



Kharazmi University

Comparison of singular value analysis set procedure and singular spectral analysis method based on new path matrix in reconstruction and forecasting of time series

Masoud Yarmohammadi¹, Maryam Movahedifar²

1. Corresponding Author, Department of Statistics, Payame Noor University (PNU), Tehran, Iran.
E-mail: masyar@pnu.ac.ir
2. Institute for Statistics, University of Bremen, Bibliothekstraße 1, Bremen, 28359, Germany.
E-mail: movahedm@uni-bremen.de

Article Info

Article type:

Research Article

Article history:

Received: 29 March 2020

Received in revised form:

14 September 2021

Accepted: 16 March 2022

Published online:

6 February 2024

Keywords:

Singular value analysis set,
Singular spectrum
analysis,
Singular value decomposition,
Hankel matrix.

ABSTRACT

Introduction

Singular spectrum analysis (SSA) is a new powerful method in time series analysis. This non-parametric method is very suitable due to its unique features, such as not requiring the assumptions of time series stationarity, normality of residuals, and its applicability to short time series. The main purpose of SSA method is to decompose time series into interpretable components such as trend, oscillatory component, and unstructured noise. The basis of SSA is singular value decomposition of the trajectory matrix built on the time series. In the basic SSA method the frequency of observations which used in the trajectory matrix is different and so there may be an error in reconstructing and forecasting the time series, especially at the beginning and end of the series. It occurs because the magnitude of eigenvalues, eigenvectors, and consequently, reconstruction and forecasting of future values of time series, is directly related to the trajectory matrix. The purpose of this paper is to improve the trajectory matrix of SSA method to increase the accuracy of the reconstructed time series and forecasting results, which is called singular spectrum decomposition (SSD). In this paper, SSA and SSD methods and their properties are briefly introduced and then the performance of SSD method over SSA method in time series reconstruction and forecasting for simulated and real data is discussed.

Material and Methods

Singular Spectrum Analysis (SSA). SSA is a nonparametric method that proves useful in extracting structural characteristics from time series data. It is particularly effective in situations where the series is non-stationary, and traditional modeling assumptions do not apply. The basic SSA method involves two key stages: decomposition and reconstruction. At the decomposition stage, the time series is broken down into a sum of independent and interpretable components, including trends and unstructured noise. The reconstruction stage involves piecing together a noise-free series from the decomposed components. Notably, the frequency of observations used in the trajectory matrix varies in SSA.

Singular Spectrum Decomposition (SSD). SSD is an iterative method built upon the SSA time series decomposition framework. In SSD, the selection of the embedding dimension and the choice of principal components for

reconstruction and forecasting are fully data-driven. This characteristic makes SSD an adjustable decomposition method, enhancing its adaptability to different data patterns. The trajectory matrix structure in SSD introduces a notable departure from SSA. In the new structure, all observations of the initial series appear L times, with L representing a parameter related to the window length. This change, along with other mathematical adjustments, results in a trajectory matrix with $L(L-1)$ times more observations than the previous SSA method.

Recurrent SSA (R-SSA) and R-SSD. Two forecasting variations within SSA are Vector SSA (SSA-V) and Recurrent SSA (R-SSA). While SSA-V has demonstrated efficiency in various instances, there is a recognized need to enhance the R-SSA forecasting approach. This paper introduces an innovative algorithm named R-SSD, a modification of R-SSA. R-SSD generates its coefficients from a modified trajectory matrix, aiming to provide more accurate forecasts. Both basic SSA and the new SSD method consist of two stages: decomposition and reconstruction. The key distinction lies in the frequency of observations used in the trajectory matrix, with SSD introducing a structure where observations appear L times. Additionally, SSD employs a fixed weight function based on the window length L , deviating from the trapezoid-like weight function in SSA.

Results and discussion

The comparative analysis between R-SSD and basic R-SSA forecasting algorithms involves the use of both simulated and real data. Evaluation criteria such as Root Relative Mean Squared Error (RRMSE) and Root Mean Absolute Deviation (RMAD) are employed to assess the performance of the algorithms. The results confirm the proposed R-SSD approach's promising output for reconstructing noise-free signals. Furthermore, R-SSD consistently outperforms basic R-SSA across different window lengths and horizons. The efficiency of the R-SSD method is further evaluated using the KSPA test, providing a clear indication that R-SSD is more powerful than its competing R-SSA model.

Conclusion

This paper contributes to the field of time series analysis by emphasizing the significance of SSA, introducing the innovative SSD method, and proposing the R-SSD forecasting algorithm. SSA's nonparametric nature and wide-ranging applications make it a valuable tool for extracting signals from time series data. The introduction of SSD as an iterative approach enhances the adaptability and data-driven decision-making capabilities of SSA. The development of R-SSD further strengthens the forecasting capabilities, providing a more accurate alternative to basic R-SSA. The comprehensive evaluation and comparison between SSA and SSD, backed by mathematical insights and empirical results, offer valuable insights for researchers and practitioners engaged in forecasting and data analysis.

How to cite: Yarmohammadi, M., Movahedifar, M., (2023). Comparison of singular value analysis set procedure and singular spectral analysis method based on new path matrix in reconstruction and forecasting of time series. *Mathematical Researches*, 9 (3), 75 - 93.



© The Author(s).

Publisher: Kharazmi University

مقایسه روش مجموعه تحلیل مقادیر تکین و روش تجزیه طیفی تکین بر پایه ماتریس مسیر جدید در بازسازی و پیش‌بینی سری‌های زمانی

مسعود یارمحمدی^۱، مریم موحدی فر^۲

۱. نویسنده مسئول، گروه آمار، دانشگاه پیام نور، تهران، ایران. رایانامه: masyar@pnu.ac.ir

۲. موسسه آمار، دانشگاه برمن، Bibliothekstraße 1، برمن، ۲۸۳۵۹، آلمان. رایانامه: movahedm@uni-bremen.de

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	تحلیل مجموعه‌ی مقادیر تکین (SSA) روشی نسبتاً جدید و قدرتمند در حوزه‌ی تحلیل سری‌های زمانی است. روش SSA یک روش ناپارامتری است که به دلیل دارا بودن ویژگی‌هایی منحصر به فرد، نظیر عدم نیاز به برقراری فرض‌های مانایی سری زمانی و نرمال بودن مانده‌ها، مورد توجه بسیاری از پژوهشگران در حوزه‌ی تحلیل سری‌های زمانی و اقتصادسنجی قرار گرفته است. هدف اصلی روش تحلیل مجموعه مقادیر تکین (SSA) تجزیه سری‌های زمانی به اجزای تفسیرپذیر مانند روند، مولفه نوسانی و نوفه بدون ساختار است. مبنای SSA تجزیه مقدار تکین ماتریس مسیر ساخته شده بر روی سری زمانی است. ماتریس مسیر به کار رفته در روش تحلیل مجموعه مقادیر تکین، طوری طراحی شده که فراوانی مشاهدات بردار سری زمانی اولیه در ماتریس مسیر با یکدیگر برابر نیستند و لذا بیم آن می‌رود که در بازسازی و پیش‌بینی سیگنال استخراج شده از نوفه، بخصوص در ابتدا و انتهای سری، خطا وجود داشته باشد، چرا که بزرگی مقادیر ویژه، بردارهای ویژه و در نتیجه تجزیه، بازسازی و پیش‌بینی مقادیر آینده سری زمانی ارتباط مستقیم با ماتریس مسیر دارد. هدف مقاله حاضر، بهبود و ارتقای ماتریس مسیر ساخته شده در روش SSA به منظور افزایش دقت سری زمانی نتیجه شده پس از بازسازی سری اولیه می‌باشد که روش تجزیه طیفی تکین (SSD) نام‌گذاری می‌شود. در این مقاله ضمن معرفی اجمالی هر دو روش و ویژگی‌های آن‌ها، کارایی روش SSD نسبت به روش SSA در بازسازی و پیش‌بینی سری زمانی برای داده‌های شبیه‌سازی شده و واقعی مورد بحث و بررسی قرار می‌گیرد.
تاریخ دریافت: ۱۳۹۹/۱/۱۰	
تاریخ بازنگری: ۱۴۰۰/۶/۲۳	
تاریخ پذیرش: ۱۴۰۰/۱۲/۲۵	
تاریخ انتشار: ۱۴۰۲/۱/۱۷	
واژه‌های کلیدی: روش تحلیل مجموعه‌ی مقادیر تکین، روش تجزیه طیفی تکین، تجزیه مقدار تکین، ماتریس هنکل.	

استناد: یارمحمدی، مسعود، موحدی فر، مریم (۱۴۰۲). مقایسه روش مجموعه تحلیل مقادیر تکین و روش تجزیه طیفی تکین بر پایه ماتریس مسیر جدید در بازسازی و پیش‌بینی سری‌های زمانی. *پژوهش‌های ریاضی*، ۹ (۳)، ۷۵ - ۹۳.



۱. مقدمه

روش تحلیل مجموعه‌ی مقادیر تکین^۱ (SSA) یک روش پرکاربرد ناپارامتری در تحلیل سری‌های زمانی محسوب شده و نیازمند وجود شروط مانایی سری زمانی و نرمال بودن مانده‌ها نمی باشد. از طرف دیگر کم بودن تعداد مشاهدات محدودیت جدی برای آن ایجاد نمی کند. به دلیل دارا بودن چنین مزیت‌هایی، SSA کاربردهای وسیعی در بسیاری از شاخه‌های علوم از جمله هواشناسی، ژئوفیزیک، علوم دریایی، پزشکی، بیمه، مهندسی، پردازش تصویر، پردازش سیگنال، اقتصاد سنجی و ریاضیات مالی دارد [۱-۳]. روش SSA به دنبال تجزیه‌ی سری زمانی به سیگنال شامل مولفه‌ی روند^۲، و مولفه‌های هارمونیک^۳ و خطای تصادفی^۴ (نویز) می باشد. به منظور انجام چنین تجزیه‌ای، از روش تجزیه‌ی مقدار تکین^۵ که روشی شناخته شده و پرکاربرد در جبر خطی است، استفاده می شود. پس از آن، سری زمانی به کمک اجزای غیر تصادفی (روند و مولفه‌های فصلی) که سیگنال نامیده می شوند، بازسازی می شود. پس از بازسازی سری می توان به روش بازگشتی^۶، پیش بینی‌هایی در مورد آینده‌ی سری زمانی انجام داد. در این مقاله، نخست مرور مختصری بر روش تحلیل مجموعه مقادیر تکین (SSA) داشته و سپس روش تجزیه طیفی تکین^۷ (SSD) معرفی خواهد شد. در ادامه، یک مطالعه تجربی انجام شده و دقت بازسازی هر دو مدل برای داده‌های واقعی و شبیه‌سازی شده مورد مقایسه قرار می گیرد.

۲. تاریخچه

اولین اثر چاپ شده که می توان به عنوان مبدا پیدایش روش (SSA) در نظر گرفت، به قرن هجدهم برمی گردد (دپرونی، ۱۷۹۵) [۴]. ولی ورود و ظهور اصلی روش SSA با انتشار مقالات برومهد و کینگ (۱۹۸۶ a, ۱۹۸۶ b) بود [۵-۶]. پس از آن رشد و توسعه‌ی روش SSA و به کارگیری آن در علوم مختلف شدت گرفت. اولین کتاب در مورد کاربرد روش SSA در تحلیل سری‌های زمانی توسط السنر و تسونیس (۱۹۹۶) به رشته‌ی تحریر در آمد [۷]. پس از آن مجموعه‌ای از مقالات آمارشناسان دانشگاه سن پترزبورگ در مورد روش SSA ، در کتابی به زبان روسی گردآوری و چاپ شد (دانیلوف و ژیگلجوسکی، ۱۹۹۷) [۸]. بعد از انتشار این کتاب، توجه بسیاری از پژوهشگران در حوزه‌ی تحلیل سری‌های زمانی به روش SSA جلب شد و تا کنون مقالات زیادی در علوم مختلف به این روش پرداخته و کاربردهای فراوانی برای آن یافته‌اند. از دیگر منابع مهم در این زمینه می توان به (گولیاندینا و همکاران، ۲۰۰۱)، (گولیاندینا و ژیگلجوسکی، ۲۰۱۳) و (صانعی و حسنی، ۲۰۱۶) اشاره کرد [۹-۱۱]. برای کسب اطلاعات بیشتر در مورد SSA و دیگر کاربردهای آن به [۱۲-۱۶] مراجعه نمایید.

¹ Singular Spectrum Analysis

² Trend

³ Harmonic components

⁴ Noise

⁵ Singular Value Decomposition

⁶ Recurrent

⁷ Singular Spectrum Decomposition

۳. مروری بر روش تحلیل مجموعه مقادیر تکین (SSA)

در روش SSA با استفاده از دیدگاه سری زمانی کلاسیک فرض می شود که یک سری زمانی ترکیبی از مؤلفه‌های روند (تغییرات درازمدت در میانگین)، تغییرات دوره‌ای، تغییرات فصلی و تغییرات نامنظم است. این روش مؤلفه‌های فوق را با استفاده از تکنیک‌هایی از سیستم‌های دینامیکی، جبر ماتریسی و آمار چند متغیری به طور مجزا استخراج کرده و سپس مؤلفه‌های به دست آمده را مبنای تحلیل‌های بعدی قرار می‌دهد.

۳-۱ مراحل روش SSA

روش SSA از دو مرحله تشکیل شده است: تجزیه و بازسازی. هر کدام از این مراحل نیز شامل دو گام هستند. فرض کنید $Y_N = \{y_1, y_2, \dots, y_N\}$ یک سری زمانی مشاهده شده به طول N بوده و L نیز عددی صحیح، که طول پنجره^۱ نامیده می‌شود، باشد به طوری که $1 < L < N$. مراحل SSA عبارتند از:

۱- تجزیه^۲. این مرحله شامل دو گام است: نشانیدن و تجزیه‌ی مقدار تکین.

الف) نشانیدن^۳: در این گام، ابتدا سری زمانی Y_N را به K زیر سری تبدیل می‌کنیم، به طوری که $K = N - L + 1$. زیر سری i-ام به ازای $i = 1, \dots, K$ ، به صورت $X_i = (y_i, \dots, y_{i+L-1})^T$ تعریف می‌شود. دقت کنید که زیر سری i-ام یک بردار ستونی با L مولفه است و گاهی اوقات بردار L-تاخیر^۴ نامیده می‌شود. سپس این بردارهای L-تاخیری را به صورت ستونی در کنار هم می‌نشانیم تا تشکیل یک ماتریس $L \times K$ دهند. این ماتریس که با X نشان داده می‌شود، ماتریس مسیر^۵ نامیده شده به طوری که عناصر روی قطرهای فرعی ماتریس X با هم برابر بوده و آن را ماتریس هنکل^۶ می‌نامند و عبارت است از:

$$X = \begin{pmatrix} y_1 & y_2 & y_3 & \dots & y_K \\ y_2 & y_3 & y_4 & \dots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \dots & y_N \end{pmatrix}$$

ب) تجزیه‌ی مقدار تکین (SVD): در این گام، تجزیه‌ی مقدار تکین ماتریس مسیر را به دست می‌آوریم. فرض کنید $\lambda_1, \dots, \lambda_L$ مقادیر ویژه‌ی ماتریس XX^T بوده که به صورت نزولی مرتب شده‌اند ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$) و U_1, \dots, U_L نیز بردارهای یکامتعاد ویژه‌ی متناظر با مقادیر ویژه‌ی $\lambda_1, \dots, \lambda_L$ باشند. تعریف می‌کنیم

¹ Window Length

² Decomposition

³ Embedding

⁴ L-Lagged vector

⁵ Trajectory Matrix

⁶ Hankel Matrix

۲- بازسازی^۱. این مرحله شامل دو گام است: گروه بندی و میانگین گیری قطری.

در این صورت SVD ماتریس X به صورت $V_i = X^T U_i / \sqrt{\lambda_i}$ و $d = \max\{i, \lambda_i > 0\}$ نوشته می شود وقتی که $X_i = \sqrt{\lambda_i} U_i V_i^T$ به ازای $i = 1, \dots, d$. سه تایی $(\sqrt{\lambda_i}, U_i, V_i)$ را سه تایی ویژه می نامند.

ج) گروه بندی^۲: در این گام، پس از اینکه SVD ماتریس مسیر به دست آمد، مجموعه‌ی اندیس‌های $\{1, \dots, d\}$ را به m زیر مجموعه‌ی I_1, \dots, I_m افراز می کنیم. فرض کنید $I = \{i_1, \dots, i_p\}$. در این صورت ماتریس X_I متناظر با گروه I به صورت $X_I = X_{i_1} + \dots + X_{i_p}$ تعریف می شود. مثلاً اگر $I = \{1, 4, 5\}$ آنگاه $X_I = X_1 + X_4 + X_5$. بدین ترتیب می توان X_I را به ازای $I = I_1, \dots, I_m$ به دست آورد. در این صورت از SVD ماتریس X نتیجه می گیریم که $X = X_{I_1} + \dots + X_{I_m}$. شیوه‌ی انتخاب مجموعه‌های I_1, \dots, I_m ، گروه بندی سه تایی ویژه نامیده می شود.

د) میانگین گیری قطری^۳: هدف اصلی در این گام، تبدیل هر ماتریس X_{I_j} ، $j = 1, \dots, m$ ، از گام گروه بندی به یک سری زمانی به طول N است. همان طور که در گام نشانیدن اشاره شد، با در اختیار داشتن یک ماتریس هنکل می توان سری زمانی متناظر با آن را به دست آورد؛ ولی ماتریس‌های X_{I_j} که در مرحله‌ی گروه بندی به دست می آیند دارای خاصیت هنکلی نیستند. هنکل سازی ماتریس X_{I_j} به وسیله‌ی میانگین گیری روی عناصر قطرهای فرعی انجام می شود، بدین معنی که همه‌ی عناصر روی یک قطر فرعی را با میانگین عناصر همین قطر فرعی جایگزین می کنیم. بدین ترتیب ماتریس X_{I_j} به یک ماتریس هنکل تبدیل می شود. پس از این تبدیل، زیرسری بازسازی شده‌ی $\{\tilde{y}_1^{(j)}, \dots, \tilde{y}_N^{(j)}\}$ به طول N به دست می آید. بنابراین با توجه به SVD ماتریس X ، سری زمانی اصلی $Y_N = \{y_1, y_2, \dots, y_N\}$ به صورت مجموع m زیر سری بازسازی شده به شکل $y_t = \sum_{j=1}^m \tilde{y}_t^{(j)}$ برای $t = 1, \dots, N$ بازسازی می شود.

۲-۳ پارامترهای روش SSA

به طور کلی روشی یکتایی برای تعیین مقدار L وجود ندارد؛ ولی با این حال، به دلیل اینکه تجزیه‌ی مقدار تکین ماتریس‌های مسیر با طول پنجره‌های L و $K = N - L + 1$ معادل هم هستند، همواره می توان فرض کرد $L \leq N/2$ است. اگر سری زمانی شامل یک مولفه‌ی دوره‌ای با دوره تناوب P باشد، پارامتر L را طوری انتخاب می کنیم که L/P عددی صحیح شود. دومین پارامتر تعداد سه تایی‌های ویژه‌ای است که تبیین کننده‌ی سیگنال بوده و آن را با r نشان داده و پارامتر

¹ Reconstruction

² Grouping

³ Diagonal Averaging

بازسازی^۱ می‌نامند. برای تعیین r می‌توان از اطلاعات مربوط به مرحله i گروه بندی استفاده کرد. مثلاً فرض کنید سه تایی های ویژه i اول و دوم مربوط به روند سری زمانی باشد و سه تایی های ویژه i سوم تا ششم مربوط به مولفه های فصلی سری زمانی بوده و بقیه i سه تایی های ویژه نیز مربوط به نوفه باشند. در این حالت داریم $r = 6$ ، (گولیاندینا و ژینگلجاوسکی، ۲۰۱۳) [۱۷].

۳-۳ پیش بینی به روش بازگشتی

به منظور دستیابی به پیش بینی هایی در مورد آینده i سری زمانی، فرض می‌کنیم که ساختار سری زمانی به وسیله i مدل هایی به نام روابط بازگشتی خطی^۲ (LRR) قابل توصیف است. گوییم سری زمانی $Y_N = \{y_1, y_2, \dots, y_N\}$ به وسیله i یک (LRR) تولید شده است هرگاه ضرایب a_1, \dots, a_d وجود داشته باشند به طوری که

$$y_{i+d} = \sum_{k=1}^d a_k y_{i+d-k} \quad \text{و} \quad 1 \leq N \leq N-d, a_d \neq 0, d < N$$

برای (گولیاندینا و ژینگلجاوسکی، ۲۰۱۳) [۱۷]. برای کسب اطلاعات بیشتر در مورد نظریه و کاربرد LRR در روش SSA به (گولیاندینا و همکاران، ۲۰۰۱) [۲۲] مراجعه کنید. فرض کنید i مجموعه i سه تایی های ویژه i انتخاب شده در مرحله i گروه بندی بوده و $U_i \in \mathbb{R}^L, i \in I$ بردارهای ویژه i متناظر باشند. همچنین فرض کنید $\underline{U}_i \in \mathbb{R}^{L-1}$ برداری شامل اولین $L-1$ مولفه i بردار U_i بوده و π_i آخرین مولفه i بردار U_i باشد و $v^2 = \sum_{i \in I} \pi_i$. گیریم $\tilde{Y}_N = \{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N\}$ نیز سری بازسازی شده به وسیله i باشد. اگر فضای خطی تولید شده توسط بردارهای $U_i \in \mathbb{R}^L, i \in I$ را با K نشان دهیم، آنگاه می‌توان ثابت کرد که آخرین مولفه i بردار $Z = (z_1, \dots, z_L)^T \in K$ یک ترکیب خطی از مولفه های قبلی z_1, \dots, z_{L-1} است (گولیاندینا و همکاران، ۲۰۰۱) [۲۲]. به عبارت دیگر داریم $z_L = a_1 z_{L-1} + \dots + a_{L-1} z_1$ وقتی که بردار ضرایب $R = (a_{L-1}, \dots, a_1)^T$ به صورت $R = 1/1 - v^2 (\sum_{i \in I} \pi_i \underline{U}_i)$ تعریف می‌شود. پیش بینی به روش بازگشتی را می‌توان در دو مرحله i زیر خلاصه کرد:

الف: سری زمانی $Z_{N+h} = \{z_1, \dots, z_{N+h}\}$ را به صورت زیر تعریف کنید:

$$z_i = \begin{cases} \tilde{y}_i & \rightarrow i = 1, \dots, N \\ \sum_{j=1}^{L-1} a_j z_{i-j} & \rightarrow i = N+1, \dots, N+h \end{cases}$$

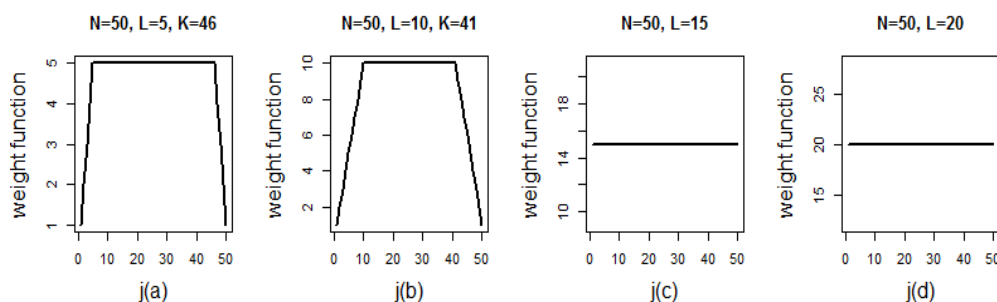
ب: مقادیر z_{N+1}, \dots, z_{N+h} پیش بینی های h گام جلوتر برای سری زمانی Y_N خواهد بود.

¹ Reconstruction Parameter

² Linear Recurrent Relations

۴. روش بهبود یافته تحلیل مجموعه مقادیر تکین^۱ (SSD)

همانطور که می‌دانیم، در مرحله نشانیدن مربوط به روش SSA، ماتریس مسیر طوری طراحی می‌شود که فراوانی مشاهدات بردار سری زمانی اولیه در این ماتریس با یکدیگر برابر نیستند. به طوری که از مشاهده اول تا مشاهده $(L-1)$ -ام، فراوانی مشاهدات بکار رفته در ماتریس مسیر در بازه ۱ تا $(L-1)$ می‌باشد و از مشاهده L -ام تا $(N-L+1)$ -ام فراوانی مشاهدات بکار رفته در ماتریس مسیر برابر L است و پس از آن یعنی از مشاهده $(N-L+2)$ تا مشاهده N -ام فراوانی مشاهدات از $(L-1)$ تا ۱ تغییر می‌کند. اگر تابع وزنی $\omega_j^{L,N}$ ، $j = 1, \dots, N$ ، را به عنوان تابعی از میزان طول پنجره و تعداد کل مشاهدات در نظر بگیریم، به طوری که به ازای هر j ، $\omega_j^{L,N}$ معرف تعداد مشاهدات به کار رفته در ماتریس مسیر باشد، گراف مربوط به ماتریس مسیر شبیه یک دوزنقه خواهد شد (شکل ۱، (a, b))، که بنابر بزرگی اندازه L ، اندازه قاعده کوچک دوزنقه تغییر می‌کند. لذا در روش SSA با توجه به برابر نبودن وزن تمامی مشاهدات در ماتریس مسیر، بیم آن می‌رود که در بازسازی و پیش‌بینی سری زمانی خصوصا در رابطه با مشاهدات مربوط به ابتدا و انتهای سری زمانی، خطا رخ دهد. دلیل این امر آن است که بزرگی مقادیر ویژه، بردارهای ویژه و در نتیجه تجزیه، بازسازی و پیش‌بینی مقادیر آینده سری زمانی ارتباط مستقیم با ماتریس مسیر دارد. در روش SSD تمام مقادیر بردار اولیه در سطرهاى ماتریس تکرار شده و لذا تمامی مشاهدات در ماتریس مسیر از فراوانی یکسانی بر حسب مقدار طول پنجره L برخوردار هستند. در این حالت اگر گراف مربوط به ماتریس مسیر را رسم کنیم شبیه مستطیلی با عرضی برابر L خواهد بود (شکل ۱، (c, d)).



شکل ۱. تابع وزنی $(\omega_j^{L,N})$ به عنوان تابعی از L برای $N = 50$ در $SSA(a, b)$ و $SSD(c, d)$.

در ماتریس مسیر مربوط به روش SSD که با X_{SSD} در ادامه ارائه شده است، بلوک سمت چپ مطابق با ماتریس مسیر اولیه شده در روش SSA و بلوک A داده شده در سمت راست، ماتریس اضافه شده به ماتریس مسیر SSA است که منجر به تکرار تمام مشاهدات بردار اولیه در تمامی سطرهاى ماتریس جدید شده است.

¹ Singular Spectrum Decomposition

$$X_{SSD} = (X|A) = \left(\begin{array}{cccc|ccc} y_1 & y_2 & \dots & y_K & y_{K+1} & \dots & y_N \\ y_2 & y_3 & \dots & y_{K+1} & y_{K+2} & \dots & y_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \dots & y_N & y_1 & \dots & y_{L-1} \end{array} \right)$$

در این حالت ماتریس مسیر بدست آمده هنوز تکین نبوده و در نتیجه باید از روش دیگری برای میانگین‌گیری قطری استفاده کرد. در این مرحله پس از انتخاب r برای بازسازی سری زمانی، اگر z_{ij} را به عنوان مولفه‌های ماتریس بازسازی شده در نظر بگیریم، آنگاه مولفه $t - i$ ام سری نتیجه شده از میانگین تمام z_{ij} هایی به دست می‌آید که $i + j = t + 1 + N$ یا $i + j = t + 1$ وقتی که $i + j < N$ و $i + j = t + 1$ وقتی که $i + j \geq N$. لذا قابل توجه است که در این روش میانگین‌گیری قطری، بر خلاف روش SSA، از تعداد مشاهدات ثابتی برای بازسازی تمامی مولفه‌های سری استفاده می‌کند.

۴-۱ الگوریتم پیش‌بینی به روش بازگشتی در روش SSD (R-SSD)

در روش تعمیم‌یافته SSA که با SSD شناخته می‌شود، ورودی‌های به کار رفته در مراحل پیش‌بینی سری زمانی، به جز ماتریس مسیر به کار رفته در آن، با روش SSA تفاوتی ندارند. از قسمت الف پیش‌بینی به روش بازگشتی در روش SSA، واضح است که ضرایب a_j نقشی اساسی در پیش‌بینی را بازی می‌کنند. همچنین با توجه به اینکه این ضرایب ارتباط مستقیم با بردارهای ویژه U_i ، XX^T دارند، لذا انتظار می‌رود که با بهبود ماتریس مسیر ضرایب به دست آمده بتوانند دقت پیش‌بینی را افزایش دهند.

فرض کنید در ماتریس مسیر X_{SSD} ، $\tilde{\lambda}_1, \dots, \tilde{\lambda}_L$ مقادیر ویژه $X_{SSD} X_{SSD}^T$ و $\tilde{U}_1, \dots, \tilde{U}_L$ بردارهای ویژه متناظر باشند. در این صورت ضرایب بازسازی شده R-SSD به صورت $\tilde{R} = 1/1 - \tilde{v}^2 (\sum_{i \in I} \tilde{\pi}_i \tilde{U}_i)$ محاسبه می‌شوند، به طوری که \tilde{U}_i برداری شامل L-1 بردار اول \tilde{U} ، $\tilde{\pi}_i$ برابر آخرین مولفه بردار \tilde{U}_i و $\tilde{v}^2 = \sum_{i \in I} \tilde{\pi}_i^2$ می‌باشد. سرانجام الگوریتم پیش‌بینی بازسازی شده R-SSD به صورت زیر محاسبه می‌شود:

$$z_i = \left\{ \begin{array}{ll} \tilde{y}_i & \rightarrow i = 1, \dots, N \\ \sum_{j=1}^{L-1} \tilde{a}_j z_{i-j} & \rightarrow i = N+1, \dots, N+h \end{array} \right\}$$

به طوری که \tilde{y}_i ، سری بازسازی شده بر اساس روش SSD می‌باشد و مقادیر z_{N+1}, \dots, z_{N+h} پیش‌بینی‌های h گام جلوتر برای سری زمانی Y_N خواهد بود.

۵. مطالعه تجربی

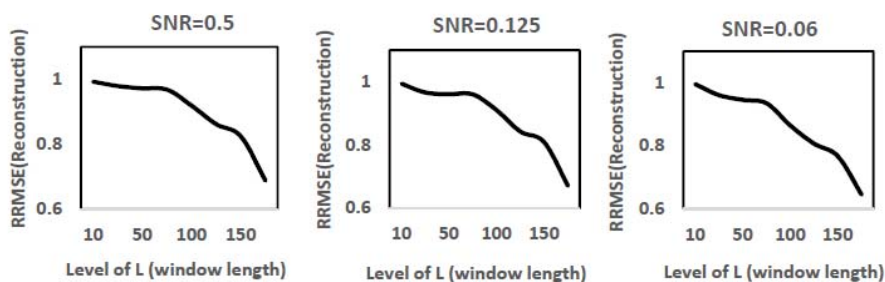
در این بخش بر مبنای داده‌های شبیه‌سازی شده و داده‌های واقعی، کارایی دو روش SSA و SSD با یکدیگر مقایسه می‌شوند. معیار مورد استفاده برای مقایسه عملکرد بازسازی‌های دو روش عبارتند از:

$$RRMSE = \frac{RMSE_{SSD}}{RMSE_{SSA}}, \quad RMAD = \frac{MAD_{SSD}}{MAD_{SSA}}$$

به طوری که $RMSE^1$ جذر میانگین مربعات خطا و MAD^2 میانگین قدر مطلق خطا بوده و اگر $RRMSE < 1$ یا $RMAD < 1$ باشد، نشان دهنده آن است که میزان خطای تولید شده از روش SSD کمتر از روش SSA می‌باشد و بالعکس.

۵-۱. داده‌های شبیه‌سازی شده

مثال ۱. سری سینوسی ساده $y_t = \sin(2\pi t / 3) + \varepsilon_t$ را برای $t = 1, \dots, 300$ در نظر می‌گیریم. از ۲۰۰ مشاهده اول برای بازسازی سری و از ۱۰۰ مشاهده بعدی برای پیش‌بینی سری استفاده خواهد شد. با توجه به اینکه در این مدل رتبه ماتریس مسیر برابر با ۲ است لذا، $r = 2$ انتخاب می‌شود. برای بررسی تاثیر میزان خطا در مقایسه‌ها، نسبت واریانس سیگنال به واریانس خطا (SNR) در سه سطح ۰/۵، ۰/۱۲۵ و ۰/۰۶ در نظر گرفته شد. همچنین در تولید نمونه‌ها از اندازه نمونه $N = 400$ استفاده شده است. نتایج RRMSE در بازسازی سری زمانی مربوطه در شکل ۲ نشان می‌دهد که در تمامی سطوح طول پنجره L ، روش SSD بر روش SSA غالب بوده و دارای RMSE کمتری می‌باشد. همچنین با افزایش طول پنجره میزان RRMSE نیز کاهش یافته و لذا برای طول پنجره‌های بزرگتر روش SSD کارا تر می‌باشد.

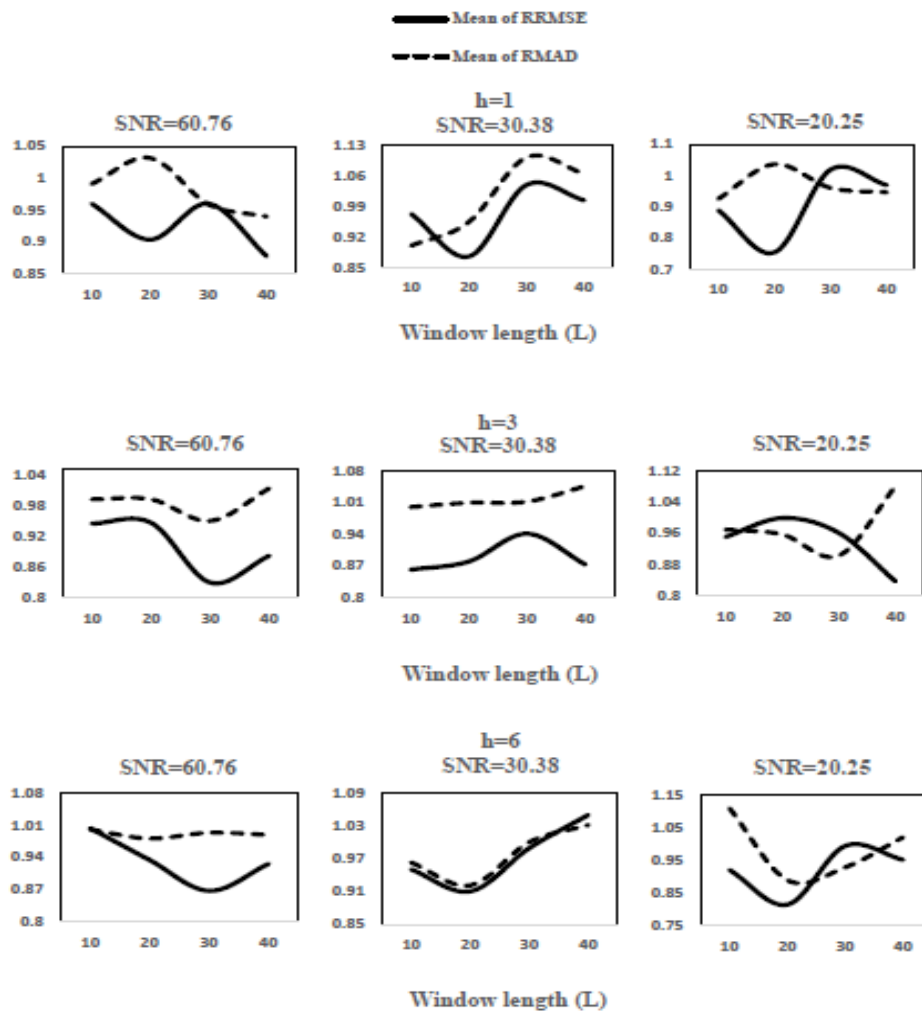


شکل ۲. نمودارهای RRMSE در بازسازی مثال ۱.

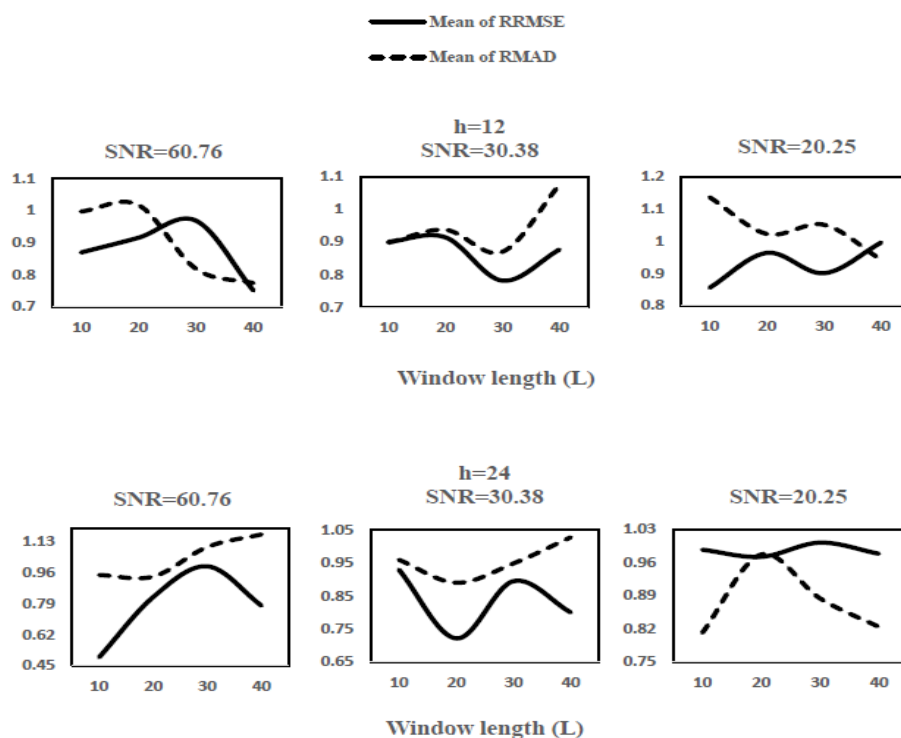
¹ Root Mean Square Error

² Mean Absolute Error

شکل‌های ۳ و ۴، مقادیر RRMSE و RMAD را برای مقادیر $h=1,3,6,12,24$ نشان می‌دهد. نمودارهای مربوطه نشان می‌دهد که مینیمم مقدار RRMSE و RMAD زمانی رخ می‌دهد که طول پنجره $L=40$ ، $h=12,24$ و $SNR=60.76$ باشند. همچنین می‌توان از شکل‌های ۳ و ۴ نتیجه گرفت که به طور کلی مقادیر RRMSE و RMAD کمتر از یک بوده که نمایانگر عملکرد بهتر روش SSD نسبت به روش اصلی SSA می‌باشد. همچنین کمترین مقدار RRMSE زمانی رخ می‌دهد که $L=10$ و $h=24$ باشند.



شکل ۳. نمودارهای RRMSE و RMAD برای پیش بینی مثال ۱ و ($h=1,3,6$)



شکل ۴. نمودارهای RRMSE و RMAD برای پیش‌بینی مثال ۱ و $(h=12,24)$

به منظور مقایسه دقت پیش‌بینی دو روش SSD و SSA کلاسیک از آزمون دقت پیش‌بینی کلموگروف-اسمیرنوف^۱ (KSPA) استفاده می‌کنیم. به این منظور اگر $F_{\varepsilon_{i+h}}^{m_1}$ و $F_{\varepsilon_{i+h}}^{m_2}$ به ترتیب نمایانگر توابع توزیع تجربی خطاهای پیش‌بینی از دو روش باشند به طوریکه $\mathcal{E}_{i+h}^{m_1}$ و $\mathcal{E}_{i+h}^{m_2}$ قدر مطلق و یا مربع خطاهای پیش‌بینی به دست آمده از دو روش m_1 و m_2 باشند، آزمون دو طرفه KSPA شرط هم‌توزیعی دو روش m_1 و m_2 را آزمون می‌کند و همچنین در آزمون یک طرفه راست KSPA، با رد فرض صفر نتیجه می‌شود که با $(1-\alpha)\%$ اطمینان مدل m_1 ، خطای تصادفی کمتری نسبت به مدل m_2 دارد. به عبارت دیگر:

$$\left\{ \begin{array}{l} H_0 : F_{\varepsilon_{i+h}}^{m_1}(x) = F_{\varepsilon_{i+h}}^{m_2}(x) \\ H_1 : F_{\varepsilon_{i+h}}^{m_1}(x) \neq F_{\varepsilon_{i+h}}^{m_2}(x) \end{array} \right\}, \quad \left\{ \begin{array}{l} H_0 : F_{\varepsilon_{i+h}}^{m_1}(x) \leq F_{\varepsilon_{i+h}}^{m_2}(x) \\ H_1 : F_{\varepsilon_{i+h}}^{m_1}(x) > F_{\varepsilon_{i+h}}^{m_2}(x) \end{array} \right\}$$

¹ Kolmogorov-Smirnov Predictive Accuracy

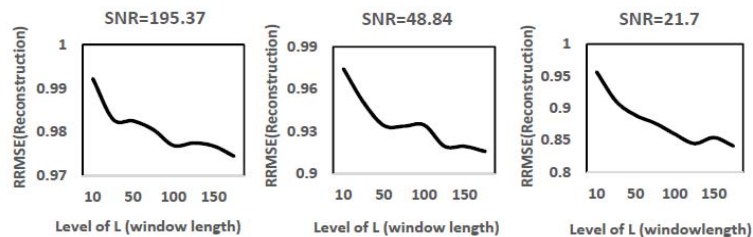
در این حالت با در نظر گرفتن $m_1 = SSD$ و $m_2 = SSA$ برای مثال ۱، نتایج آزمون KSPA در جدول ۱، برای $L=10$ و $h=24$ گزارش شده است.

جدول ۱. نتایج آزمون KSPA برای $h=24$ در مثال ۱.

آزمون	دو طرفه	یک طرفه راست
KSPA	< 0.01	< 0.01

نتایج جدول ۱ نشان می‌دهد که با توجه به آزمون دو طرفه، هم توزیعی خطاهای پیش‌بینی دو روش SSD و SSA اصلی در سطح اطمینان ۰/۹۵ رد می‌شود. همچنین نتایج آزمون یک طرفه نشان می‌دهد که با ۰/۹۰ اطمینان روش SSD خطاهای تصادفی کمتری را نسبت به روش SSA در پیش‌بینی سری زمانی دارا می‌باشد.

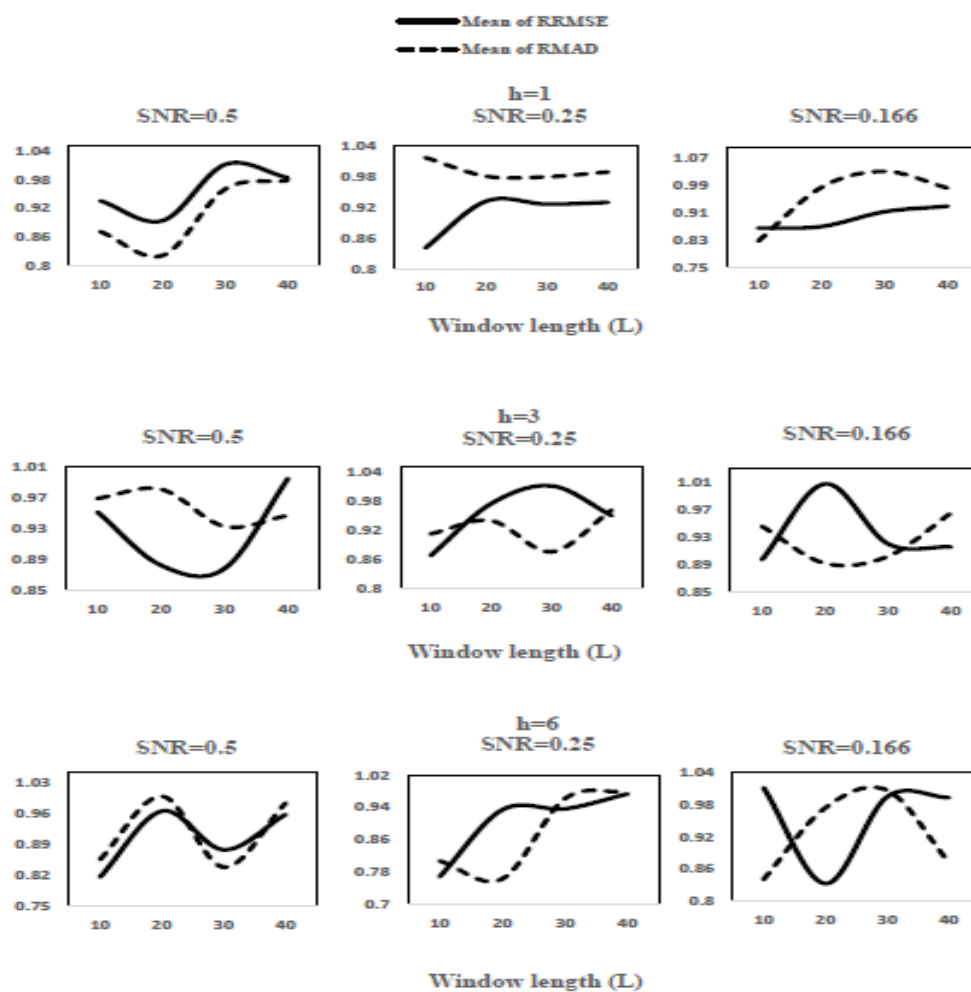
مثال ۲. سری نمایی ساده $y_t = \exp(0.01t) \times \varepsilon_t$ را برای $t = 1, \dots, 400$ در نظر می‌گیریم. با توجه به اینکه در این مدل رتبه ماتریس مسیر برابر با ۱ است لذا، $r=1$ انتخاب می‌شود. برای بررسی تاثیر میزان خطا در مقایسه‌ها، نسبت واریانس سیگنال به واریانس خطا (SNR) در سه سطح ۱۹۵/۳۷، ۴۸/۸۴ و ۲۱/۷ در نظر گرفته شد. همچنین در تولید نمونه‌ها از اندازه نمونه $N=400$ استفاده شده است. نتایج RRMSE در بازسازی سری مربوطه در شکل ۵ نشان داده شده است. نتایج به دست آمده نشان می‌دهد که در تمامی سطوح L، روش SSD بر روش SSA غالب بوده و دارای RMSE کمتری می‌باشد. همچنین نتایج نشان می‌دهد با افزایش L میزان RRMSE نیز کاهش می‌یابد. لذا برای طول پنجره‌های بزرگتر روش جدید SSD بسیار کارا تر می‌باشد.



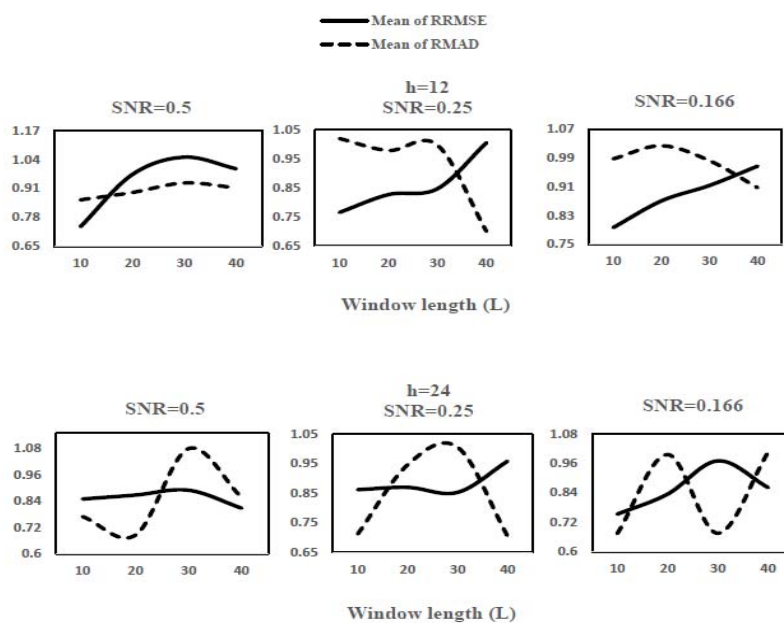
شکل ۵. نمودارهای RRMSE در بازسازی مثال ۲.

شکل‌های ۶ و ۷، مقادیر RRMSE و RMAD را برای مقادیر $h=1,3,6,12,24$ نشان می‌دهد. نمودارهای مربوطه نشان می‌دهد که مینیمم مقدار RRMSE و RMAD زمانی رخ می‌دهد که طول پنجره $L=10$ ، $h=24$ و $SNR=0.166$ و همچنین

زمانی که طول پنجره $L=20$ ، $h=24$ و $SNR=0.5$ باشند. همچنین می‌توان از شکل‌های ۶ و ۷ می‌توان نتیجه گرفت که به طور کلی مقادیر RRMSE و RMAD کمتر از یک بوده که نمایانگر عملکرد بهتر روش SSD نسبت به روش اصلی SSA است.



شکل ۶. نمودارهای RRMSE و RMAD برای پیش بینی مثال ۲ و (h=1,3,6)



شکل ۷. نمودارهای RRMSE و RMAD برای پیش بینی مثال ۲ و (h=12,24)

۲-۵. داده واقعی

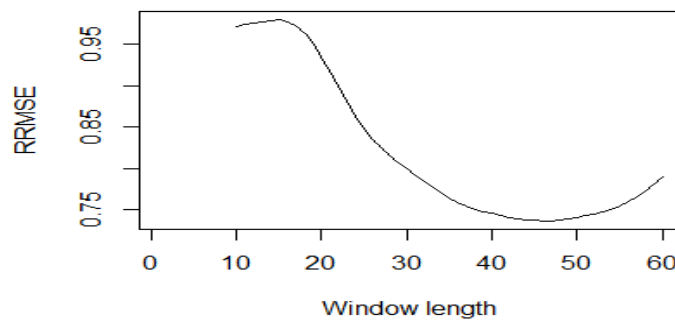
مثال ۳. داده‌های قیمت نفت اوپک (دلار به ازای هر بشکه) که از فروردین ۱۳۸۸ لغایت فروردین ۱۳۹۸ که در بانک داده‌های اقتصادی و مالی^۱ ارایه شده، مورد بررسی قرار می‌گیرد. بر اساس نمودار این داده‌ها در شکل ۸، روند تغییرات قیمت در بازه ۱۳۸۸ تا ۱۳۹۱ افزایشی و پس از آن سیر کاهشی تا سال ۱۳۹۵ و سپس دارای روند افزایشی تا سال ۱۳۹۸ می‌باشد.



شکل ۸. نمودار سری زمانی قیمت نفت اوپک (دلار به ازای هر بشکه)

¹ <https://databank.mefa.ir>

با به کارگیری اولین سه تایی ویژه در بازسازی این سری زمانی و با بکار بردن طول پنجره‌های مختلف، مقادیر RRMSE محاسبه شده و مقادیر آن در شکل ۹ نشان داده شده است. همانطور که دیده می‌شود برای تمامی مقادیر طول پنجره نسبت RRMSE به دست آمده کمتر از ۱ می‌باشد که نشان دهنده برتری روش جدید SSD نسبت به روش پایه SSA در بازسازی سری α است.



نتایج RRMSE در پیش‌بینی سری زمانی، برای سطوح مختلف طول پنجره، L ، و با به کارگیری اولین سه تایی ویژه اول، در جدول ۲ نشان داده شده است. نتایج به دست آمده نشان می‌دهد که به طور کلی، در تمامی سطوح L و h روش SSD دارای خطای کمتری نسبت به روش SSA می‌باشد و لذا کارایی SSD نسبت به SSA تایید می‌شود. نتایج این جدول حاکی از آن است که کمترین مقادیر RRMSE مربوط به زمانی است که طول پنجره رو به افزایش است به طوری که کمترین مقدار $0/32$ زمانی رخ می‌دهد که $L=45$ و $h=24$ است.

جدول ۲. مقادیر RRMSE در پیش‌بینی سری قیمت نفت اوپک

L \ h	۱۵	۲۰	۳۰	۴۰	۴۵
۱	۰/۷۵	۰/۷۷	۰/۸۵	۰/۵۹	۰/۵۶
۳	۰/۷۹	۰/۷۴	۰/۶۱	۰/۵	۰/۴۴
۶	۰/۷۶	۰/۷۱	۰/۶	۰/۵	۰/۴۳
۱۲	۰/۷۲	۰/۶۷	۰/۶۱	۰/۵	۰/۴۲
۲۴	۰/۷۲	۰/۷۳	۰/۶۱	۰/۳۸	۰/۳۲

در ادامه در جدول‌های ۳ و ۴ نتایج آزمون KSPA، برای سطح $L=40$ ، و همچنین $h=6,12,24$ ، ارایه شده است.

جدول ۳. نتایج آزمون KSPA برای $h=24$ و $L=15$ در مثال ۳.

آزمون	دو طرفه	یک طرفه راست
KSPA	۰/۰۰	۰/۰۰۱

نتایج جدول ۳ نشان می‌دهد که با توجه به آزمون دو طرفه، هم توزیعی خطاهای پیش‌بینی دو روش SSD و SSA اصلی در سطح اطمینان ۰/۹۵ رد می‌شود. همچنین نتایج آزمون یک طرفه نشان می‌دهد که با ۰/۹۵٪ اطمینان روش SSD خطاهای تصادفی کمتری را نسبت به روش SSA در پیش‌بینی سری زمانی دارا می‌باشد.

جدول ۴. نتایج آزمون KSPA برای $h=6,12$ و $L=40$ در مثال ۳.

آزمون	دو طرفه	یک طرفه راست
KSPA	۰/۰۱	۰/۰۰

نتایج جدول ۴ نشان می‌دهد که با توجه به آزمون دو طرفه، هم توزیعی خطاهای پیش‌بینی دو روش SSD و SSA اصلی در سطح اطمینان ۰/۹۵ رد می‌شود. همچنین نتایج آزمون یک طرفه نشان می‌دهد که با ۰/۹۵٪ اطمینان روش SSD خطاهای تصادفی کمتری را نسبت به روش SSA در پیش‌بینی سری زمانی دارا می‌باشد.

۶. نتیجه‌گیری

در این مقاله نخست روش تحلیل مقادیر تکین (SSA) و سپس روش بهبود یافته SSD بر پایه روش SSA معرفی شده است. برای مقایسه میزان دقت سری بازسازی شده و دقت پیش‌بینی توسط این دو روش، ابتدا از داده‌های شبیه‌سازی شده و سپس داده‌های مربوط به قیمت نفت اوپک در بازه زمانی اول فروردین ۱۳۸۸ لغایت اول فروردین ۱۳۹۸ استفاده شده است. نتایج حاصل شده در بازسازی و پیش‌بینی سری‌ها، با استفاده از هر دو مجموعه داده‌های شبیه‌سازی شده و واقعی نشان داد که استفاده از روش SSD نسبت به روش پایه SSA مقدار ریشه میانگین مربع خطا RRMSE و میانگین قدرمطلق خطا RMAD را به مقدار قابل توجهی کاهش می‌دهد. از این رو روش SSD می‌تواند به عنوان یک انتخاب بهتر نسبت به روش پایه SSA در تحلیل سری‌های زمانی به کار رود.

References

1. H. Hassani, Singular spectrum analysis: methodology and comparison. *J. Data Sci.* **5** (2) (2007), 239–257.
2. S. Sanei, M. Ghodsi, and H. Hassani, An adaptive singular spectrum analysis approach to murmur detection from heart sounds, *Med. Eng. Phys.* **33** (3) (2011), 362–367.
3. H. Xie, T. Guo, B. Sivakumar, A. W. Liew, and S. Dokos, Symplectic geometry spectrum analysis of nonlinear time series, *Proc. R. Soc. A* **470** 20140409, (2014).
4. G. de Prony, Essai expérimental et analytique sur les lois de la dilatabilité des fluids élastiques et sur celles de la force expansive de la vapeur de l'eau et la vapeur de l'alkool à différentes températures. *J de l'Ecole Polytechnique* **1**(2), (1795), 24–76.
5. D. S. Broomhead, and G. P. King, Extracting qualitative dynamics from experimental data. *Physica D*, **20** (1986a), 217-236.
6. D. S., Broomhead, G. P. King, and E. R. Pike, On the qualitative analysis of experimental dynamical systems. In: Sarkar S (ed), *Nonlinear Phenomena and Chaos*. Adam Hilger, Bristol, (1986b), 113-144.
7. J. B., Elsner and A. A. Tsonis, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*, New York: Plenum Press, 1996.
8. D. Danilov, and A. Zhigljavsky, (Eds.), *Principal Components of Time Series: the 'Caterpillar' method*, University of St. Petersburg Press. (In Russian.), 1997.
9. N. Golyandina, V. Nekrutkin and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and related techniques*. Chapman & Hall/CRC, 2001.
10. N. Golyandina, and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Springer Briefs in Statistics. Springer-Verlag, 2013.
11. S. Sanei and H. Hassani, *Singular Spectrum Analysis of Biomedical Signals*, Taylor Francis/CRC, 2016.
12. H. Hassani, S. Heravi and A. Zhigljavsky, Forecasting European Industrial Production with Singular Spectrum Analysis. *International Journal of Forecasting* **25** (2009), 103– 118.
13. H. Hassani, A. Dionisio, and M. Ghodsi, The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. *Nonlinear Analysis: Real World Applications* **11** (2010), 492–502.
14. A. Zhigljavsky, H. Hassani and H. Heravi, Forecasting European Industrial Production with Multivariate Singular Spectrum Analysis, *International Institute of Forecasters*, 2008–2009 SAS/IIF Grant, (2009).
15. H. Hassani and D. D. Thomakos, A review on Singular Spectrum Analysis for economic and financial time series, *Statistics and its interface*, **3** (2010), 377–397.
16. H. Hassani, A. S. Soofi and A. Zhigljavsky, Predicting inflation dynamics with singular spectrum analysis, **176** (3) (2013), 743-760.

17. N. Golyandina and A. Zhigljavsky, Singular spectrum analysis for time Series. Springer, (2013).
18. H. Hassani, R. Mahmoudvand and M. Zokaei, Separability and window length in singular spectrum analysis. *Comptes Rendus Mathematique*, **349** (17-18) (2011), 987–990.
19. R. Mahmoudvand, N. Najari, and M. Zokaei, On the optimal parameters for reconstruction and forecasting in singular spectrum analysis. *Communications in Statistics- Simulation and Computation* **42** (4) (2013), 860–870.
20. H. Hassani, A. Webster, E. S. Silva and S. Heravi, Forecasting U.S. tourist arrivals using optimal singular spectrum analysis. *Tourism Management* **46** (2015), 322–335.
21. N. Alharbi, and H. Hassani, A new approach for selecting the number of the Eigen values in singular spectrum analysis. *Journal of the Franklin Institute* **353** (2016), 1–16.
22. N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, Analysis of Time Series Structure: SSA and related techniques. London: Chapman & Hall/CRC, (2001).