
2026	5887.66	7506.76	2080.01
2027	5917.58	7597.89	2103.93
2028	5947.49	7689.02	2127.85
2029	5977.41	7780.15	2151.77
2030	6007.33	7871.27	2175.69
2031	6037.24	7962.40	2199.61
2032	6067.16	8053.53	2223.53
2033	6097.07	8144.66	2247.45
2034	6126.99	8235.78	2271.37

US oil production is projected to increase slowly from approximately 5,858 in 2025 to 6,127 in 2034, while Saudi Arabia tends to remain stable at a high level, increasing from 7,416 to 8,236, and Iraq shows a moderate increase from 2,056 to 2,271 over the same period.



(a) (b) (c)

Figure 7. Forecasting Results Plot International Oil Production
(a) Forecasting Results Plot-USA, (b) Forecasting Results Plot-Saudi Arabia, (c) Forecasting Results Plot-Iraq

For the USA, the forecast indicates a continued upward trend in oil production following historical fluctuations, reflecting recovery and growth in recent years. For Saudi Arabia, the production series shows a strong and stable increasing trend, and the forecast suggests sustained growth driven by consistent production capacity. For Iraq, the data exhibit high historical volatility. However, the forecast projects a positive upward trend, indicating potential production expansion despite past instability.

To evaluate the forecasting performance of the SSA–Trend Regression hybrid model, three statistical metrics were utilized, the coefficient of determination (R^2), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). The R^2, MAPE, and RMSE values obtained in this research are as follows:

Table 4. Model Accuracy Table

Country	R^2	MAPE	RMSE
USA	0.480	18.735	799.998
Saudi Arabia	0.980	57.134	334.241
Iraq	0.751	42.369	354.296

The results indicate that Saudi Arabia achieves the highest R^2 value (0.98), reflecting an exceptionally stable long-term production trend. This is consistent with previous studies highlighting the stabilizing role of OPEC policies and long-term state-led production planning [1], [4]. However, Saudi Arabia also exhibits the highest MAPE (57.134%), which suggests that while the model captures the overall trend direction perfectly, the absolute percentage deviation remains high due to historical production shifts and large-scale output adjustments.

In contrast, the United States shows the lowest R^2 (0.480), indicating high production variability driven by the dynamic shale oil sector and its acute sensitivity to global market prices [1]. Interestingly, the USA achieves the lowest MAPE (18.735%), suggesting that despite the lower trend correlation, the model's predictions remain relatively close to the actual values in percentage terms.

Iraq's intermediate R^2 value (0.751) and MAPE (42.369%) align with its dual nature, a strong long-term capacity development trend periodically disrupted by short-term geopolitical instability [1]. The RMSE values further confirm these dynamics, where the USA's high RMSE (799.998) reflects large-scale volume fluctuations, while Saudi Arabia and Iraq show more concentrated error margins around their mean production levels. Overall, these findings confirm that production stability and the underlying drivers of the energy sector significantly influence the performance and error structure of trend-based forecasting models.

4. CONCLUSION

Based on the Singular Spectrum Analysis (SSA) decomposition, this study reveals that the world's top three oil producers exhibit distinct variance structures. The United States shows a high concentration of variance in the first component (67.88%), indicating a dominant and relatively stable long-term trend. In contrast, Saudi Arabia displays a more distributed variance across the first five components (totaling 72.6%), reflecting a major trend influenced by periodic production cycles under OPEC policies. Meanwhile, Iraq presents the flattest variance distribution (61.4% across the first five components), underscoring high production volatility driven by geopolitical instability.

This research successfully addresses the existing research gap by formulating a hybrid SSA–Trend Regression framework. This integration proves theoretically superior as it leverages SSA's non-parametric de-noising capabilities with the parametric stability of trend regression, effectively handling non-stationary energy data. The final regression models for the 2025–2034 horizon project a sustained growth in oil production for all three nations, albeit with different growth rates and stability levels.

The robustness of the proposed model is validated through multiple performance metrics. Saudi Arabia ($R^2 = 0.98$) exhibits the most predictable trend, while the United States ($R^2 = 0.48$) remains the most sensitive to market-driven fluctuations. These findings underscore the profound impact of geological, technological, and geopolitical factors on global energy stability. Ultimately, this study provides a significant practical contribution for global policymakers, offering a stable and accurate tool for long-term strategic energy planning and oil production projections for the coming decade.

Author Contributions

Ega Saherti: Conceptualization, Software, Data Curation, Formal Analysis, Visualization, Validation, Writing–Original Draft, Writing–Review and Editing. Salwa Azzah Imtiyaz: Conceptualization, Methodology, Investigation, Data Curation, Project Administration, Resources, Visualization, Validation, Writing–Original Draft, Writing–Review and Editing. Gumgum Darmawan: Conceptualization, Software, Supervision, Validation, Visualization. Budi Nurani Ruchjana: Supervision, Validation, Writing–Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

We would like to express our sincere appreciation to the anonymous reviewers for their valuable comments and constructive feedback, which greatly improved the quality of this paper. We also extend our appreciation to our colleagues who contributed their insights and assistance during the writing process, enabling this paper to reach its final form.

Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Declaration of Generative AI and AI-assisted technologies

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript, including for writing, editing, data analysis, or the creation of tables and figures.

REFERENCES

[1] L. Cozzi and Tim Gould, "WORLD ENERGY OUTLOOK 2023," France, 2023. [Online].
[2] World Health Organization, "STRATEGIC HEALTH INFRASTRUCTURE INVESTMENTS TO SUPPORT UNIVERSAL HEALTH COVERAGE," 2023. [Online].
[3] G. T. Wilson, "TIME SERIES ANALYSIS: FORECASTING AND CONTROL," *J. Time Ser. Anal.*, vol. 37, no. 5, pp. 709–711, Sep. 2016, doi: https://doi.org/10.1111/jtsa.12194
[4] J. D. Hamilton, "UNDERSTANDING CRUDE OIL PRICES," Cambridge, Nov. 2008. [Online].doi: https://doi.org/10.3386/w14492
[5] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *ANALYSIS OF TIME SERIES STRUCTURE: SSA AND RELATED TECHNIQUES*. Monographs on Statistics and Applied Probability, 2001.doi: https://doi.org/10.1201/9781420035841
[6] H. Hassani, "SINGULAR SPECTRUM ANALYSIS: METHODOLOGY AND COMPARISON," *Journal of Data Science*, vol. 5, pp. 239–257, 2007, doi: https://doi.org/10.6339/JDS.2007.05(2).396.
[7] D. H. Schoellhamer, "SINGULAR SPECTRUM ANALYSIS FOR TIME SERIES WITH MISSING DATA," *Geophys. Res. Lett.*, vol. 0, no. 0, p. 0, 2001.
[8] T. Knapik, A. Ratiarison, and H. Razafindralambo, "AN EXPERIMENTAL EVALUATION OF CHOICES OF SSA FORECASTING PARAMETERS," *Revue Africaine de Recherche en Informatique et Mathématiques Appliquées*, vol. 40, Mar. 2024, doi: https://doi.org/10.46298/arima.9641.
[9] N. Golyandina, "PARTICULARITIES AND COMMONALITIES OF SINGULAR SPECTRUM ANALYSIS AS A METHOD OF TIME SERIES ANALYSIS AND SIGNAL PROCESSING," *St. Petersburg State University*, Jan 2021, doi: https://doi.org/10.1002/wics.1487.
[10] G. Aydin, "REGRESSION MODELS FOR FORECASTING GLOBAL OIL PRODUCTION," *Pet. Sci. Technol.*, vol. 33, pp. 1822–1828, Nov. 2015, doi: 10.1080/10916466.2015.1101474.
[11] US Energy Information Administration, "OIL AND PETROLEUM PRODUCTS EXPLAINED: WHERE OUR OIL COMES FROM," US Energy Information Administration. Accessed: Nov. 27, 2025. [Online].
[12] M. Roser and H. Ritchie, "OIL PRODUCTION," Our World in Data. Accessed: Nov. 27, 2025. [Online].
[13] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, "ANALYSIS OF TIME SERIES STRUCTURE: SSA AND RELATED TECHNIQUES," *Monographs on Statistics and Applied Probability*, vol. 90, p. 320, Jan. 2001, doi: https://doi.org/10.1201/9781420035841.
[14] H. Hassani, A. Webster, E. Silva, and S. Heravi, "FORECASTING U.S. TOURIST ARRIVALS USING OPTIMAL SINGULAR SPECTRUM ANALYSIS," *Tour. Manag.*, vol. 46, pp. 322–335, Feb. 2015, doi: https://doi.org/10.1016/j.tourman.2014.07.004.
[15] X. Zhang, X. Jiang, and Y. Li, "PREDICTION OF AIR QUALITY INDEX BASED ON THE SSA-BILSTM-LIGHTGBM MODEL," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-32775-2.
[16] J. Wang *et al.*, "SINGULAR SPECTRUM ANALYSIS (SSA) BASED HYBRID MODELS FOR EMERGENCY AMBULANCE DEMAND (EAD) TIME SERIES FORECASTING," *IMA Journal of Management Mathematics*, vol. 35, no. 1, pp. 45–64, Jan. 2024, doi: https://doi.org/10.1093/imaman/dpad019.
[17] Enerdata, "WORLD CRUDE OIL PRODUCTION STATISTICS," Enerdata. Accessed: Nov. 27, 2025. [Online].
[18] N. Golyandina and A. Zhigljavsky, *SINGULAR SPECTRUM ANALYSIS FOR TIME SERIES*. Heidelberg: Springer Berlin, 2013 doi: https://doi.org/10.1007/978-3-642-34913-3.
[19] MIchael. Gilliland, *FORECASTING WITH SAS : SPECIAL COLLECTION*. SAS Institute, 2020.
[20] H. Utami, Y. W. Sari, S. Subanar, A. Abdurakhman, and G. Gunardi, "PERAMALAN BEBAN LISTRIK DAERAH ISTIMEWA YOGYAKARTA DENGAN METODE SINGULAR SPECTRUM ANALYSIS (SSA)," *MEDIA STATISTIKA*, vol. 12, no. 2, pp. 214–225, Dec. 2019, doi: https://doi.org/10.14710/medstat.12.2.214-225.
[21] F. Paris, R. Simonella, and R. Casalini, "SINGULAR SPECTRUM ANALYSIS: BACK TO BASICS," *SSRN*, 2025, doi: https://doi.org/10.2139/ssrn.5136637.
[22] S. B. Swart, A. R. den Otter, and C. J. C. Lamoth, "SINGULAR SPECTRUM ANALYSIS AS A DATA-DRIVEN APPROACH TO THE ANALYSIS OF MOTOR ADAPTATION TIME SERIES," *Biomed. Signal Process. Control*, vol. 71, no. Part A, p. 103068, Jan. 2022, doi: https://doi.org/10.1016/j.bspc.2021.103068.
[23] J. I. Melinda, "PENERAPAN MODEL SINGULAR SPECTRUM ANALYSIS DALAM MEMPREDIKSI INFLASI DI INDONESIA AKIBAT COVID-19," Skripsi, UIN Syarif Hidayatullah, Depok, 2023.
[24] M. Leonard and B. Elsheimer, "AUTOMATIC SINGULAR SPECTRUM ANALYSIS AND FORECASTING," *SAS Institute Inc*, 2017.

[25] N. Golyandina and A. Zhigljavsky, “SINGULAR SPECTRUM ANALYSIS FOR TIME SERIES,” Heidelberg: Springer Berlin, 2013, ch. 2, pp. 11–70. doi: https://doi.org/10.1007/978-3-642-34913-3_2.
[26] M. Ghil *et al.*, “ADVANCED SPECTRAL METHODS FOR CLIMATIC TIME SERIES,” *Reviews of Geophysics*, vol. 40, no. 1, pp. 3-1-3–41, 2002, doi: https://doi.org/10.1029/2000RG000092.
[27] K. Kume and N. Nose-Togawa, “SPECTRAL STRUCTURE OF SINGULAR SPECTRUM DECOMPOSITION FOR TIME SERIES,” *ArXiv*, Jul. 2015.
[28] N. Golyandina and A. Shlemov, “SEMI-NONPARAMETRIC SINGULAR SPECTRUM ANALYSIS WITH PROJECTION,” *Stat. Interface*, vol. 10, no. 1, pp. 47–57, Jul. 2015, doi: https://doi.org/10.4310/SII.2017.v10.n1.a5.
[29] E. Purnama, “APLIKASI METODE SINGULAR SPECTRUM ANALYSIS (SSA) PADA PERAMALAN CURAH HUJAN DI PROVINSI GORONTALO,” *JAMBURA JOURNAL OF PROBABILITY AND STATISTICS*, vol. 3, no. 2, pp. 1–10, 2022, doi: 10.34312/jjps.v3i2.16537.
[30] TCD Group, “SINGULAR SPECTRUM ANALYSIS (SSA),” UCLA. Accessed: Nov. 27, 2025. [Online].
[31] N. Golyandina and E. Osipov, “THE ‘CATERPILLAR’-SSA METHOD FOR ANALYSIS OF TIME SERIES WITH MISSING VALUES,” *Statistical Planning and Inference*, vol. 137, no. 8, Feb. 2006, [Online].doi: https://doi.org/10.1016/j.jspi.2006.05.014
[32] A. Zhigljavsky, “SINGULAR SPECTRUM ANALYSIS FOR TIME SERIES: INTRODUCTION TO THIS SPECIAL ISSUE,” *Stat. Interface*, vol. 3, pp. 255–258, 2010, [Online].doi: https://doi.org/10.4310/SII.2010.v3.n3.a1.
[33] GistaT Group, “TIME SERIES ANALYSIS AND FORECASTING, CATERPILLAR SSA METHOD,” GistaT Group. Accessed: Nov. 27, 2025. [Online].
[34] H. Hassani and R. Mahmoudvand, *SINGULAR SPECTRUM ANALYSIS USING R*. London: Palgrave MacMillan, 2018.doi: https://doi.org/10.1057/978-1-137-40951-5
[35] X. Xu, M. Zhao, and J. Lin, “DETECTING WEAK POSITION FLUCTUATIONS FROM ENCODER SIGNAL USING SINGULAR SPECTRUM ANALYSIS,” *ISA Trans.*, vol. 71, pp. 440–447, Nov 2017, doi: https://doi.org/10.1016/j.isatra.2017.09.001.