Journal of Statistical Sciences, Autumn and Winter, 2022
Vol. 16, No. 2, pp 373-395
DOI: 10.29252/jss.16.2.373

Improving the Recurrent Forecasting of Singular Spectrum Analysis in Structural Time Series Models Using Data Filtration and Weighting Algorithm

Zabihi Moghadam, R.[1], Yarmohammadi, M.[1], Hassani, H.[2], Nasiri, P.[1]
[1]Department of Statistics, Payame Noor University, Tehran, Iran.
[2]Institute for Energy Management and Planning, University of Tehran, Tehran, Iran.

Corresponding author: M. Yarmohammadi, masyar@pnu.ac.ir

Received: 8/1/2021 **Revised:** 3/11/2022 **Accepted and Published Online:** 13/11/2022.

Introduction

Among different forecasting methods, the singular spectrum analysis (SSA) is a powerful nonparametric technique with both filtering and forecasting capabilities. The SSA method breaks down the observational series into two components, i.e. noise and signal, using the eigenvalues and eigenvectors of the trajectory matrix. Then it calculates the forecast by reconstructing the original series using the signal component and a recursive linear relationship from the original series. Since the real data are not noise-free, the linear recurrent relation (LRR) coefficients obtained from the eigenvectors of the trajectory matrix are also contaminated with noise and lead to a reduction in the forecast accuracy. Thus, to improve the performance of the recursive forecasting methods, some actions have been taken into account. In this paper, a hybrid method is proposed to improve the recursive forecasting performance of SSA using the Kalman filter algorithm. Then, the effectiveness of this method was compared with the reconstructed SSA-R, the SSA weighting algorithm, and the basic SSA method using the root mean square error criterion.

Material and Methods

For a time series y_t with constant window length L, there are L-1, SSA-R co-

efficient $\phi_1, \ldots, \phi_{L-1}$ are obtained from eigenvectors of the trajectory matrix. Therefore if the observational series includes noise, the estimated coefficients $\phi_1, \ldots, \phi_{L-1}$ are not accurate and will reduce the accuracy of the prediction. We will use the state-space equations and Kalman filter algorithm to reduce noise, improve our prediction, and define the KF-SSA-R method. Another approach that can be used to improve the recursive prediction of SSA method is the reconstructed SSA. In this method, the prediction coefficients are obtained from the re-execution of the SSA method for the reconstructed series. In the last method, to improve the prediction of SSA method, we use a weighting algorithm. In this method, to obtain the coefficients, 2/3 observations are used to calculate the weights.

Results and Discussion
To investigate the effect of data refinement, the proposed methods have been compared using simulation studies and for the gas consumption in England. For the impact of noise on the performance of the methods, the signal-to-noise ratio is used for different window lengths and prediction horizons. To compare the introduced methods, the $RRMSE$ is used.

Conclusion
The simulation studies and real data results show that the $KF - SSA - R$ method compared to the basic SSA, Reconstructed SSA and weighted SSA have better performance, especially when the window length is small. When the window length is considerable, the weighted SSA method is more efficient than the primary SSA method in the data obtained from the structural model. Compared with the reconstructed SSA and original SSA methods, when the window length is small, we conclude that the reconstructed SSA method performs better.

Keywords: Forecasting, Kalman Filter, Singular Spectrum Analysis, Recurrent Forecasting.

Mathematics Subject Classification (2010): 62M10, 62M20.

©The Author(s). The Publisher is Iranian Statistical Society.
This is an open access article distributed under the terms and conditions of (CC BY-NC 4.0)

مجله علوم آماری، پاییز و زمستان ۱۴۰۱
جلد ۱۶، شماره ۲، ص ۳۷۳ – ۳۹۵
DOI: 10.29252/jss.16.2.373
مقاله پژوهشی

بهبود پیش‌بینی بازگشتی تحلیل مجموعه مقادیر تکین در مدل‌های سری زمانی ساختاری با استفاده از پالایش داده‌ها و الگوریتم وزنی

رضا ذبیحی مقدم[۱]، مسعود یارمحمدی [۱]، حسین حسنی[۲]، پرویز نصیری[۱]

[۱]گروه آمار، دانشگاه پیام نور، تهران، ایران
[۲]موسسه پژوهش در مدیریت و برنامه‌ریزی انرژی، دانشگاه تهران، تهران، ایران

نویسنده مسئول : مسعود یارمحمدی، masyar@pnu.ac.ir

تاریخ دریافت: ۱۴۰۰/۱۰/۱۸ تاریخ بازنگری: ۱۴۰۱/۸/۱۲ تاریخ پذیرش و انتشار: ۱۴۰۱/۸/۱۴

چکیده: تحلیل مجموعه مقادیر تکین (SSA) یکی از روش‌های ناپارامتری توانمند برای کاهش نویز، مدل‌سازی و پیش‌بینی سری‌های زمانی است. این روش نیازمند فرض‌های محدود کننده مانند نرمال بودن خطاها، کم بودن تعداد دادها و مانایی سری زمانی نیست. در سال‌های اخیر تلاش‌های مستمری از جانب محققان در جهت بهبود این روش خصوصاً در زمینه پیش‌بینی سری‌های زمانی صورت گرفته است. در این مقاله روش ترکیبی برای بهبود پیش‌بینی روش SSA با استفاده از الگوریتم پالایش کالمن در مدل‌های ساختاری معرفی می‌شود. سپس کارایی این روش و چند روش تعمیم‌یافته SSA با روش SSA پایه با استفاده از معیار جذر میانگین توان دوم خطا مورد مقایسه قرار می‌گیرد. برای انجام این مقایسه‌ها، از داده‌های شبیه‌سازی شده و داده‌های واقعی میزان مصرف گاز در انگلستان استفاده شده است. نتایج نشان می‌دهند که روش ترکیبی معرفی شده از دقت بیشتری نسبت به سایر روش‌ها برخوردار است.

واژه‌های کلیدی: پالایش کالمن، فضای حالت، تحلیل مجموعه مقادیر تکین، پیش‌بینی بازگشتی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62M10، 62M20.

©نویسند(گان). ناشر انجمن آمار ایران است.
این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

۱ مقدمه

یکی از موضوعات مهم در تحلیل مشاهدات سری زمانی پیش‌بینی مشاهدات جدید است. روش‌های مختلفی برای پیش‌بینی وجود دارد. یکی از آن‌ها روش ناپارامتری تحلیل مجموعه مقادیر تکین [۱] (SSA) با قابلیت صافی کردن و پیش‌بینی مناسب بوده و شامل دو شیوه‌ی پیش‌بینی بازگشتی (SSA-R) و پیش‌بینی برداری (SSA-V) است. روش SSA برخلاف سایر روش‌های کلاسیک پیش‌بینی نیازمند نرمال بودن خطاها، برقراری شروط مانایی و زیاد بودن تعداد داده‌ها نیست. معرفی روش SSA توسط برومهید و کینگ (۱۹۸۶) صورت گرفته و به تدریج تلاش‌های مستمری برای بهبود آن انجام شده است (حسنی و همکاران ، ۲۰۱۱؛ آیدین و همکاران ، ۲۰۱۱؛ بایل و همکاران ، ۲۰۱۴؛ لیو و همکاران ، ۲۰۱۴؛ وانگ و همکاران ، ۲۰۱۵؛ محمودوند و رودریگز ، ۲۰۱۶).

روش SSA ابتدا سری زمانی را در قالب یک ماتریس خاص، که به آن ماتریس مسیر [۲] گویند، در می‌آورد. سپس طی انجام مراحلی سری مشاهدات را به دو مؤلفه نویز و سیگنال تقسیم کرده و پیش‌بینی را با استفاده از بازسازی سری اصلی از مؤلفه سیگنال و یک رابطه خطی بازگشتی (LRR) از این سری محاسبه می‌کند. ضرایب این رابطه خطی بازگشتی از بردارهای ویژه ماتریس مسیر بدست می‌آیند. حال با توجه به اینکه مشاهدات سری زمانی اغلب دارای نویز هستند، در نتیجه ضرایب رابطه خطی بازگشتی بدست آمده از بردارهای ویژه ماتریس مسیر نیز آلوده به نویز شده و منجر به کاهش دقت پیش‌بینی می‌شوند. بنابراین به منظور بهبود کارایی روش‌های پیش‌بینی بازگشتی نیازمند انجام اقدامات و اصلاحات مناسب می‌باشیم. در این راستا حسنی و همکاران (۲۰۱۷) الگوریتمی جدیدی را به نام SSA-R بازسازی شده معرفی کردند که در آن ضرایب SSA-R از بردارهای ویژه سری بازسازی شده بدست می‌آید. آنها کارایی این روش را با روش SSA پایه بر اساس معیار جذر میانگین توان دوم خطا (RMSE) با استفاده از مطالعات شبیه‌سازی شده و داده‌های سری زمانی واقعی مقایسه کرده و نشان دادند که روش بازسازی شده SSA-R برای مقادیر کوچک طول پنجره [۳] (L) کارایی بیشتری نسبت به روش SSA پایه دارد. محمودوند و رودریگز (۲۰۱۷) مدلی جدید بر پایه روش SSA-R را معرفی کردند که از تعداد ضرایب بهینه کمتری نسبت به روش SSA-R پایه در رابطه بازگشتی خطی LRR استفاده می‌کند. در مطالعه دیگری، روش جدیدی برای استخراج سیگنال در مدل‌های نمایی با نویز زیاد توسط حسنی و کلانتری (۲۰۱۸) ارائه شده است. اخیراً، کلانتری و همکاران (۲۰۲۰) از یک الگوریتم وزن دهی به ضرایب LRR استفاده کرده و کارایی آن را با روش SSA پایه براساس ملاک RMSE مقایسه کرده و برتری الگوریتم وزن‌دهی را نسبت به روش پیش‌بینی SSA-R برای سری‌های نمایی و سینوسی نتیجه گرفتند.

در بخش ۲ به منظور بهبود کارایی پیش‌بینی بازگشتی روش SSA، به معرفی مدل‌های ساختاری، مدل فضای حالت و الگوریتم پالایش کالمن می‌پردازیم. در بخش ۳ روش SSA و روش‌های مختلف پیش‌بینی بازگشتی SSA با استفاده از الگوریتم پالایش کالمن، SSA بازسازی شده (حسنی و همکاران ، ۲۰۱۷) و SSA وزنی (کلانتری و همکاران ، ۲۰۲۰) بیان می‌شود. در بخش ۴ با انجام شبیه‌سازی و استفاده از داده‌های واقعی کارایی

[1] Singular Spectrum Analysis
[2] Trajectory matrix
[3] Window Length

روش‌های پیشنهادی و سایر روش‌ها با استفاده از معیار جذر میانگین توان دوم خطا مورد مقایسه قرار می‌گیرد.

۲ مدل‌های ساختاری، فضای حالت و الگوریتم پالایش کالمن

مدل‌های ساختاری: در یک مدل ساختاری مجموعه‌ای از مؤلفه‌های مشاهده نشده در مدل سری زمانی قرار می‌گیرند که هر کدام تفسیری خاص از عوامل تأثیرگذار را نشان می‌دهند. استفاده از مدل‌های ساختاری باعث می‌شود که سری مشاهدات را به مؤلفه‌های غیر قابل مشاهده که باعث درک بهتر مشخصات دینامیکی سری شده، تجزیه نموده و پایه‌ای مؤثر برای تعدیلات فصلی ایجاد کرد. در ادامه می‌توان پیش‌بینی‌های بهینه‌ای را با استفاده از الگوریتم‌های خاص به دست آورد.

فرم مدل‌های ساختاری برای مشاهدات $\{y_t : t = ۱, \dots, n\}$ به‌صورت

$$y_t = \mu_t + \psi_t + \gamma_t + \epsilon_t, \qquad \epsilon_t \sim N(۰, \sigma_\epsilon^۲)$$

تعریف می‌شود، که در آن y_t مشاهده در لحظه t، μ_t مؤلفه روند، ψ_t مؤلفه دوره، γ_t مؤلفه اثرات فصلی و ϵ_t مؤلفه تغییرات نامنظم هستند و می‌توانند به صورت معادلات

$$\begin{aligned} \mu_t &= \mu_{t-۱} + \beta_{t-۱} + \eta_t, & \eta_t &\sim N(۰, \sigma_\eta^۲) \\ \beta_t &= \beta_{t-۱} + \zeta_t, & \zeta_t &\sim N(۰, \sigma_\zeta^۲) \end{aligned}$$

$$\begin{pmatrix} \psi_t \\ \psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-۱} \\ \psi_{t-۱}^* \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^* \end{pmatrix}$$

$$\sum_{j=۰}^{s-۱} \gamma_{t-j} = \omega_t, \qquad \omega_t \sim N(۰, \sigma_\omega^۲)$$

بیان شوند. در معادلات فوق β_t شیب، λ_c فرکانس مؤلفه دوره، ρ عامل میرایی دوره، s تعداد فصل‌ها در یک دوره، η_t، ζ_t، κ_t، κ_t^* و ω_t عوامل اغتشاشی و مستقل از یکدیگر و همچنین مستقل از اغتشاش ϵ_t هستند. همچنین مؤلفه‌های روند، دوره و اثرات فصلی مستقل از یکدیگر هستند.

مدل‌های فضای حالت: این مدل‌ها اولین بار توسط کالمن (۱۹۶۰) معرفی شده و بسیاری از مدل‌های خطی و غیرخطی را در بر می‌گیرند. مهم‌ترین الگوریتم در مدل‌های فضای حالت الگوریتم پالایش کالمن است که برای برآورد پارامترها و پیش‌بینی نیازی به شروط مانایی و وارون‌پذیری ندارد. مدل فضای حالت برای مشاهدات سری زمانی $\{y_t : t = ۱, \dots, N\}$ شامل یک معادله اندازه و یک معادله انتقال مرتبط به معادله اندازه از داده‌های مشاهده

شده به یک بردار حالت است، که این بردار حالت در معادله انتقال از یک فرآیند مارکوف به دست می‌آید. معادلات فضای حالت برای $t = ۱, \ldots, N$ به صورت

$$y_t = Z_t \alpha_t + \epsilon_t, \qquad \epsilon_t \sim N(۰, \sigma_\epsilon^۲) \tag{۱}$$

$$\alpha_t = \mathbf{T}_t \alpha_{t-۱} + \eta_t, \qquad \eta_t \sim N(۰, \mathbf{Q}_t) \tag{۲}$$

تعریف می‌شوند، که در آن y_t متغیر مشاهده شده، Z_t یک بردار $۱ \times m$، $\boldsymbol{T}_t$ ماتریس انتقال با بعد ماتریسی $m \times m$ ، α_t بردار حالت $m \times ۱$ ، ϵ_t و η_t به ترتیب اغتشاش‌های معادله اندازه و معادله انتقال هستند. معادلات فضای حالت با بردار حالت اولیه $\alpha_۰$ با میانگین $E(\alpha_۰) = a_۰$ و واریانس $Var(\alpha_۰) = \boldsymbol{p}_۰$ شروع می‌شود. در معادلات فضای حالت اغتشاش‌های ϵ_t و η_t از یکدیگر مستقل هستند. این اغتشاش‌ها از بردار حالت اولیه $\alpha_۰$ نیز ناهمبسته هستند. یعنی

$$E(\epsilon_t \eta_s') = ۰, \qquad E(\eta_t \alpha_۰') = ۰, \qquad E(\epsilon_t \alpha_۰') = ۰, \qquad s, t = ۱, \ldots, n$$

در معادلات فضای حالت ماتریس‌های $\boldsymbol{T}_t$، $\boldsymbol{Q}_t$ و Z_t ماتریس‌های سیستمی نامیده می‌شوند. اگر این ماتریس‌ها برحسب زمان ثابت باشند، مدل زمان–همگن نامیده می‌شود که مدل‌های مانا را در بر می‌گیرند.

مثال ۱. ساده‌ترین مدل ساختاری مدل روند خطی موضعی با معادلات

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, && \epsilon_t \sim N(۰, \sigma_\epsilon^۲) \\ \mu_t &= \mu_{t-۱} + \beta_{t-۱} + \eta_t, && \eta_t \sim N(۰, \sigma_\eta^۲) \\ \beta_t &= \beta_{t-۱} + \zeta_t, && \zeta_t \sim N(۰, \sigma_\zeta^۲) \end{aligned}$$

است، که در آن مدل فضای حالت عبارت است از

$$\alpha_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}, Z_t = \begin{pmatrix} ۱ & ۰ \end{pmatrix}, \quad \boldsymbol{T}_t = \begin{pmatrix} ۱ & ۱ \\ ۰ & ۱ \end{pmatrix}, \boldsymbol{Q}_t = \begin{pmatrix} \sigma_\eta^۲ & ۰ \\ ۰ & \sigma_\zeta^۲ \end{pmatrix}.$$

پالایش کالمن: از معروف‌ترین الگوریتم‌ها برای تجزیه و تحلیل معادلات فضای حالت، الگوریتم پالایش کالمن است. در مدل فضای حالت گاوسی، پالایش کالمن مجموعه‌ای از روش‌های بازگشتی برای بدست آوردن برآورد بردار α_t، پالایش کردن و پیش‌بینی داده‌های سری زمانی براساس مشاهدات $Y_t = \{y_۱, \ldots, y_t\}$ است. به منظور اجرا کردن الگوریتم پالایش کالمن فرض کنید $Y_t = \{y_۱, \ldots, y_t\}$ و $a_{t-۱} = a_{t-۱|t-۱} = E(\alpha_{t-۱}|Y_{t-۱})$ بهترین برآوردکننده بر اساس مشاهدات $Y_{t-۱}$ برای $\alpha_{t-۱}$ باشد و $\boldsymbol{p}_{t-۱} = \boldsymbol{p}_{t-۱|t-۱} = Var\left[\alpha_{t-۱}|Y_{t-۱}\right]$

نیز نشان‌دهنده ماتریس کوواریانس با بعد $m \times m$ از خطای این برآوردکننده باشد. حال اگر a_{t-1} و $\boldsymbol{p}_{t-1}$ در اختیار باشند، آنگاه برآورد α_t و ماتریس کوواریانس برآورد $\boldsymbol{p}_t$ با توجه به اطلاعات موجود تا زمان $t-1$ به‌صورت

$$a_{t|t-1} = E(\alpha_t|Y_{t-1}) = \boldsymbol{T}_t E(\alpha_{t-1}|Y_{t-1}),$$
$$\boldsymbol{p}_{t|t-1} = Var(\alpha_t|Y_{t-1}) = \boldsymbol{T}_t Var(\alpha_{t-1}|Y_{t-1})\boldsymbol{T}_t' + \boldsymbol{Q}_t.$$

بوده و معادلات پیش‌بینی نامیده می‌شوند. مقدار پیش‌بینی y_t در لحظه $t-1$ یا مقدار y_t بدون نویز به‌صورت

$$\hat{y}_{t|t-1} = E(y_t|Y_{t-1}) = Z_t a_{t|t-1}, \qquad (۳)$$

به‌دست می‌آید. آنگاه خطای پیش‌بینی یا نوساز[1] در لحظه $t-1$ به صورت

$$\upsilon_t = y_t - \hat{y}_{t|t-1} = Z_t(\alpha_t - a_{t|t-1}) + \epsilon_t, \quad t = 1, \ldots, n$$

است با ماتریس کوواریانس $F_t = Var(\upsilon_t|Y_{t-1}) = Z_t \boldsymbol{p}_{t|t-1} Z_t' + \sigma_t$[2] است. وقتی y_t در لحظه t مشاهده شود، برآوردگر α_t و ماتریس کوواریانس آن در قالب معادلات به هنگام به‌صورت

$$\begin{aligned} a_t &= a_{t|t} = E(\alpha_t|Y_t) = a_{t|t-1} + \boldsymbol{p}_{t|t-1} Z_t' F_t^{-1} \upsilon_t, \\ \boldsymbol{p}_t &= \boldsymbol{p}_{t|t} = \boldsymbol{p}_{t|t-1} - \boldsymbol{p}_{t|t-1} Z_t' F_t^{-1} Z_t \boldsymbol{p}_{t|t-1}', \end{aligned}$$

ارائه می‌شود. معادلات پیش‌بینی و معادلات به‌هنگام تواماً الگوریتم پالایش کالمن را می‌سازند (شاموی و استوفر ، ۲۰۱۱؛ ذبیحی مقدم و همکاران ، ۱۳۹۷).

۳ مروری بر روش SSA

روش SSA یک روش ناپارامتری برای تجزیه و تحلیل سری‌های زمانی بوده که می‌تواند مشاهدات سری زمانی را به چندین مؤلفه تقسیم کرده و در ادامه برای انجام پیش‌بینی مورد استفاده قرار گیرد. این روش از دو مرحله تجزیه و بازسازی تشکیل شده است که هر یک از این مراحل شامل دو گام هستند.

مرحله ۱. تجزیه: مرحله تجزیه شامل دو گام نشانیدن[2] و تجزیه مقدار تکین (SVD) است.

[1] Innovation
[2] Embedding

گام ۱. نشاندن: در گام اول از مرحله تجزیه یعنی نشاندن، ابتدا مشاهدات سری زمانی $Y_N = \{y_1, \ldots, y_N\}$ به صورت ماتریس $\mathbf{X} = \{X_1, \ldots, X_K\}$ که $X_i = (y_i, \ldots, y_{i+L-1})'$ نوشته می‌شود. لازم به ذکر است به ماتریس $\mathbf{X}$ که در آن L طول پنجره به صورت $2 \leq L \leq N/2$ و $K = N - L + 1$ بوده و همچنین عناصر در قطرهای فرعی آن با هم برابرند، ماتریس مسیر و یا ماتریس هنکل[1] گویند.

گام ۲. تجزیه مقدار تکین: در این گام ماتریس مسیر به مجموع ماتریس‌هایی با رتبه یک تجزیه می‌شود. فرض کنید $\lambda_1, \ldots, \lambda_L$ مقادیر ویژه ماتریس $\mathbf{X}\mathbf{X}'$ باشند که به صورت نزولی $(\lambda_1 \geq \ldots \geq \lambda_L \geq 0)$ مرتب شده‌اند و $U_1, \ldots, U_L$ نیز بردارهای یکه متعامد ویژه متناظر با مقادیر ویژه $\lambda_1, \ldots, \lambda_L$ و $d = \max\{i : \lambda_i > 0\}$ باشند. در این صورت SVD ماتریس X به صورت $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$ نوشته می‌شود، که $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i'$ و $V_i = \mathbf{X}' U_i / \sqrt{\lambda_i}$ برای $i = 1, \ldots, d$ هستند. همچنین سه تایی $(\sqrt{\lambda_i}, U_i, V_i)$ را سه تایی ویژه[2] می‌نامند.

مرحله ۲. :بازسازی این مرحله از دو گام گروه بندی و میانگین گیری قطری تشکیل شده است.

گام ۱. گروه بندی: در این گام، هدف تشخیص مؤلفه‌های سیگنال و نویز است. پس از محاسبه SVD، ماتریس‌های ابتدایی $\mathbf{X}_i$ به چندین گروه تقسیم شده و در هر گروه با هم جمع می‌شوند. بعنوان نمونه فرض کنید، اندیس‌های $i_1 \ldots i_r$ متناظر با گروه سیگنال و $I_r = \{i_1 \ldots i_r\}$ باشد، در نتیجه ماتریس $\mathbf{X}_{I_r}$ متناظر با گروه I_r یا گروه سیگنال‌ها به صورت $\mathbf{X}_{I_r} = \mathbf{X}_{i_1} + \ldots + \mathbf{X}_{i_r}$ تعریف می‌شود.

گام ۲. میانگین گیری قطری: در این گام، ماتریس‌های بدست آمده از مرحله گروه‌بندی به ماتریس هنکل تبدیل می‌شوند و سپس به صورت یک سری زمانی نوشته می‌شوند.

دو پارامتر مهم در روش SSA وجود دارند. اولین پارامتر طول پنجره L است که در گام نشاندن برای تشکیل ماتریس مسیر مورد نیاز است. انتخاب نادرست این پارامتر به تجزیه و گروه‌بندی نامناسب سری زمانی منجر خواهد شد. متاسفانه روشی یکتا برای تعیین L وجود ندارد ولی به هر حال، مجموعه‌ای از اصول و قواعد کلی که از پشتوانه نظری و کاربردی خوبی برخوردار بوده وجود دارند که می‌توانند در انتخاب مناسب پارامتر L راهگشا باشند. دومین پارامتری که نقش اساسی و کلیدی در بازسازی سیگنال یک سری زمانی ایفا می‌کند، تعداد سه تایی‌های ویژه‌ای است که تبیین کننده سیگنال بوده و در بازسازی آن به کار می‌روند. معمولاً این پارامتر را با r نشان داده و پارامتر بازسازی می‌نامند. برای تعیین r می‌توان از اطلاعات نهفته در مقادیر ویژه و بردارهای ویژه متناظر استفاده کرد. روش SSA پس از انجام مراحل تجزیه و بازسازی، قابلیت محاسبه پیش‌بینی را با استفاده از چندین روش پیش‌بینی دارد. در ادامه پیش‌بینی با SSA-R معرفی می‌شود. برای اطلاع در مورد انتخاب پارامترهای L و r و آشنایی با سایر روش‌های پیش‌بینی SSA به گولیاندینا و ژیگلجاوسکی (۲۰۱۳) و صانعی و حسنی (۲۰۱۶) مراجعه شود.

---

[1] Hankel matrix
[2] Eigentriples

۳.۱ پیش‌بینی بازگشتی SSA

به منظور انجام پیش‌بینی با روش SSA برای مشاهدات سری زمانی $Y_N = \{y_1, \ldots, y_N\}$، فرض کنید $I_r = \{i_1, \ldots, i_r\}$ مجموعه سه تایی ویژه متناظر با سیگنال‌ها، r تعداد سه تایی ویژه اول، $U_i \in \mathbb{R}^L$ بردارهای ویژه متناظر با مقادیر ویژه از سه تایی ویژه اول، $\underline{U}_i \in \mathbb{R}^{L-1}$، تعداد $L-1$ مؤلفه از بردارهای ویژه U_i، π_i آخرین مؤلفه از بردارهای ویژه U_i و $\tilde{Y}_N = \{\tilde{y}_1, \ldots, \tilde{y}_N\}$ سری بازسازی شده براساس I_r باشند. در این صورت پیش‌بینی مشاهدات جدید سری زمانی با استفاده از روش SSA به صورت

$$z_i = \begin{cases} \tilde{y}_i, & i = 1, \ldots, N \\ \sum_{j=1}^{L-1} \phi_j z_{i-j}, & i = N+1, \ldots, N+h. \end{cases} \quad (۴)$$

به دست می‌آید، که در آن $z_{N+1}, \ldots, z_{N+h}$ پیش‌بینی‌های بدست آمده تا h گام به جلو و $\phi_1, \ldots, \phi_{L-1}$ ضرایب پیش‌بینی روش بازگشتی SSA هستند، که به صورت

$$R = (\phi_{L-1}, \ldots, \phi_1)' = \frac{1}{1-\nu^2} \sum_{i \in I_r} \pi_i \underline{U}_i \quad (۵)$$

به دست می‌آیند و $\nu^2 = \sum_{i \in I_r} \pi_i^2$ است.

۳.۲ پیش‌بینی بازگشتی SSA-R با استفاده از پالایش داده‌ها و الگوریتم وزنی

در این بخش روش‌های مختلف پیش‌بینی بازگشتی SSA با استفاده از الگوریتم پالایش کالمن، SSA بازسازی شده (حسنی و همکاران ، ۲۰۱۷) و SSA وزنی (کلانتری و همکاران ، ۲۰۲۰) بررسی می‌شود.

الف- روش ترکیبی پیش‌بینی بازگشتی SSA-R با استفاده از الگوریتم پالایش کالمن: فرض کنید سری زمانی شامل دو مؤلفه سیگنال و نویز به صورت $y_t = s_t + \epsilon_t$ باشد، که در آن $\{y_t : t = 1, \ldots, N\}$ مشاهدات، s_t مؤلفه سیگنال و ϵ_t مؤلفه نویز هستند. همان طور که اشاره شد، برای پیش‌بینی مشاهدات جدید هر سری زمانی با طول پنجره ثابت L، تعداد $L-1$ ضریب $\phi_1, \ldots, \phi_{L-1}$ وجود دارد، که با استفاده از بردارهای ویژه ماتریس مسیر X بدست می‌آیند و نقش مهمی در پیش‌بینی دارند. در روش SSA، اگر سری زمانی شامل نویز نباشد، یعنی $y_t = s_t$، مقدار دقیق این ضرایب را می‌توان بدست آورد، اما در داده‌های واقعی با توجه اینکه مشاهدات معمولاً دارای نویز هستند، ماتریس مسیر نیز شامل نویز بوده و بردارهای ویژه بدست آمده از آن نیز نویزی هستند. بنابراین نتیجه می‌گیریم که ضرایب پیش‌بینی نیز شامل نویز بوده و از دقت کافی برخوردار نیستند. لذا هنگامی که داده‌های سری زمانی شامل نویز باشند این ضرایب موجب کاهش دقت پیش‌بینی می‌شوند. بنابراین برای دستیابی به داده‌های با نویز کمتر و بهبود پیش‌بینی روش SSA-R می‌توان از معادلات فضای حالت و الگوریتم های پالایش

کالمن بهره جست. با استفاده از این ایده، روش ترکیبی برای پیش‌بینی SSA براساس پالایش کالمن، که با نماد $(KF-SSA-R)$ نشان داده می‌شود، به دست می‌آید.

فرض کنید $\{y_t : t = ۱ \dots N\}$ مشاهدات اولیه سری زمانی باشند که معادلات فضای حالت برای آن‌ها به صورت معادلات (۱) و (۲) تعریف شده و $\{\hat{y}_t : t = ۱ \dots N\}$ یک سری زمانی با نویز کم باشد که به وسیله الگوریتم پالایش کالمن معادلات فضای حالت در معادله (۳) تولید شده است. با جایگزینی سری زمانی کم نویز $\{\hat{y}_t : t = ۱ \dots N\}$ به جای مشاهدات $\{y_t : t = ۱ \dots N\}$ در روش SSA پایه، فرض کنید $\tilde{\mathbf{X}}$ ماتریس مسیر بدست آمده برای سری زمانی $\{\hat{y}_t : t = ۱ \dots N\}$ از گام نشانیدن مرحله تجزیه روش SSA پایه، $\tilde{\lambda}_۱ \dots \tilde{\lambda}_L$ و $\tilde{U}_۱ \dots \tilde{U}_L$ به ترتیب مقادیر ویژه و بردارهای ویژه ماتریس $\tilde{\mathbf{X}}\tilde{\mathbf{X}}'$ باشند. همچنین فرض کنید I_r مجموعه سه تایی ویژه انتخاب شده باشد، در این صورت ضرایب پیش‌بینی روش KF-SSA-R با استفاده از معادله (۵) به صورت

$$R = (\tilde{\phi}_{L-۱}, \dots, \tilde{\phi}_۱)' = \frac{۱}{۱ - \tilde{\nu}^۲} \sum_{i \in I_r} \tilde{\pi}_i \underline{\tilde{\mathbf{U}}}_i,$$

به‌دست می‌آیند، که بردار $\underline{\tilde{\mathbf{U}}}_i$ شامل $L-۱$ مؤلفه از بردار $\tilde{U}_i$، $\tilde{\pi}_i$ آخرین مؤلفه از بردار $\tilde{U}_i$ و $\tilde{\nu}^۲ = \sum_{i \in I_r} \tilde{\pi}_i^۲$ هستند. در نتیجه پیش‌بینی h گام به جلو برای روش KF-SSA-R با استفاده از معادله (۴) به صورت

$$z_i = \begin{cases} \tilde{\tilde{y}}_i, & i = ۱, \dots, N. \\ \sum_{j=۱}^{L-۱} \tilde{\phi}_j z_{i-j}, & i = N+۱, \dots, N+h. \end{cases}$$

تعریف می‌شود، که در آن $\tilde{\tilde{y}}_i$ برای $i = ۱, \dots, N$ سری بازسازی شده بوسیله مجموعه سه تایی ویژه I_r است. لازم به ذکر است که نویز زیاد در داده‌های سری زمانی موجب نادقیق بودن در انتخاب سه تایی ویژه شده و استفاده از روش KF-SSA-R می‌تواند به دقیق‌تر شدن سه تایی ویژه کمک نماید. همچنین روش KF-SSA-R باعث تجزیه بهتر مشاهدات سری به مؤلفه‌های غیر قابل مشاهده شده و در نتیجه می‌تواند موجب بهبود پیش‌بینی شود.

ب- روش SSA بازسازی شده: در روش SSA بازسازی شده از آنجایی که سری بازسازی شده دارای نویز کمتری نسبت به سری اصلی است، ضرایب پیش‌بینی از سری بازسازی شده بدست می‌آیند. فرض کنید $\{\hat{y}_t : t = ۱ \dots N\}$ سری زمانی بازسازی شده بوسیله I_r از روش SSA باشند. $\tilde{\mathbf{X}}$ ماتریس مسیر بدست آمده از این سری بازسازی شده و $\tilde{\lambda}_۱ \dots \tilde{\lambda}_L$ و $\tilde{U}_۱ \dots \tilde{U}_L$ به ترتیب مقادیر ویژه و بردارهای ویژه ماتریس $\tilde{\mathbf{X}}\tilde{\mathbf{X}}'$ باشند. همچنین فرض کنید I_r مجموعه سه تایی ویژه انتخاب شده باشد، در این صورت ضرایب پیش‌بینی روش SSA-R بازسازی شده به صورت

$$R = (\tilde{\phi}_{L-۱}, \dots, \tilde{\phi}_۱)' = \frac{۱}{۱ - \tilde{\nu}^۲} \sum_{i \in I_r} \tilde{\pi}_i \underline{\tilde{\mathbf{U}}}_i,$$

به‌دست می‌آیند، که در آن بردار $\underline{\tilde{U}}_i$ شامل $L-1$ مؤلفه از بردار $\tilde{U}_i$، $\tilde{\pi}_i$ آخرین مؤلفه از بردار $\tilde{U}_i$ و $\tilde{\nu}^2 = \sum_{i\in I_r} \tilde{\pi}_i^2$ هستند. در نتیجه پیش‌بینی h گام به جلو به صورت

$$z_i = \begin{cases} \tilde{\tilde{y}}_i, & i = 1, \ldots, N \\ \sum_{j=1}^{L-1} \tilde{\phi}_j z_{i-j}, & i = N+1, \ldots, N+h. \end{cases}$$

تعریف می‌شود، که در آن $\tilde{\tilde{y}}_i$ سری بازسازی شده بوسیله مجموعه سه‌تایی ویژه I_r و $z_{N+1}, \ldots, z_{N+h}$ پیش‌بینی‌های به‌دست آمده تا h گام به جلو از روش بازگشتی SSA بازسازی شده، هستند.

ج- روش SSA وزنی: یکی از روش‌های بهبود پیش‌بینی، وزن دار کردن مناسب ضرایب پیش‌بینی روش SSA است (کلانتری و همکاران ، ۲۰۲۰). فرض کنید $I_r = \{i_1, \ldots, i_r\}$ مجموعه سه تایی ویژه متناظر با سیگنال‌ها، r تعداد سه تایی ویژه اول، $U_i \in \mathbb{R}^L$ بردارهای ویژه متناظر با مقادیر ویژه از سه تایی ویژه اول، $\underline{U}_i \in \mathbb{R}^{L-1}$، تعداد $L-1$ مؤلفه از بردارهای ویژه U_i، π_i آخرین مؤلفه از بردارهای ویژه U_i و $\tilde{Y}_N = \{\tilde{y}_1, \ldots, \tilde{y}_N\}$ سری بازسازی شده براساس I_r باشند. در این صورت ضرایب پیش‌بینی SSA وزنی به صورت

$$R_w^{(h)} = w_h(\phi_{L-1}, \ldots, \phi_1)' = \frac{w_h}{1-\nu^2} \sum_{i\in I_r} \pi_i \underline{U}_i,$$

به دست می‌آیند، که در آن w_h وزن پیش‌بینی در h گام به جلو بوده و به صورت $\frac{y_{N+h}}{z_{N+h}}$ تعریف می‌شود. بنابراین پیش‌بینی h گام به جلو برای روش SSA-R وزنی را می‌توان به صورت

$$z_i = \begin{cases} \tilde{y}_i, & i = 1, \ldots, N \\ \sum_{j=1}^{L-1} w_{i-N} \phi_j z_{i-j}, & i = N+1, \ldots, N+h. \end{cases}$$

نوشت، که در آن $z_{N+1}, \ldots, z_{N+h}$ پیش‌بینی‌های به‌دست آمده تا h گام به جلو در پیش‌بینی روش بازگشتی SSA وزنی است. لازم به ذکر است، این روش هنگامی قابل اجراست که مقادیر y_{N+h} و z_{N+h} قابل دسترس باشند. اما با توجه اینکه مشاهدات سری زمانی $Y_N = \{y_1, \ldots, y_N\}$ فرض شده است، دسترسی به y_{N+h} وجود ندارد. برای رفع این مشکل و استفاده از این روش در افق پیش‌بینی h، وزن w_h برابر با میانه رابطه $w_h^{(i)} = \frac{y_{M+h+i-1}}{z_{M+h+i-1}}$ به ازای $i = 1, \ldots, N-M-h+1$ و $M = \frac{2}{3}N$ درنظر گرفته شده است.

۴ کاربرد

در این بخش، کارایی روش‌های KF-SSA-R، SSA-R بازسازی شده، SSA-R وزنی و SSA-R پایه با استفاده از سری‌های زمانی شبیه‌سازی شده و داده‌های واقعی مورد مقایسه قرار می‌گیرند. برای این منظور سری زمانی $Y_N = \{y_1, \ldots, y_N\}$ به دو مجموعه داده تقسیم می‌شود. از یک مجموعه داده برای بازسازی و از مجموعه داده دیگر برای به دست آوردن پیش‌بینی استفاده می‌شود. برای انجام این مقایسه‌ها از معیار RMSE استفاده شده است. بدیهی است روش مطلوب‌تر، روشی است که از کمترین مقدار RMSE برخوردار باشد. برای مقایسه آسان‌تر کارایی روش KF-SSA-R با سایر روش‌های مطرح شده از نسبت RMSE به صورت

$$RRMSE_h = \frac{RMSE_h(\text{KF-SSA-R})}{RMSE_h(\text{سایر روش‌ها})}$$

استفاده می‌شود. همچنین برای مقایسه روش‌های مطرح شده با روش SSA-R پایه از نسبت RMSE به صورت

$$RRMSE_h = \frac{RMSE_h(\text{روش‌های مطرح شده})}{RMSE_h(\text{SSA-R})}$$

استفاده شده است. بعنوان نمونه برای مقایسه روش KF-SSA-R با روش SSA-R پایه نسبت RMSE به صورت

$$\begin{aligned} RRMSE_h(KF-SSA-R, SSA-R) &= \frac{RMSE_h(\text{KF-SSA-R})}{RMSE_h(\text{SSA-R original})} \\ &= \sqrt{\frac{\sum_{t=M}^{N-h}(y_{t+h} - \hat{y}_{t+h|t})^2}{\sum_{t=M}^{N-h}(y_{t+h} - \hat{\hat{y}}_{t+h|t})^2}} \end{aligned}$$

تعریف می‌شود، که N، M و h به ترتیب طول سری زمانی Y_N، طول سری بازسازی شده و افق پیش‌بینی بوده و $\hat{y}_{t+h|t}$ و $\hat{\hat{y}}_{t+h|t}$ به ترتیب پیش‌بینی‌های h گام به جلو بدست آمده از SSA-R و KF-SSA-R هستند. اگر $RRMSE_h < 1$ باشد، نتیجه می‌گیریم که روش KF-SSA-R از کارایی بیشتری نسبت به روش SSA-R برخوردار است. لازم به ذکر است که برای انجام محاسبات مربوط به روش SSA و پالایش کالمن، از بسته‌های $Rssa$ و $Stats$ در نرم افزار R استفاده شده است.

۴.۱ مطالعات شبیه‌سازی

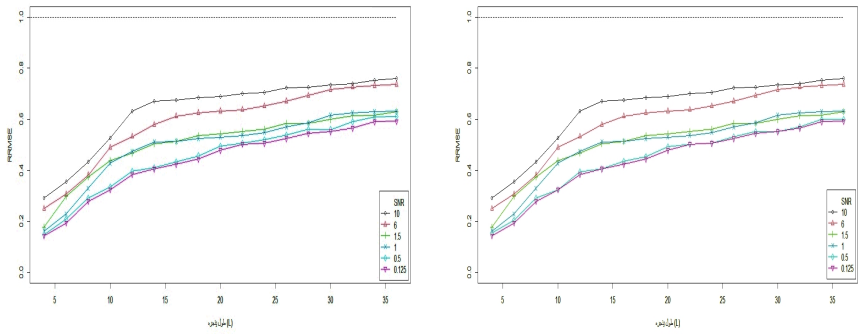
در مطالعات شبیه‌سازی، ۱۵۰ داده با استفاده از مدل‌های مختلف ساختاری تولید شده است. از $M = 100$ مشاهده اول برای بازسازی و از ۵۰ مشاهده بعدی برای پیش‌بینی استفاده می‌شود. در کلیه روش‌های ارائه شده، تعداد سه تایی ویژه (r)، برای بازسازی و پیش‌بینی بر اساس رتبه ماتریس مسیر به دست آمده و شبیه‌سازی برای هر یک از

مدل‌ها در ۱۰۰۰ تکرار انجام شده و میانگین $RRMSE$ محاسبه می‌گردد. همچنین به منظور ارزیابی تأثیر سطح نویز بر نتایج پیش‌بینی، از انواع مختلف نسبت واریانس سیگنال به نویز (SNR) برای افق‌های مختلف پیش‌بینی استفاده شده و بدین منظور مقادیر $SNR = ۰/۱۲۵, ۰/۵, ۱, ۱/۵, ۶, ۱۰$ درنظر گرفته شده است.

مثال ۲. مدل روند خطی موضعی

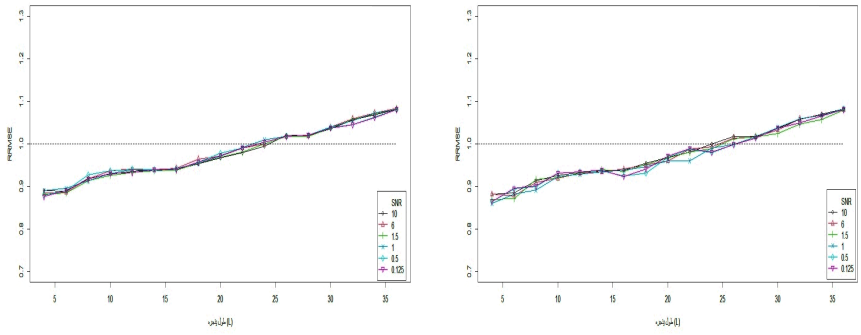
$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim N(۰, \sigma_\epsilon^۲)\\
\mu_t &= \mu_{t-۱} + \beta_{t-۱} + \eta_t, & \eta_t &\sim N(۰, \sigma_\eta^۲)\\
\beta_t &= \beta_{t-۱} + \zeta_t, & \zeta_t &\sim N(۰, \sigma_\zeta^۲)
\end{aligned}
$$

را در نظر بگیرید، که در آن $\mu_۰ = ۲/۵$، $\beta_۰ = ۰/۰۰۵$، ϵ_t مؤلفه نویز، η_t و ζ_t مؤلفه‌های اغتشاشی مستقل از یکدیگر هستند. در کلیه روش‌های مطرح شده با توجه به رتبه ماتریس مسیر، دو مقدار ویژه اول برای بازسازی و پیش‌بینی انتخاب شده است، یعنی $(r = ۲)$. شکل ۱، برای افق‌های پیش‌بینی $h = ۲, ۱۲$، نمودارهای $RRMSE$ روش KF-SSA-R نسبت به روش SSA-R را برای انواع مختلف SNR و طول پنجره‌های مختلف نشان می‌دهد. بر اساس این نمودارها می‌توان نتیجه گرفت، که روش KF-SSA-R از کارایی بیشتری نسبت به روش SSA-R هنگامی که مقادیر SNR یا مقادیر طول پنجره (L) کوچک باشد، برخوردار است. شکل ۲، نمودارهای $RRMSE$ روش SSA-R بازسازی شده نسبت به روش SSA-R را برای انواع مختلف SNR در طول پنجره‌های مختلف برای افق‌های پیش‌بینی $h = ۲, ۱۲$ نشان می‌دهد. با بررسی این نمودارها می‌توان نتیجه گرفت که روش SSA-R بازسازی شده نسبت به روش SSA-R هنگامی که طول پنجره کوچک باشد کارایی بیشتری دارد. شکل ۳، نمودارهای $RRMSE$ روش SSA-R وزنی نسبت به روش SSA-R را برای انواع مختلف SNR در طول پنجره‌های مختلف برای افق‌های پیش‌بینی $h = ۲, ۱۲$ نشان می‌دهد و بیانگر آن است که برای طول پنجره بزرگ، روش SSA-R وزنی نسبت به SSA-R پایه کارایی بیشتری دارد. شکل‌های ۴ و ۵، نمودارهای $RRMSE$ مربوط به روش KF-SSA-R نسبت به SSA-R بازسازی شده و SSA-R وزنی را برای انواع مختلف SNR در طول پنجره‌های مختلف، برای افق‌های پیش‌بینی $h = ۲, ۱۲$ نشان می‌دهند. بررسی این نمودارها نشان می‌دهد که برای مقادیر طول پنجره کوچک، پیش‌بینی روش KF-SSA-R از کارایی بیشتری نسبت به روش‌های SSA-R بازسازی شده و SSA-R وزنی برخوردار است.



(ب) (الف)

شکل ۱. $RRMSE$ روش KF-SSA-R نسبت به SSA-R، SNR الف- برای انواع مختلف افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$

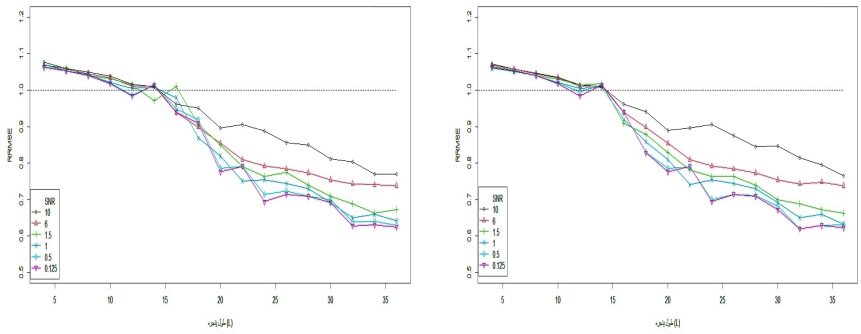


(ب) (الف)

شکل ۲. $RRMSE$ روش SSA-R بازسازی شده نسبت به SSA-R پایه، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$

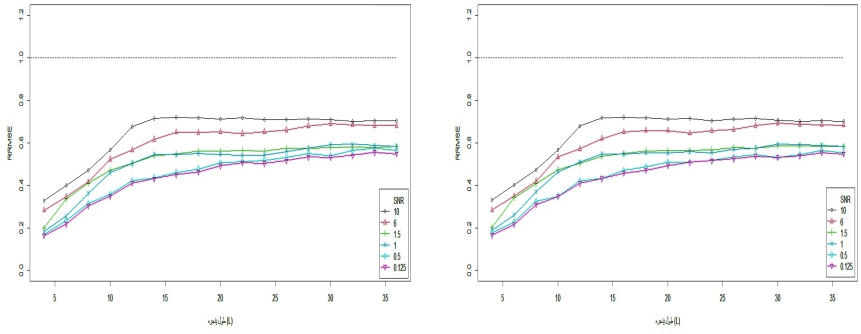
مثال ۳. مدل فصلی ساختگی سه ماهه

$$
\begin{aligned}
y_t &= \mu_t + \gamma_t + \epsilon_t, && \epsilon_t \sim N(۰, \sigma_\epsilon^۲)\\
\mu_t &= \mu_{t-۱} + \beta_{t-۱} + \eta_t, && \eta_t \sim N(۰, \sigma_\eta^۲)\\
\beta_t &= \beta_{t-۱} + \zeta_t, && \zeta_t \sim N(۰, \sigma_\zeta^۲)\\
\sum_{j=۰}^{۳} \gamma_{t-j} &= \omega_t, && \omega_t \sim N(۰, \sigma_\omega^۲)
\end{aligned}
$$



(ب) (الف)

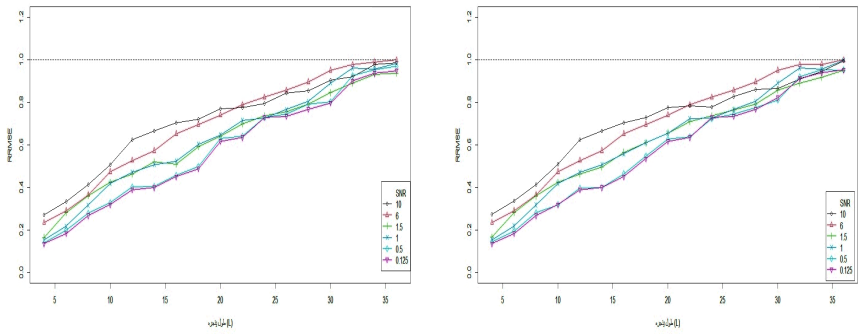
شکل ۳. $RRMSE$ روش SSA-R وزنی نسبت به SSA-R پایه، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$



(ب) (الف)

شکل ۴. $RRMSE$ روش KF-SSA-R نسبت به SSA-R بازسازی شده، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$

را در نظر بگیرید. در معادلات فوق γ_t مؤلفه اثرات فصلی ، $\mu_\circ = ۲/۵$، $\beta_\circ = ۰/۰۰۵$ ، η_t، ζ_t و ω_t عامل‌های اغتشاشی مستقل از یکدیگر و نیز مستقل از اغتشاش ϵ_t هستند. هم چنین مؤلفه‌های روند، دوره و اثرات فصلی مستقل از یکدیگرند. درکلیه روش‌های مطرح شده با توجه به رتبه ماتریس مسیر، پنج مقدار ویژه اول برای بازسازی و پیش‌بینی انتخاب شده است، به عبارت دیگر $(r = ۵)$. شکل‌های ۶، ۷ و ۸، نمودارهای $RRMSE$ روش‌های مطرح شده نسبت به روش SSA-R را برای انواع مختلف SNR در طول پنجره‌های مختلف و افق‌های پیش‌بینی



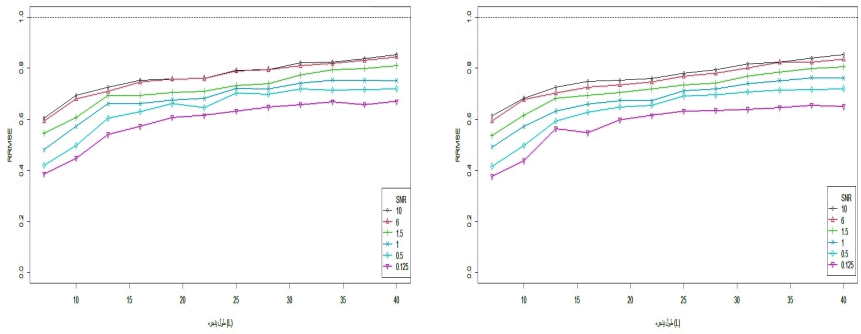
(ب) (الف)

شکل ۵. $RRMSE$ روش KF-SSA-R نسبت به SSA-R وزنی، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$

$h = ۲, ۱۲$ نشان می‌دهد. بر اساس نمودارهای $RRMSE$ رسم شده در شکل ۶، می‌توان نتیجه گرفت که روش KF-SSA-R از کارایی بیشتری نسبت به روش SSA-R هنگامی که مقادیر SNR یا مقادیر طول پنجره (L) کوچک باشد، برخوردار است. از شکل ۷ می‌توان نیز نتیجه گرفت که روش SSA-R بازسازی شده نسبت به SSA-R پایه هنگامی که طول پنجره کوچک باشد کارایی بیشتری دارد. همچنین از شکل ۸ می‌توان نتیجه گرفت، هنگامی که طول پنجره بزرگ باشد، روش SSA-R وزنی نسبت به SSA-R پایه کارایی بیشتری دارد. شکل‌های ۹ و ۱۰، نمودارهای $RRMSE$ روش KF-SSA-R نسبت به روش SSA-R بازسازی شده و روش SSA-R وزنی را برای انواع مختلف SNR در طول پنجره‌های مختلف، برای افق‌های پیش‌بینی $h = ۲, ۱۲$ نشان می‌دهد. از بررسی این نمودارها، می‌توان نتیجه گرفت برای مقادیر طول پنجره کوچک روش KF-SSA-R از کارایی بیشتری نسبت به روش‌های SSA-R بازسازی شده و SSA-R وزنی برخوردار است.

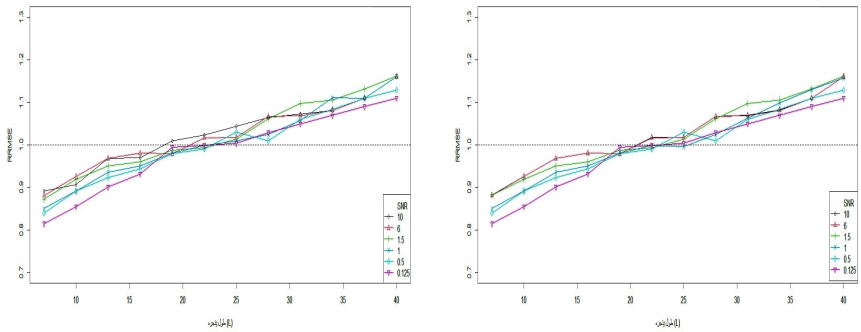
۴.۲ تحلیل داده‌های میزان مصرف گاز در انگلستان

در این بخش برای بررسی کارایی روش پیشنهادی از داده‌های میزان مصرف گاز در کشور انگلستان که در نرم افزار R با نام $UKgas$ قابل دسترس هستند، استفاده شده است. شکل ۱۱ نمودار سری زمانی این داده‌ها را برای سه ماهه اول سال ۱۹۶۰ تا سه ماهه چهارم سال ۱۹۸۶ نشان می‌دهد. به منظور تجزیه و تحلیل این داده‌ها از ۹۲ مشاهده اول برای بازسازی و از ۱۶ مشاهده باقیمانده برای پیش‌بینی استفاده شده است. شکل ۱۲ رفتار مقادیر ویژه ماتریس مسیر این داده‌ها را نشان می‌دهد. با توجه به این نمودار و دلایل ارائه شده توسط حسنی و محمودوند (۲۰۱۸) در کلیه روش‌های مطرح شده، از هفت مقدار ویژه اول برای بازسازی و پیش‌بینی استفاده می‌شود. جداول ۱، ۲ و ۳ به ترتیب مقادیر $RRMSE$ روش‌های $KF-SSA-R$، $SSA-R$ بازسازی شده و SSA-R



(ب) (الف)

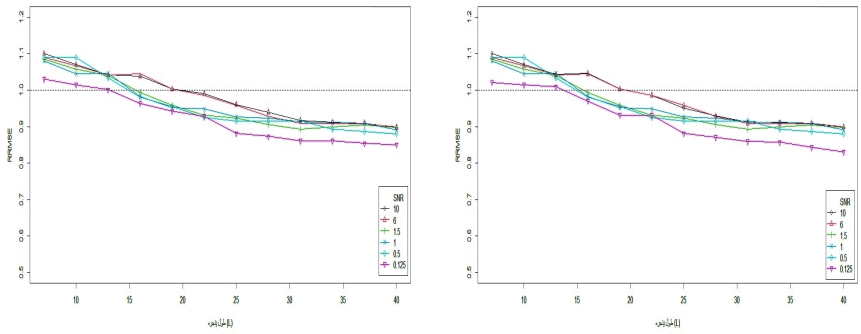
شکل ۶. $RRMSE$ روش KF-SSA-R نسبت به SSA-R، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$



(ب) (الف)

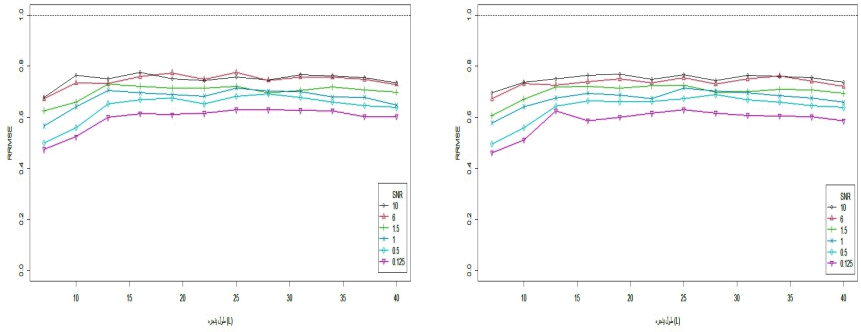
شکل ۷. $RRMSE$ روش SSA-R بازسازی شده نسبت به SSA-R پایه، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$

وزنی نسبت به SSA-R را برای افق‌های مختلف پیش‌بینی h ارائه می‌کنند. در هر یک از افق‌های پیش‌بینی، از مقادیر مختلف طول پنجره (L) استفاده شده است. بر اساس نتایج بدست آمده از جدول ۱، می‌توان نتیجه گرفت که هنگامی که مقادیر طول پنجره کوچک است روش KF-SSA-R از کارایی بیشتری نسبت به روش SSA-R دارد. نتایج ارائه شده در جدول ۲ حاکی از آن است که روش SSA-R بازسازی شده نسبت به SSA-R هنگامی که طول پنجره کوچک باشد کاراتر است. همچنین از جدول ۳ می‌توان نتیجه گرفت، هنگامی که طول پنجره بزرگ باشد، روش SSA-R وزنی نسبت به روش SSA-R از کارایی بیشتری برخوردار است. جداول ۴ و ۵ به ترتیب



(ب) (الف)

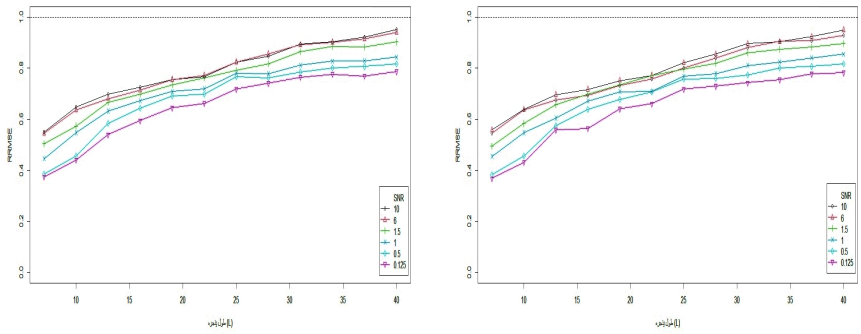
شکل ۸. $RRMSE$ روش SSA-R وزنی نسبت به SSA-R پایه، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$



(ب) (الف)

شکل ۹. $RRMSE$ روش KF-SSA-R نسبت به SSA-R بازسازی شده، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$

مقادیر $RRMSE$ روش KF-SSA-R نسبت به روش‌های SSA-R بازسازی شده و SSA-R وزنی برای مقادیر مختلف افق‌های پیش‌بینی ارائه می‌کنند. در هر یک از افق‌های پیش‌بینی، از مقادیر مختلف طول پنجره استفاده شده است. بر اساس نتایج $RRMSE$ به دست آمده از جداول ۴ و ۵، می‌توان نتیجه گرفت که برای مقادیر کوچک طول پنجره روش KF-SSA-R دارای کارایی بیشتری نسبت به روش SSA-R بازسازی شده و روش SSA-R وزنی داشته و به نوعی با نتایج قبلی به دست آمده برای سری‌های شبیه‌سازی شده مطابقت دارد.



(الف) (ب)

شکل ۱۰. $RRMSE$ روش KF-SSA-R نسبت به SSA-R وزنی، الف- برای افق پیش‌بینی $h = ۲$، ب- برای افق پیش‌بینی $h = ۱۲$

جدول ۱. مقدار $RRMSE$ روش KF-SSA-R نسبت به SSA-R برای مقادیر مختلف طول پنجره (L) در افق‌های پیش‌بینی مختلف

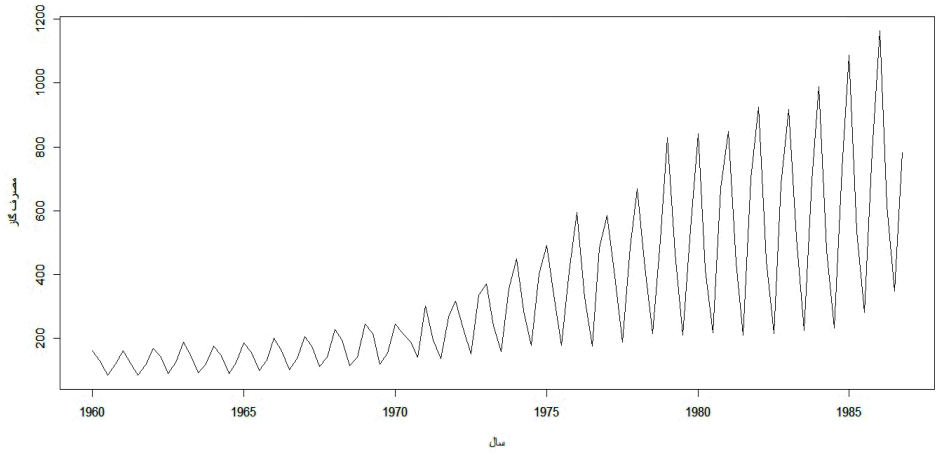
h	$RRMSE(L)$					
۱	۰/۹۶(۴۶)	۰/۹۵(۳۶)	۰/۹۳(۳۱)	۰/۹۱(۲۳)	۰/۸۹(۱۷)	۰/۸۹(۱۲)
۲	۰/۹۶(۴۴)	۰/۹۵(۳۷)	۰/۹۳(۳۰)	۰/۹۱(۲۴)	۰/۹۰(۲۰)	۰/۹۰(۱۱)
۳	۰/۹۶(۴۴)	۰/۹۶(۳۸)	۰/۹۴(۳۲)	۰/۹۰(۲۵)	۰/۹۰(۲۱)	۰/۹۰(۱۵)
۴	۰/۹۵(۴۳)	۰/۹۳(۳۵)	۰/۹۰(۲۴)	۰/۸۹(۲۰)	۰/۸۹(۱۵)	۰/۸۹(۱۲)
۶	۰/۹۳(۴۲)	۰/۹۱(۳۳)	۰/۹۰(۲۹)	۰/۸۷(۲۱)	۰/۸۷(۱۴)	۰/۸۶(۱۱)

جدول ۲. مقدار $RRMSE$ روش SSA-R بازسازی شده نسبت به SSA-R برای مقادیر مختلف طول پنجره (L) در افق‌های پیش‌بینی مختلف

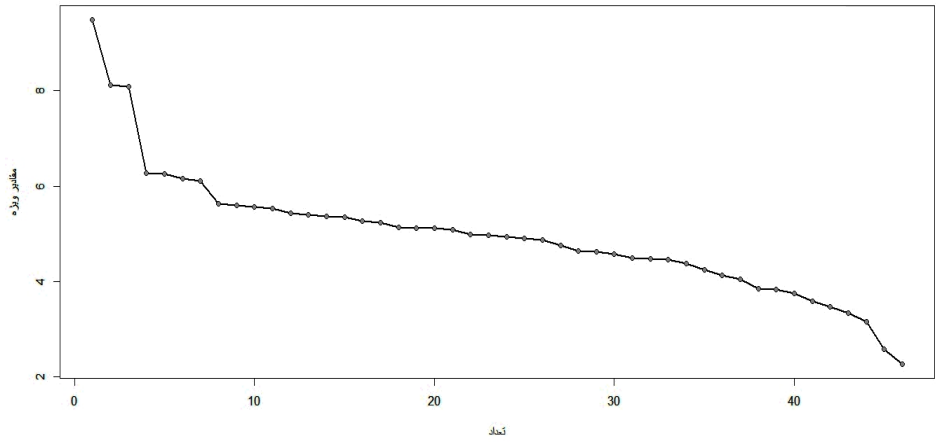
h	$RRMSE(L)$					
۱	۱/۰۲(۴۶)	۱/۰۱(۳۵)	۱/۰۱(۳۱)	۰/۹۹(۲۳)	۰/۹۸(۱۹)	۰/۹۸(۱۴)
۲	۱/۰۳(۴۴)	۱/۰۲(۳۷)	۱/۰۳(۳۰)	۰/۹۹(۲۴)	۰/۹۹(۲۰)	۰/۹۸(۱۵)
۳	۱/۰۲(۴۴)	۱/۰۲(۳۶)	۰/۹۸(۲۴)	۰/۹۸(۲۱)	۰/۹۸(۱۶)	۰/۹۹(۱۱)
۴	۱/۰۱(۴۲)	۱/۰۱(۳۷)	۰/۹۸(۳۱)	۰/۹۶(۱۹)	۰/۹۶(۱۵)	۰/۹۸(۱۲)
۶	۱/۰۱(۴۲)	۰/۹۵(۲۹)	۰/۹۴(۲۳)	۰/۹۴(۲۰)	۰/۹۵(۱۴)	۰/۹۶(۱۱)

بحث و نتیجه‌گیری

در این مقاله به منظور بهبود پیش‌بینی بازگشتی بدست آمده از روش SSA هنگامی که مشاهدات آلوده به نویز هستند، روش مبتنی بر معادلات فضای حالت و الگوریتم پالایش کالمن پیشنهاد گردید. همچنین دو روش دیگر برای بهبود



شکل ۱۱. نمودار سری زمانی داده‌های مصرف گاز در انگلستان از سه ماهه اول سال ۱۹۶۰ تا سه ماهه چهارم سال ۱۹۸۶



شکل ۱۲. نمودار رفتار مقادیر ویژه ماتریس مسیر

پیش‌بینی روش SSA با استفاده از پالایش داده‌ها و وزن دهی به ضرایب پیش‌بینی بازگشتی نیز معرفی شد. سپس برای ارزیابی آن‌ها با روش SSA پایه از معیار RMSE در مدل‌های ساختاری از مطالعات شبیه‌سازی شده و داده‌های واقعی استفاده شد. در این ارزیابی برای بررسی کارایی و تأثیر نویز بر نتایج پیش‌بینی از SNR متفاوت در افق‌های

جدول ۳. مقدار $RRMSE$ روش SSA-R وزنی نسبت به SSA-R برای مقادیر مختلف طول پنجره (L) در افق‌های پیش‌بینی مختلف

h	$RRMSE(L)$					
۱	۰/۹۸(۴۶)	۰/۹۸(۳۸)	۱/۰۱(۳۲)	۱/۰۳(۱۹)	۱/۰۴(۱۴)	۱/۰۴(۱۲)
۲	۰/۹۹(۴۴)	۰/۹۹(۳۷)	۱/۰۲(۳۰)	۱/۰۴(۲۰)	۱/۰۴(۱۵)	۱/۰۴(۱۱)
۳	۰/۹۹(۴۳)	۰/۹۹(۳۶)	۱/۰۲(۳۱)	۱/۰۲(۲۳)	۱/۰۴(۱۵)	۱/۰۴(۱۲)
۴	۰/۹۸(۴۲)	۰/۹۸(۳۷)	۱/۰۲(۲۴)	۱/۰۲(۲۰)	۱/۰۳(۱۵)	۱/۰۴(۱۱)
۶	۰/۹۸(۴۲)	۰/۹۹(۳۵)	۰/۹۹(۲۹)	۱/۰۲(۲۴)	۱/۰۲(۱۹)	۱/۰۳(۱۵)

جدول ۴. مقدار $RRMSE$ روش KF-SSA-R نسبت به SSA-R بازسازی شده برای مقادیر مختلف طول پنجره (L) در افق‌های پیش‌بینی مختلف

h	$RRMSE(L)$					
۱	۰/۹۴(۴۶)	۰/۹۴(۳۶)	۰/۹۲(۳۱)	۰/۹۲(۲۳)	۰/۹۰(۱۷)	۰/۹۰(۱۲)
۲	۰/۹۳(۴۴)	۰/۹۳(۳۷)	۰/۹۲(۲۹)	۰/۹۲(۲۴)	۰/۹۰(۲۰)	۰/۹۱(۱۱)
۳	۰/۹۴(۴۴)	۰/۹۴(۳۸)	۰/۹۳(۳۲)	۰/۹۱(۲۵)	۰/۹۱(۲۱)	۰/۸۹(۱۵)
۴	۰/۹۴(۴۳)	۰/۹۳(۳۷)	۰/۹۳(۲۴)	۰/۹۳(۲۰)	۰/۹۳(۱۵)	۰/۹۰(۱۲)
۶	۰/۹۳(۴۱)	۰/۹۴(۳۳)	۰/۹۴(۲۹)	۰/۹۲(۲۱)	۰/۹۱(۱۴)	۰/۸۹(۱۱)

جدول ۵. مقدار $RRMSE$ روش KF-SSA-R نسبت به SSA-R وزنی برای مقادیر مختلف طول پنجره (L) در افق‌های پیش‌بینی مختلف

h	$RRMSE(L)$					
۱	۰/۹۸(۴۶)	۰/۹۶(۳۶)	۰/۹۲(۳۱)	۰/۸۹(۲۳)	۰/۸۶(۱۷)	۰/۸۶(۱۲)
۲	۰/۹۷(۴۴)	۰/۹۶(۳۷)	۰/۹۱(۳۰)	۰/۸۸(۲۴)	۰/۸۶(۲۰)	۰/۸۶(۱۱)
۳	۰/۹۷(۴۴)	۰/۹۷(۳۸)	۰/۹۲(۳۲)	۰/۸۸(۲۵)	۰/۸۷(۲۱)	۰/۸۶(۱۵)
۴	۰/۹۷(۴۳)	۰/۹۴(۳۵)	۰/۸۸(۲۴)	۰/۸۷(۲۰)	۰/۸۶(۱۵)	۰/۸۶(۱۲)
۶	۰/۹۵(۴۲)	۰/۹۲(۳۳)	۰/۹۱(۲۹)	۰/۸۵(۲۱)	۰/۸۴(۱۴)	۰/۸۳(۱۱)

پیش‌بینی مختلف استفاده شد. نتایج مقایسه‌ها نشان می‌دهد، هنگامی که مشاهدات آلوده به نویز بوده و همچنین دارای مؤلفه‌های غیر قابل مشاهده مانند روند یا اثرات فصلی باشند، روش KF-SSA-R در مقایسه با روش SSA پایه، SSA بازسازی شده و SSA وزن دهی، هنگامی که طول پنجره کوچک باشد از کارایی بیشتری برخوردار است. دلیل این موضوع آن است که روش KF-SSA-R باعث تجزیه بهتر مؤلفه‌ها شده و در نتیجه موجب بهبود پیش‌بینی می‌شود. همچنین می‌توان نتیجه گرفت، هنگامی که طول پنجره مقدار بزرگی داشته باشد، روش SSA وزنی نسبت به روش SSA پایه در داده‌های تولید شده از مدل ساختاری کارایی بیشتری دارد. در مقایسه با روش‌های SSA بازسازی شده و SSA پایه، هنگامی که طول پنجره مقدار کوچکی داشته باشد، می‌توان نتیجه گرفت که روش SSA بازسازی شده کارایی بیشتری دارد.

تقدیر و تشکر

نویسندگان مقاله از سردبیر محترم مجله علوم آماری و داوران محترم همچنین ویراستار گرامی که باعث ارتقای مقاله شدند، کمال تشکر و قدردانی را دارند.

مراجع

ذبیحی مقدم، ر.، چینی‌پرداز، ر. و پرهام، غ. (۱۳۹۷)، شناسایی انواع تغییرات تأثیرگذار بر رفتار مدل‌های سری زمانی ساختاری با معادلات فضای حالت، مجله علوم آماری،۱۲(۱)،۱۴۳-۱۶۳.

Aydin, S., Saraoglu, H. M. and Kara, S. (2011), Singular Spectrum Analysis of Sleep EEG in Insomnia, Journal of Medical Systems, **35**(4), 457-461.

Bail, K. L., Gipson, J. M. and MacMillan, D. S. (2014), Quantifying the Correlation Between the MEI and LOD Variations by Decomposing LOD with Singular Spectrum Analysis, Earth on the Edge: Science for a Sustainable Planet International Association of Geodesy Symposia, **139**, 473-477.

Broomhead, D. S., King, G. P. (1986), Extracting Qualitative Dynamics from Experimental Data, Phys. D, Nonlinear Phenom, **20**, 217-236.

Golyandina, N. and Zhigljavsky, A. (2013), Singular Spectrum Analysis for Time Series, Springer Briefs in Statistics, Springer.

Hassani, H. and Kalantari, M. (2018), A Novel Signal Extraction Approach for Filtering and Forecasting Noisy Exponential Series, Comptes Rendus Mathematique, **356**, 563-570.

Hassani, H., Kalantari, M. and M. Yarmohammadi, M. (2017), An Improved SSA Forecasting Result based on a Filtered Recurrent Forecasting Algorithm, Comptes Rendus Mathematique, **355**, 1026-1036.

Hassani, H., Xu, Z. and Zhigljavsky, A. (2011), Singular Spectrum Analysis based on the Perturbation Theory, Nonlinear Analysis: Real World Applications, **12**, 2752-2766.

Hassani, H. and Mahmoudvand, R. (2018), Singular Spectrum Analysis Using R: Palgrave Advanced Texts in Econometrics.

Kalantari, M., Hassani, H. and Silva, E. S. (2020), Weighted Linear Recurrent Forecasting in Singular Spectrum Analysis, Fluctuation and Noise Letters, **19**(1), 2050010 (14 pages).

Kalman, R. E. (1960), A New Approach to Linear filtering and Prediction Problems, Trans. ASME, Journal of Basic Engineering, **83**, 35-45.

Liu, K., Law, S. S., Xia, Y. and Zhu, X. Q. (2014), Singular Spectrum Analysis for Enhancing the Sensitivity in Structural Damage Detection, Journal of Sound and Vibration, **333**(2), 392-417.

Mahmoudvand, R. and Rodrigues, P. C. (2017), A New Parsimonious Recurrent Forecasting Model in Singular Spectrum Analysis, Journal of Forecasting, 1-10. https://doi.org/10.1002/for.2484.

Mahmoudvand, R. and Rodrigues, P. C. (2016), Missing Value Imputation in Time Series Using Singular Spectrum Analysis, International Journal of Energy and Statistics, **4**(1), 1650005.

Sanei, S. and Hassani, H. (2016), Singular Spectrum Analysis of Biomedical Signals. Taylor & Francis/CRC.

Shumway, R. H. and Stoffer, D. S. (2011), Time Series Analysis and Application, 3rd ed, Springer, New York.

Wang, R., Ma, H. G., Liu, G. Q. and Zuo, D. G. (2015), Selection of Window Length for Singular Spectrum Analysis, Journal of the Franklin Institute, **352**, 1541-1560.