

Evaluation of the performance of the Singular Spectrum Analysis (SSA) algorithm in reconstructing missing data with different intensities in the hourly Land Surface Temperature time series

Hadi Zare khormizi¹, Mohammad Jafari^{2*}, Hamidreza Ghafarian malamiri¹, Ali Tavili¹, Hamidreza Keshtkar¹

¹. Department of Arid and Mountainous Regions Reclamation, Faculty of Natural Resources, University of Tehran, Karaj, Iran.

². Associate Professor, Department of Geography, Department of Environmental Planning, Yazd University, Yazd, Iran.

*Corresponding author: jafary@ut.ac.ir

Introduction and purpose: Generating Land Surface Temperature (LST) data with temporal and spatial continuity is in great demand for hydrology, meteorology, ecology, environment, and, etc. studies. Approximately, 70 to 90 percent of the Earth is covered by clouds at any given moment. Therefore, clouds, by creating an obstacle, absorb part of the thermal energy emitted from the earth by affecting thermal infrared energy, creating gaps and outliers in LST time series data. Removing the effect of cloud cover is always a common problem in the field of using satellite images. The purpose of this research is to evaluate the performance of Multi-channel Singular Spectrum Analysis (M-SSA) in order to reconstruct gaps and remove outlier data due to the cloud coverage in the hourly LST time series of the Meteosat-9 satellite.

Materials and methods: The study area in the present research was whole Iran. Also, the hourly LST time series of the SEVIRI sensor from the Meteosat-9 geostationary satellite in 2022 was used. At first, using SSA software and the Monte Carlo test, the window size and the number of significant components of an hourly LST time series were determined. Then, using the identified significant components, LST time series were reconstructed using M-SSA algorithm. Reconstruction error in clear sky conditions with available time series data and reconstruction error in cloudy sky conditions by creating artificial missing data (artificial cloud) with intensities of 10, 20, 30, ..., 90% in time series were evaluated using root mean square error (RMSE) and coefficient of determination (R^2) statistics.

Results: On average, in Iran, 20.0% of the hourly LST time series in 2022 was lost due to cloud cover, and the highest percentage of lost data was observed at the edge of the Caspian Sea. The results of analyzing the annual hourly LST time series in a window size of 96 hours with the Monte Carlo test showed that components 1 to 5 are significant components of this time series. These components control 99.0% of the LST time series variance. The frequency of the first, second-third, and fourth-fifth components are respectively 1, 1.04 and 1.08 cycles per image. The first component indicates annual periodic changes, the second and third components indicate 24-hour or daily temperature changes, and the fourth and fifth components indicate 12-hour periodic temperature changes. Based on the results, the RMSE and the R^2 between the original and the reconstructed data in clear sky conditions were 1.38 and 0.99 Kelvin, respectively. Also, in cloudy sky conditions, the RMSE error up to the level of 8% of randomly lost data (artificial cloud) was always less than 2.1 Kelvin.

Discussion and conclusion: The main key to reconstructing time series with periodic behavior is to identify significant periodic components and trends. In hourly LST time series, annual, 24-, 12- and 6-hour periods are the most important components of the time series. These components are formed due to the rotation of the earth around itself and the sun and the deviation of its axis. Therefore, these components are generally the same for the reconstruction of hourly LST time series in the major part of the globe. Based on the findings, M-SSA algorithm can be effective in reconstructing lost data with

large distance in LST time series due to consideration of periodic components and trends as well as using temporal and spatial correlation. One of the significant cases in reconstructing the effect of cloud cover in the present study and many other studies is the reconstruction of LST with the clear sky condition. Therefore, reconstruction of LST under cloud cover can be a challenge and suggestion for further studies in the future.

keywords: Cloud cover, Empirical Orthogonal Functions, Singular Spectrum Analysis, Temporal-spatial interpolation, Time series

ارزیابی عملکرد الگوریتم تجزیه تحلیل طیفی منفرد در بازسازی داده‌های از دست رفته با شدت‌های مختلف در

سری زمانی ساعتی دمای سطح زمین

هادی زارع خورمیزی^۱، محمد جعفری^{۱*}، حمیدرضا غفاریان مالامیری^۲، علی طویلی^۱، حمیدرضا کشت‌کار^۱

۱. گروه احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران.

۲. گروه جغرافیا، بخش برنامه‌ریزی محیطی، دانشگاه یزد، یزد، ایران.

*نویسنده عهده دار مکاتبات: jafary@ut.ac.ir

مقدمه و هدف: تولید داده‌های دمای سطح زمین (LST) با پیوستگی زمانی و مکانی برای مطالعات هیدرولوژی، هواشناسی، اکولوژی، زیست محیطی و غیره تقاضای زیادی دارد. تقریباً ۶۰ تا ۷۵ درصد کره زمین در هر لحظه توسط ابرها پوشیده شده است. بنابراین ابرها با ایجاد مانعی باعث جذب بخشی از انرژی حرارتی ساطع شده از زمین می‌شوند و با تاثیر بر انرژی مادون قرمز حرارتی منجر به ایجاد داده‌های از دست رفته و نامعتبر در سری‌های زمانی LST می‌شوند. رفع اثر پوشش ابر همواره یک مسئله رایج در زمینه استفاده از تصاویر ماهواره‌ای می‌باشد. هدف از پژوهش حاضر ارزیابی عملکرد الگوریتم تجزیه تحلیل طیفی منفرد-چندکاناله (M-SSA) به منظور بازسازی داده‌های از دست رفته و دور افتاده توسط پوشش ابر در سری زمانی LST ساعتی ماهواره Meteosat-9 می‌باشد.

مواد و روش‌ها: منطقه مطالعاتی در پژوهش حاضر کشور ایران در نظر گرفته شد. همچنین، از سری زمانی LST ساعتی سنجنده SEVIRI از ماهواره زمین ثابت Meteosat-9 در سال ۲۰۲۲ استفاده شد. در ابتدا با استفاده از نرم‌افزار SSA و آزمون مونت کارلو، اندازه پنجره و تعداد مولفه‌های معنی‌دار یک سری زمانی LST ساعتی تعیین شد. سپس با استفاده از مولفه‌های معنی‌دار شناسایی شده، سری‌های زمانی LST با استفاده از الگوریتم M-SSA بازسازی شد. خطای بازسازی در شرایط آسمان صاف با داده‌های موجود سری زمانی و خطای بازسازی در شرایط آسمان ابری با ایجاد داده‌های از دست رفته مصنوعی (ابر مصنوعی) با شدت‌های ۱۰، ۲۰، ۳۰، ...، ۹۰ درصد در سری زمانی با استفاده از آماره‌های خطای جذر میانگین مربعات (RMSE) و ضریب تبیین (R^2) اندازه گیری شد.

نتایج: به طور میانگین در سطح ایران ۲۵/۵ درصد از سری زمانی LST ساعتی در سال ۲۰۲۲ توسط پوشش ابر از دست رفته است که بالاترین میزان درصد داده‌های از دست رفته در حاشیه دریای خزر مشاهده شد. نتایج تجزیه و تحلیل سری زمانی LST ساعتی سالانه در اندازه پنجره ۹۶ ساعت با آزمون مونت کارلو نشان داد مولفه‌های ۱ تا ۵ جز مولفه‌های معنی‌دار این سری زمانی می‌باشند. این مولفه‌ها ۹۷/۵ درصد از تغییرات سری زمانی LST را تحت کنترل دارد. فرکانس مربوط به مولفه‌های اول، دوم با سوم، چهارم با پنجم به ترتیب ۰، ۰/۰۴۲ و ۰/۰۸۳ سیکل بر تصویر می‌باشد. مولفه اول نشان دهنده تغییرات دوره‌ای سالانه، مولفه دوم با سوم نشان دهنده تغییرات دوره‌ای ۲۴ ساعته یا روزانه دما، مولفه چهارم با پنجم نشان دهنده تغییرات دوره‌ای ۱۲ ساعته دما می‌باشد. بر اساس نتایج، خطای RMSE و R^2 بین داده‌های موجود و داده‌های بازسازی شده سری زمانی

LST ساعتی توسط الگوریتم M-SSA در شرایط آسمان صاف به ترتیب $1/38$ کلون و $0/99$ بود. همچنین در شرایط آسمان ابری خطای RMSE تا سطح 80 درصد داده از دست رفته به صورت تصادفی (ابر مصنوعی) همواره کمتر از $2/1$ کلون بود.

بحث و نتیجه‌گیری: کلید اصلی بازسازی سری‌های زمانی با رفتار دوره‌ای، شناسایی مولفه‌های معنی‌دار دوره‌ای و روندها می‌باشد. در سری‌های زمانی LST ساعتی، دوره‌های سالانه، 24 ، 12 و 8 ساعته از مهمترین مولفه‌های سری زمانی هستند. این مولفه‌ها ناشی از چرخش زمین به دور خود و خورشید و انحراف محور آن شکل می‌گیرند. بنابراین این مولفه‌ها عموماً برای بازسازی سری‌های زمانی LST ساعتی در قسمت عمده کره زمین یکسان هستند. بر اساس یافته‌ها، الگوریتم M-SSA با توجه به در نظر گرفتن مولفه‌های دوره‌ای و روندها و همچنین استفاده از همبستگی زمانی و مکانی می‌تواند در بازسازی داده‌های از دست رفته با فاصله بزرگ در سری‌های زمانی LST موثر باشد. یکی از موارد قابل توجه در بازسازی اثر پوشش ابر در پژوهش حاضر و بسیاری از پژوهش‌های دیگر، بازسازی دمای سطح زمین با شرط آسمان صاف است. بنابراین بازسازی دمای سطح زمین زیر پوشش ابر می‌تواند چالش و پیشنهادی برای مطالعات بیشتر در آینده باشد.

کلمات کلیدی: پوشش ابر، تجزیه تحلیل طیفی منفرد، توابع متعامد تجربی، درون‌یابی زمانی- مکانی، سری زمانی

۱- مقدمه

دمای سطح زمین^۱ (LST) نقش مهمی در فرآیندهای فیزیکی و بیولوژیکی سطح زمین در مقیاس جهانی و منطقه‌ای ایفا می‌کند (Yu et al., ۲۰۲۲). این پارامتر ارتباط نزدیکی با تغییرات در متغیرهای کلیدی اکوسیستم‌های زمین مانند محتوای بخار آب، رطوبت خاک، وضعیت تبخیر و تعرق، فرآیندهای فنولوژی و ... دارد (Yu et al., ۲۰۲۲). بنابراین، دمای سطح زمین به طور گسترده در زمینه‌های مطالعات اکولوژی، زیست محیطی، هیدرولوژی، هواشناسی، تولیدات کشاورزی و ... استفاده می‌شود (Helman et al., ۲۰۱۷; Zhou et al., ۲۰۱۷; Zare khormizi et al., ۲۰۲۱; Wang et al., ۲۰۲۱; Yu et al., ۲۰۲۲). یکی از روش‌های اندازه‌گیری دمای سطح زمین، اندازه‌گیری‌های زمینی و میدانی می‌باشد. این روش داده‌های پیوسته و قابل اعتمادی را در مکان اندازه‌گیری تولید می‌کند. با این حال این داده‌ها برای مناطق وسیع و ناهمگن با توزیع نامناسب ایستگاه‌های اندازه‌گیری قابل اعتماد نیست (Shiff et al., ۲۰۲۱). از طرفی تهیه و تولید این داده‌ها هزینه‌های مالی زیادی را در بر دارد. یکی از روش‌های نوین برای دستیابی به دمای سطح زمین استفاده از فناوری سنجنش از دور ماهواره‌ای می‌باشد. این فناوری دمای سطح زمین را با کمک سنجنده‌های حرارتی با وضوح زمانی و مکانی قابل قبول در سراسر جهان ارائه می‌دهد. در زمینه استفاده از فن‌آوری سنجنش از دور برای محاسبه دمای سطح زمین، همواره عواملی از قبیل ذرات معلق در هوا، گرد و غبار های جوی و حضور ابرها می‌توانند بر انرژی بازتابیده شده از سطح اثر قابل توجه بگذارند و به دنبال آن قرائت سنجنده‌های حرارتی را با خطا مواجه کنند (Julien & Sobrino, ۲۰۱۰). تقریباً 60 تا 75 درصد کره زمین در هر لحظه توسط ابرها پوشیده شده است (Kleipool et al., ۲۰۰۸). ابرها با ایجاد مانعی باعث جذب بخشی از انرژی حرارتی ساطع شده از زمین می‌شوند و با تاثیر بر انرژی مادون قرمز حرارتی منجر به ایجاد دمای پایین در تصاویر دمای سطح زمین می‌شوند (Ghafarian et al., ۲۰۱۸). در مراحل بعد ممکن است اثرات پوشش ابر در تصاویر دمای سطح زمین توسط الگوریتم‌های ماسک ابر^۲ شناسایی و حذف شود. در نتیجه در تصاویر و سری‌های زمانی دمای سطح زمین داده‌های از دست رفته^۳ و داده‌ها با مقادیر غیر قابل قبول^۴

^۱ Land Surface Temperature

^۲ Cloud Mask Algorithm

^۳ Gap

^۴ Outlier

حضور خواهند داشت. این مشکل باعث می‌شود روش‌های سنجش از دور با توجه به اثر پوشش ابر همواره در هر زمان و مکان قادر به ارائه دمای سطح زمین نباشند.

بیش از ۲۰ روش به منظور رفع اثر پوشش ابر و بازیابی داده‌های از دست رفته و دور افتاده در سری‌های زمانی تصاویر ماهواره‌ای (شاخص‌های گیاهی یا دمای سطح زمین) وجود دارد (Geng et al., ۲۰۱۴). به طور کلی این روش‌ها را می‌توان به سه دسته زمانی، مکانی و در نهایت زمانی - مکانی تقسیم نمود (Chen et al., ۲۰۲۲). در روش‌های زمانی، مقادیر داده‌های از دست رفته و نامعتبر بر اساس حرکت در زمان تخمین زده می‌شوند و مقادیر پیکسل‌های مجاور جغرافیایی در نظر گرفته نمی‌شوند. برای مثال، الگوریتم تجزیه و تحلیل هارمونیک سری‌های زمانی^۱ (HANTS) بر اساس رویکرد زمانی برای بازسازی سری‌های زمانی LST و شاخص NDVI^۲ بکار گرفته شده است (Zhou et al., ۲۰۱۳; Xu & Shen, ۱۹۹۶; Verhoef et al., ۲۰۱۵). روش‌های مکانی فقط مقادیر پیکسل‌های مجاور را در نظر می‌گیرند و مقادیر در دوره‌های زمانی مختلف را شامل نمی‌شوند (Pede & Mountrakis, ۲۰۱۸). از روش مکانی می‌توان روش‌های زمین آماری مانند کوکریجینگ (Zhang et al., ۲۰۱۳; Ke et al., ۲۰۰۹) یا رگرسیون‌های خطی یا چند جمله‌ای در مکان (Mukherjee et al., ۲۰۱۴) را نام برد. روش‌های مکانی به راحتی قابل تحقق‌اند، اما دارای کاستی‌های خاصی مانند عدم وضوح و دقت تصاویر بازسازی شده هستند (Cui et al., ۲۰۲۲). در روش‌های زمانی - مکانی هم تغییرات زمانی و هم تغییرات مکانی در بازیابی داده‌های ابری و از دست رفته مورد استفاده قرار می‌گیرند. از روش‌های درون‌یابی زمانی - مکانی برای بازیابی مقادیر داده‌های از دست رفته و نامعتبر در سری‌های زمانی می‌توان روش تجزیه تحلیل طیفی منفرد چند کاناله (M-SSA^۳) اشاره نمود (Ghafarian et al., ۲۰۱۸).

یکی از استراتژی‌های غالب در پر کردن فاصله داده‌های از دست رفته در روش‌های زمانی و مکانی - مکانی، شناسایی رفتار دوره‌ای و روندها و پارامترهای تغییر در سری‌های زمانی و سپس بکارگیری این پارامترها برای بازسازی سری زمانی می‌باشد (Cui et al., ۲۰۲۲). سری‌های زمانی LST دارای اجزا منظم (چرخه‌ای^۴) اجزا نامنظم (تویز) می‌باشند. اجزا یا مولفه‌های منظم ناشی از حرکت زمین به دور خود و خورشید و تغییرات فصلی ناشی از انحراف محور کره زمین و اجزا نامنظم یا نویز ناشی از فعل و انفعالات تعامل میان انرژی تابشی و انرژی آشفته محیط می‌باشد (Ghafarian et al., ۲۰۱۸). الگوریتم تجزیه تحلیل طیفی منفرد (SSA) از همبستگی زمانی (تک کاناله: SSA) و همبستگی زمانی - مکانی (چند کاناله: M-SSA) برای بازسازی سری‌های زمانی استفاده می‌نماید. این الگوریتم یک سری زمانی را به مولفه‌های ساده‌تر و قابل تفسیر مانند روندها، نوسانات دوره‌ای و نیمه دوره‌ای و نویزها تجزیه می‌کند (Golyandina et al., ۲۰۰۱).

پژوهش‌هایی به بررسی دقت و کارایی روش‌های مختلف رفع داده‌های از دست رفته و نامعتبر توسط پوشش ابر در سری‌های زمانی LST پرداخته‌اند. در مطالعه‌ای دقت روش‌های زمانی، مکانی و زمانی - مکانی در بازسازی اثر پوشش ابر در سری‌های زمانی LST سنجنده MODIS^۵ در حوضه رودخانه هیه^۶ چین مورد بررسی قرار گرفت. بر اساس نتایج این مطالعه روش درون‌یابی زمانی خطی، بالاترین دقت را با خطای جذر میانگین مربعات ۷/۱ کلوین داشت (Chen et al., ۲۰۲۲). در مطالعه‌ای دیگر الگوریتم از HANTS برای بازسازی و پر کردن فاصله داده‌های از دست رفته و نامعتبر در سری زمانی ۸ روزه LST سنجنده MODIS استفاده شد. نتایج این مطالعه نشان داده الگوریتم HANTS به طور موثر می‌تواند کیفیت داده‌های دمای سطح زمین

^۱ Harmonic Analysis of Time Series

^۲ Normalized Difference Vegetation Index

^۳ Multi-Singular Spectrum Analysis

^۴ Cycles

^۵ Trend

^۶ Moderate Resolution Imaging Spectroradiometer

^۷ Heihe river

سنجنده MODIS را بهبود بخشد (Xu & Shen, ۲۰۱۳). در مطالعه‌ای دیگر داده‌های از دست رفته و نامعتبر توسط پوشش ابر در محصول LST سنجنده MODIS با استفاده از داده‌های آسمان صاف پیسکل‌های مجاور در زمان بازسازی شد (Sun et al., ۲۰۱۷). در پژوهشی روش درون‌یابی داده‌ها از طریق توابع متعامد تجربی^۱ (DINEOF) برای بازسازی و پر کردن فاصله داده‌های از دست رفته و نامعتبر در سری‌های زمانی LST سنجنده MODIS در فلات تبت مورد استفاده قرار گرفت. در این پژوهش خطای بازسازی داده‌های از دست رفته در تصاویر LST در زمان روز و شب به ترتیب ۴/۵ و ۲/۳ کلوین بود (Zhou et al., ۲۰۱۷). در پژوهشی دیگر در فلات تبت چین، با استفاده از الگوریتم M-SSA سری زمانی داده‌های ساعتی دمای سطح زمین ماهواره زمین ثابت Fengyun-۲C بازسازی شد؛ نتایج این مطالعه نشان داد در یک سری زمانی ساعتی با میانگین ۶۳ درصد داده از دست رفته توسط پوشش ابر، خطای میانگین مطلق^۲ ۲/۲۵ درجه کلوین است (Ghafarian et al., ۲۰۱۲). در بازسازی سری‌های زمانی روزانه LST سنجنده MODIS با استفاده از الگوریتم HANTS در منطقه مطالعاتی ایران نشان داده شد، خطای جذر میانگین مربعات^۳ (RMSE) بین داده‌های غیر ابری و بازسازی شده در سری‌های زمانی LST در زمان روز و شب به ترتیب ۳/۸۷ و ۲/۶۸ کلوین می‌باشد (Ghafarian & Zare khormizi, ۲۰۱۷).

هدف از پژوهش حاضر ارزیابی الگوریتم M-SSA به منظور بازسازی داده‌های از دست رفته و دور افتاده توسط پوشش ابر در سری زمانی LST ساعتی ماهواره زمین ثابت^۹ Meteosat می‌باشد. با توجه به اینکه تولید داده‌های دمای سطح زمین با پیوستگی زمانی و مکانی برای مطالعات هیدرولوژی، هواشناسی، اکولوژی، مطالعات زیست محیطی و غیره تقاضای زیادی دارد؛ نتایج این پژوهش می‌تواند در اغلب این مطالعات با توجه به داده‌های پیوسته و خالص تولید شده کاربرد داشته باشد.

۲- مواد و روش‌ها

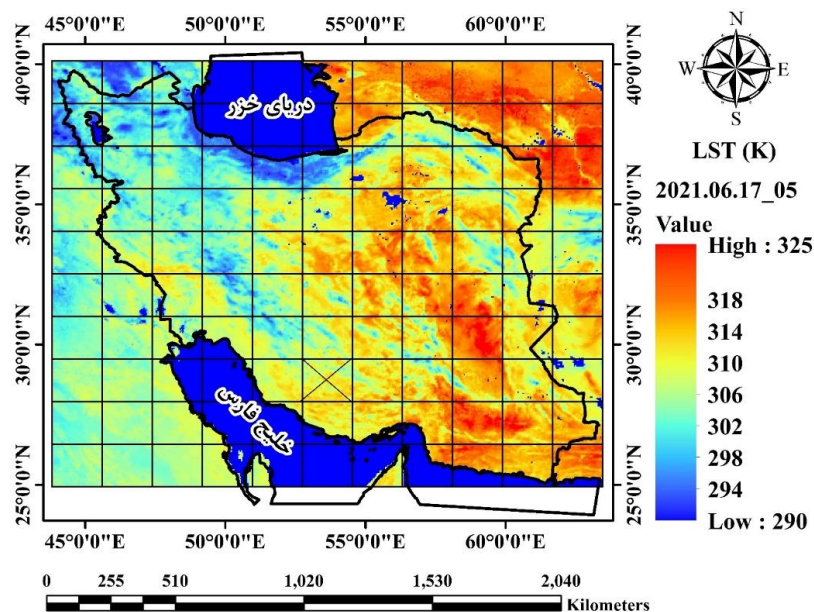
۲-۱- معرفی منطقه مورد مطالعه

منطقه مطالعاتی در پژوهش حاضر کل سطح کشور ایران در نظر گرفته شد. علت این انتخاب، داشتن یک منطقه وسیع با میزان داده‌های از دست رفته با شدت‌های مختلف در سری‌های زمانی LST ساعتی می‌باشد. شکل ۱ برای نمونه، یک تصویر LST در تاریخ ۲۷ خرداد ۱۴۰۱ (۲۰۲۲/۰۶/۱۷) در ساعت ۵ صبح را نشان می‌دهد.

^۱ Data Interpolating Empirical Orthogonal Functions (DINEOF)

^۲ Mean Absolute Error

^۳ Root Mean Square Error



شکل ۱. منطقه مورد مطالعه به همراه یک تصویر دمای سطح زمین در تاریخ ۱۴۰۱/۰۳/۲۷ در ساعت ۵ صبح، به همراه شبکه ایجاد شده برای پردازش جز به جز سری زمانی

۲-۲- تصاویر ماهواره‌ای

در پژوهش حاضر از محصول دمای سطح زمین ماهواره زمین ثابت^۱ (IODC)^۱ -Meteosat استفاده شد.^۲ این محصول دارای قدرت تفکیک مکانی ۵ کیلومتر و قدرت تفکیک زمانی ساعتی می‌باشد. مجموعه ماهواره‌های Meteosat از جمله ماهواره‌های هواشناسی با هدف پیش‌بینی و تشخیص سریع آب و هوا در چند ساعت آینده می‌باشند. ماهواره -Meteosat ۹ از ماهواره‌های نسل دوم Meteosat می‌باشد (MSG^۳). سنجنده SEVIRI^۴ ماهواره -Meteosat ۹ در ۱۲ باندطیفی و هر ۱۵ دقیقه (در حالت دید کامل^۵) تصویربرداری می‌کند. همچنین تصاویر دمای سطح زمین سنجنده SEVIRI ماهواره -Meteosat ۹ با استفاده از الگوریتم پنجره مجزای دو کاناله^۶ محاسبه می‌شوند (Martins et al., ۲۰۲۳; Freitas et al., ۲۰۱۳).

۲-۳- اجرای الگوریتم تجزیه تحلیل طیفی منفرد (SSA)

به منظور بازسازی سری زمانی یکساله LST ساعتی با ۸۷۶۰ تصویر (۳۶۵ روز × ۲۴ ساعت) در طول سال ۲۰۲۲ از نرم‌افزار SSA-MTM استفاده شد. این نرم‌افزار از لینک پیوست^۷ به صورت رایگان قابل دسترسی می‌باشد. برای بازسازی سری‌های زمانی در نرم‌افزار SSA ابتدا لازم است؛ اندازه پنجره^۸ و تعداد مولفه‌های معنی‌دار^۹ تعیین شود. به منظور تعیین این پارامترها ابتدا یک سری زمانی تکی یا برداری بدون داده از دست رفته به عنوان یک نماینده از کل سری زمانی انتخاب شد. سپس اندازه پنجره ۹۶

^۱ Indian Ocean Data Coverage (IODC)

^۲ <https://land.copernicus.eu/global/products/lst>

^۳ Meteosat Second Generation

^۴ Spinning Enhanced Visible and InfraRed Imager (SEVIRI)

^۵ Full view

^۶ Generalized Split-Window algorithm from two adjacent channels

^۷ <http://research.atmos.ucla.edu/tcd/ssa/form.html>

^۸ Window size

^۹ Significant components

ساعت در نظر گرفته شد. در مرحله بعد با استفاده از آزمون‌های مختلف در نرم‌افزار SSA-MTM نظیر آزمون مونت کارلو^۱ تعداد مولفه‌های معنی‌دار در اندازه پنجره ۹۶ ساعت مشخص شد؛ در بخش نتایج به طور مفصل به این مورد پرداخته خواهد شد. همچنین لازم به ذکر است با توجه به طولانی بودن بحث آزمون‌های نرم‌افزار SSA نظیر آزمون مونت کارلو و آزمون فرض صفر EOFs^۲ و بحث تئوری مربوط به روندها و دوره‌ها به می‌توان برای کسب اطلاعات بیشتر به پژوهش غفاریان (۲۰۱۵) مراجعه شود. اندازه پنجره در پژوهش حاضر، ۹۶ ساعت (۴ روز) در نظر گرفته شد. زیرا کوچک در نظر گرفتن اندازه پنجره در مناطق با داده‌های از دست رفته با فاصله طولانی منجر به ایجاد سری زمانی نامعتبر خواهد شد و از طرفی هنگامی که اندازه پنجره بزرگ در نظر گرفته شود به شدت منابع حافظه موقت و پردازنده کامپیوتر درگیر خواهد شد و مدت زمان بازسازی افزایش خواهد یافت. در نرم‌افزار SSA-MTM سری‌های زمانی می‌تواند به صورت برداری و ماتریسی مورد پردازش قرار گیرد. هنگامی که بخواهیم یک سری زمانی n تکی را مورد پردازش قرار دهیم از الگوریتم SSA استفاده می‌شود و هنگامی که تعداد n تا سری زمانی در کنار هم به صورت ماتریس قرار داشته باشند در حالت M -SSA^۳ پردازش می‌شوند. پس از تعیین اندازه پنجره و مولفه‌های معنی‌دار، سری‌های زمانی LST ساعتی سالانه با استفاده از M -SSA بازسازی شد. لازم به ذکر است با توجه به وسعت منطقه مورد مطالعه و اندازه پیکسل سائز تصاویر LST (۵ کیلومتر)، منطقه مورد مطالعه ۳۴۰ پیکسل در ردیف و ۴۴۰ پیکسل در ستون دارد و با توجه به داده‌های LST ساعتی، در سال ۸۷۶۰ تصویر دمای سطح زمین وجود خواهد داشت. پردازش یکجای این سری زمانی با استفاده از الگوریتم M -SSA با استفاده از کامپیوترهای معمولی امکان پذیر نیست. بنابراین در این مطالعه با توجه به طرح شبکه بندی نشان داده شده در شکل ۱، این سری زمانی به قطعات کوچکتر تقسیم و هر قسمت جداگانه پردازش شد. در ادامه مبانی نظری الگوریتم SSA توضیح داده می‌شود.

۲-۴- مبانی نظری الگوریتم تجزیه تحلیل طیفی منفرد (SSA)

تجزیه تحلیل طیفی منفرد در ابتدا توسط بروم هید و کینگ^۴، (۱۹۸۶) توسعه داده شد؛ سپس این الگوریتم توسط ووتارد و گیل^۵، (۱۹۹۲) و کاندراشو و گیل^۶، (۲۰۰۶) برای پر کردن فاصله داده‌های از دست رفته یا گم شده در سری‌های زمانی پیشنهاد داده شد. عملکرد الگوریتم تجزیه تحلیل طیفی منفرد در بازسازی سری‌های زمانی را می‌توان در چهار مرحله توصیف نمود (Ghafarian et al., ۲۰۱۸; Musial et al., ۲۰۱۱).

۱. در مرحله اول یک سری زمانی اسکالر n ، $1, 2, 3, \dots, n$ به یک ماتریس تراژکتوری^۷ چند بعدی از بردارهای $X = [f_1, \dots, f_k]$ تبدیل می‌شود که در این ماتریس $k = n - m + 1$ است (اندازه پنجره m). هر بردار این ماتریس به صورت $X_j = (f_j, \dots, f_{j+m-1})$; $j = 1, \dots, k$ تعریف می‌شود. این ماتریس در واقع حاوی رکوردهای کامل الگوهای موجود در اندازه پنجره زمانی ($m < n$) است (رابطه ۱).

^۱ Monte carlo test

^۲ Empirical Orthogonal Functions

^۳ Multi-channel SSA

^۴ Broomhead & King

^۵ Vautard & Ghil

^۶ Kondrashov & Ghil

^۷ Trajectory

$$X = \begin{pmatrix} f_1 & f_2 & f_3 & \dots & f_k \\ f_2 & f_3 & f_4 & \dots & f_{k+1} \\ f_3 & f_4 & f_5 & \dots & f_{k+2} \\ \dots & \dots & \dots & \dots & \dots \\ f_m & f_{m+1} & f_{m+2} & \dots & f_n \end{pmatrix} \quad \text{رابطه (۱)}$$

۲. در مرحله دوم ماتریس $m \times k$ تراژکتوری X با استفاده از روش SVD^1 تجزیه می‌شود. محصول این تجزیه به صورت رابطه ۲ خواهد بود.

$$X = DLE^T \quad \text{رابطه (۲)}$$

در رابطه ۲، D و E بردارهای چپ و راست تکین ماتریس X به ترتیب با اندازه $m \times m$ و $k \times k$ هستند. همچنین پارامتر L ، یک ماتریس قطری مستطیلی با ابعاد $m \times k$ است. عناصر ماتریس L ، به عنوان مقادیر منحصر به فرد، ریشه‌های مربع مقادیر ویژه^۲ ماتریس، $S = XX^T$ با اندازه $m \times m$ می‌باشند. ستون‌های ماتریس D بردارهای ویژه^۳، S هستند که به عنوان توابع متعامد تجربی (EOFs) شناخته می‌شوند. همچنین ردیف‌های ماتریس E^T بردارهای ویژه ماتریس XX^T هستند. اگر مقادیر اعداد تکین^۴ را به ترتیب صعودی به نزولی ترسیم کنیم، اغلب یک نمودار با یک شیب تند اولیه، که نشان دهنده یک سیگنال و یک سطح صاف با شیب کم بدست می‌آید که نشان دهنده نویزها است (Vautard et al., ۱۹۹۲). هر زیر مجموعه از بردارهای ویژه d (EOFs)، $1 \leq d \leq m$ ، برای هر بردار ویژه مربوطه توصیف‌های مثبتی از بهترین نمایش ماتریس X به عنوان مجموعه ماتریس‌های X_i ، $i=1, 2, \dots, d$ هستند (Golyandina et al., ۲۰۰۱).

۳. در مرحله سوم d به تعداد p زیر مجموعه متمایز تقسیم می‌شود. در نتیجه جمع همه مولفه‌ها در داخل هر زیر مجموعه به صورت رابطه ۳ است.

$$X = \sum_{n'=1}^p X_{I_{n'}} \rightarrow X_{I_n} = \sum_{i \in I_{n'}} X_i \quad \text{رابطه (۳)}$$

در یک حالت ایده آل ماتریس‌های X_{I_n} به صورتیک ماتریس هانکل^۵ هستند و در نتیجه با ماتریس‌های تراژکتوری منطبق می‌شوند.

۴. از آنجایی که حالت ایده‌آلی که در مرحله ۳ توصیف شد، معمولاً رخ نمی‌دهد، در مرحله چهارم ماتریس‌های X_{I_n} باید به شکل یک ماتریس هانکل برای انطباق به ماتریس‌های تراژکتوری تبدیل شوند. این مرحله به عنوان میانگین‌گیری قطری شناخته می‌شود. به این معنا که ماتریس اصلی می‌تواند به صورت مجموع این ماتریس‌ها بازسازی شود (رابطه ۴).

$$x_t = \sum_{n'=1}^p x_t^{n'}, \quad t=0, 1, \dots, n-1 \quad \text{رابطه (۴)}$$

در رابطه ۴، برای هر p سری $X_t^{n'}$ نتیجه میانگین قطری ماتریس X_{I_n} است.

SSA چند کاناله (M-SSA) زمانی که بیش از یک سری زمانی وجود داشته باشد، استفاده می‌شود. M-SSA تعداد L پروفایل مکانی زمان (که $L \leq m$) اندازه پنجره انتخاب شده در SSA است) را به کار می‌گیرد و از اطلاعات مکانی (برخی از مولفه‌های معنی‌دار مکانی PCA) همراه با اطلاعات زمانی (T-PCA) برای بازسازی دقیق‌تر سری‌های زمانی به ویژه زمانی که

^۱ Singular Value Decomposition

^۲ Eigenvalues

^۳ Eigenvectors

^۴ Singular Values

^۵ Hankel Matrix

سری زمانی دارای داده‌های از دست رفته طولانی هستند، استفاده می‌کند. اگر بیش از یک سری زمانی از مشاهدات داشته باشیم، در این صورت X_{li} که $i = 1, 2, \dots, N$ و $l = 1, 2, \dots, L$ می‌باشد و تعمیم SSA برای ساخت ماتریس کوواریانس چند متغیره (T) به صورت رابطه ۵ خواهد بود:

$$T = \begin{bmatrix} T_{1,1} & T_{1,2} & \dots & T_{1,L} \\ T_{2,1} & T_{2,2} & \dots & \\ \vdots & \vdots & \dots & \\ T_{L,1} & T_{L,2} & \dots & T_{L,L} \end{bmatrix} \quad \text{رابطه (۵)}$$

در رابطه ۵، $T_{1,l}$ ماتریس کوواریانس چند متغیره بین کانال l و l' است. $T_{1,l}$ مانند S برای SSA تک کاناله است ($l = 1$). با قطری کردن ماتریس T ، تعداد $L \times m$ بردار ویژه و مقدار ویژه از ماتریس T توسط M-SSA ایجاد می‌شود. هر بردار ویژه (E^k)، مولفه اصلی مکان-زمان (ST-PCs)^۱ نامیده می‌شود، که مانند SSA با انتخاب کردن p از ST-PCs، سری زمانی اصلی می‌تواند بازسازی شود.

عملکرد SSA برای پر کردن شکاف داده‌های از دست رفته در سری‌های زمانی به صورت زیر است (Ghafarian Malamiri et al., ۲۰۱۸):

در یک سری زمانی با یک پنجره به سائز (m) سری زمانی اصلی با محاسبه مقادیر میانگین بدون اریب، در مرکز قرار می‌گیرد و داده‌های از دست رفته به صورت صفر، تنظیم می‌شوند. اولین EOF توسط یک روند تکراری که الگوریتم SSA روی داده‌های صفر و متمرکز شده اعمال می‌کند، یافت می‌شود. مقادیر از دست رفته بر اساس مولفه بازسازی شده EOF فعلی، به روز رسانی می‌شوند. سپس الگوریتم SSA دوباره بر روی این مجموعه به روز شده اعمال و EOF دوباره محاسبه و مقادیر از دست رفته، بازسازی می‌شوند. این روند تا زمانی که نتایج آزمون همگرایی رضایت‌بخش باشد، تکرار می‌شود. تا زمانی که تفاوت بین دو مقدار متوالی خطای جذر میانگین مربعات نرمال شده (RMSE) کمتر از ۲/۵ درصد شود، RMSE بین داده‌های بازسازی شده جدید و مقادیر مشخص داده‌های از دست رفته از تکرار قبلی محاسبه می‌شود (Kondrashov et al., ۲۰۱۰). سپس فرآیند تکرار برای دومین EOF پیشرو، شروع می‌شود و تا زمانی که همگرایی برای EOF دوم بدست آید، ادامه می‌یابد. برای به دست آوردن مقدار بهینه برای اندازه پنجره و تعداد حالت‌های غالب SSA، برای پر کردن داده‌های از دست رفته، اعتبار سنجی متقابل اعمال می‌شود. برای انتخاب بهترین اندازه پنجره و تعداد EOFها بخشی از داده‌های موجود (انتخاب شده به صورت تصادفی) تبدیل به داده از دست رفته می‌شود و سپس خطای RMSE بازسازی، محاسبه می‌شود. در الگوریتم SSA هیچ راه مستقیمی برای حذف داده‌های نامعتبر وجود ندارد، با این حال با انتخاب تعداد کم مولفه‌های معنی‌دار، سیگنال بازسازی شده صاف می‌شود و داده‌های نامعتبر به‌طور غیرمستقیم هموار می‌شوند. اما باید توجه داشت که داده‌های نامعتبر بر سیگنال بازسازی تاثیر می‌گذارند.

۲-۵- ارزیابی عملکرد الگوریتم M-SSA

به منظور ارزیابی عملکرد الگوریتم M-SSA در بازسازی تصاویر LST از خطای جذر میانگین مربعات^۲ (RMSE) و ضریب تبیین (R^2) استفاده شد. RMSE با استفاده از رابطه ۶ و ضریب تبیین با استفاده از روابط ۷، ۸ و ۹ محاسبه شد. در این روابط نام برده، y_i ، $\hat{y}_i$ و SS_{tot} به ترتیب داده‌های واقعی و داده‌های مدل شده، مجموع مربعات خطا و مجموع مربعات کل می‌باشند.

^۱ Space-Time Principal Components

^۲ Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad \text{رابطه (۶)}$$

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} \quad \text{رابطه (۷)}$$

$$SS_{err} = \sum_{i=1}^n (x_i - y_i)^2 \quad \text{رابطه (۸)}$$

$$SS_{tot} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{رابطه (۹)}$$

عملکرد الگوریتم M-SSA در بازسازی سری زمانی ساعتی LST به دو صورت مورد بررسی قرار گرفت. در حالت اول بین داده‌های واقعی موجود در سری زمانی LST (آسمان صاف) و داده‌های بازسازی شده توسط الگوریتم M-SSA خطای RMSE و ضریب تبیین اندازه گیری شد. لازم به ذکر است در این حالت عملکرد الگوریتم M-SSA در پر کردن نقاط بدون داده (آسمان ابری) همچنان مبهم خواهد بود. در واقع اگر بخواهیم عملکرد الگوریتم M-SSA در مناطق با داده‌های از دست رفته را بررسی کنیم (خطای پر کردن)^۱؛ باید به داده از دست رفته یا داده‌های زمینی دسترسی داشته باشیم. برای رفع این مشکل و بررسی عملکرد الگوریتم M-SSA در پر کردن نقاط بدون داده یک ماتریس از سری زمانی ساعتی LST سالانه با اندازه ۱۳۶۰ در ۸۷۶۰ انتخاب شد. به عبارت دیگر ۱۳۶۰ سری زمانی LST ساعتی سالانه بدون داده دست رفته (در مناطق جنوبی ایران با کمترین شدت ابرناکی) انتخاب شد (شکل ۱، قطعه مشخص شده با ضریب). سپس در این سری زمانی ماتریسی به طور تصادفی و با شدت‌های مختلف ۱۰، ۲۰، ۳۰، ۴۰، ...، ۹۰ درصد داده‌های از دست رفته به طور مصنوعی یا ابر مصنوعی ایجاد شد. همان طور که ذکر شده این سری زمانی ۸۷۶۰ داده در سال دارد. نقطه شروع ایجاد داده از دست رفته (گپ) به صورت تصادفی از عدد ۱ تا ۸۷۶۰ انتخاب شد و سپس این روند تا رسیدن به سطح داده از دست رفته دلخواه (۱۰، ۲۰، ...، ۹۰٪) در نرم افزار MATLAB ادامه یافت. سپس این سری‌های زمانی با داده‌های از دست رفته مصنوعی با استفاده از الگوریتم M-SSA بازسازی شد و در نهایت بین داده‌های از دست رفته به صورت مصنوعی و داده‌های بازسازی شده توسط الگوریتم M-SSA خطای RMSE و ضریب تبیین محاسبه شد. در روش دوم عملکرد الگوریتم M-SSA در بازسازی داده‌های از دست رفته با شدت‌های مختلف مورد بررسی قرار گرفت.

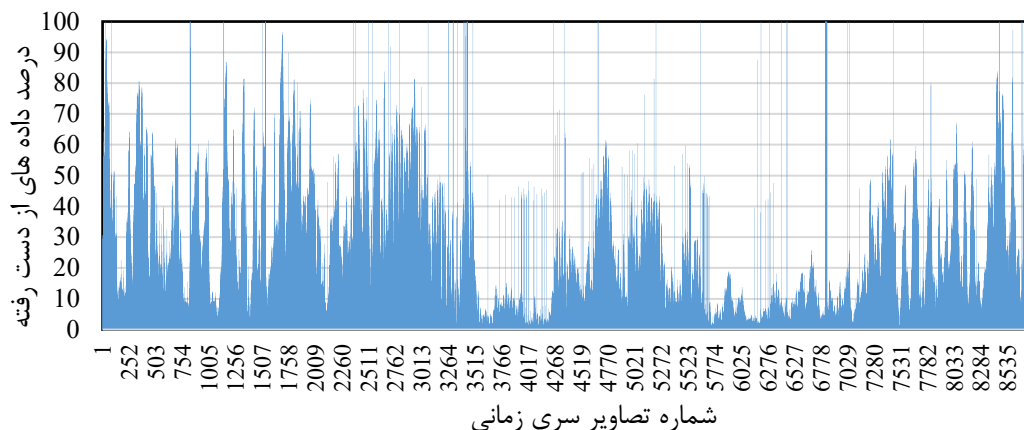
۳- یافته‌های پژوهش

۳-۱- بررسی میزان داده‌های از دست رفته در زمان و مکان

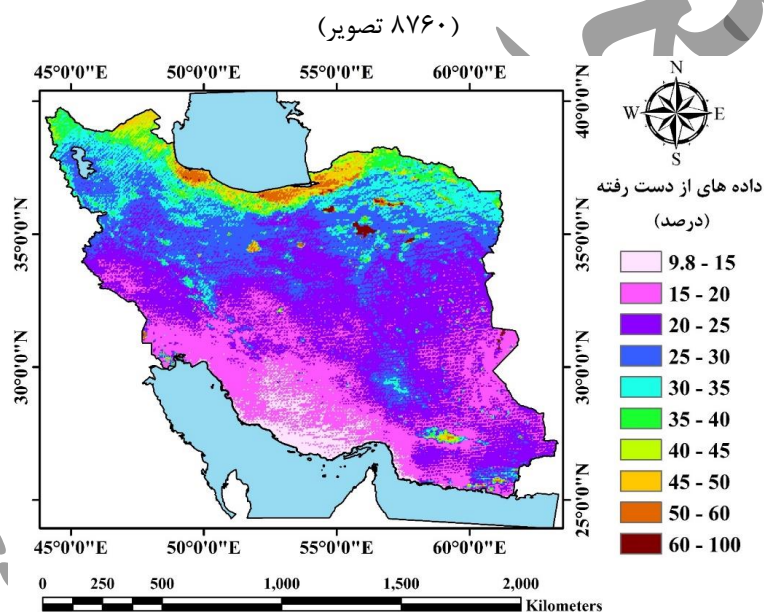
به منظور بررسی عملکرد الگوریتم M-SSA، در ابتدا میزان و نحوه پراکنش داده‌های از دست رفته در سری زمانی LST ساعتی مورد بررسی قرار گرفت. میزان داده‌های از دست رفته در هر تصویر سری زمانی LST سالانه در شکل ۲ نشان داده شده است. با توجه به شکل ۲ تقریباً در تمام سال داده‌های از دست رفته وجود دارد (پراکنش زمانی). نقشه مکانی درصد داده‌های از دست رفته در سری زمانی ساعتی دمای سطح زمین در طول یک سال در شکل ۳ نشان داده شده است. با توجه به شکل نام برده، با حرکت از مناطق شمالی ایران به سمت جنوب میزان داده‌های از دست رفته کاهش می‌یابد. به عبارت دیگر در مناطق شمالی ایران شدت ابرناکی یا روزهای ابری بیشتر از مناطق جنوبی ایران می‌باشد. حداکثر میزان داده‌های از دست رفته در سری زمانی

^۱ Gap filling error

ساعتی LST در حاشیه دریای خزر مشاهده شد. به طور میانگین در سطح ایران ۲۵/۵ درصد از سری زمانی LST ساعتی در سال ۲۰۲۲ توسط پوشش ابر از بین رفته است.



شکل ۲. میزان داده‌های از دست رفته‌ی هر تصویر (پراکنش زمانی) در سری زمانی ساعتی دمای سطح زمین در طول یک سال

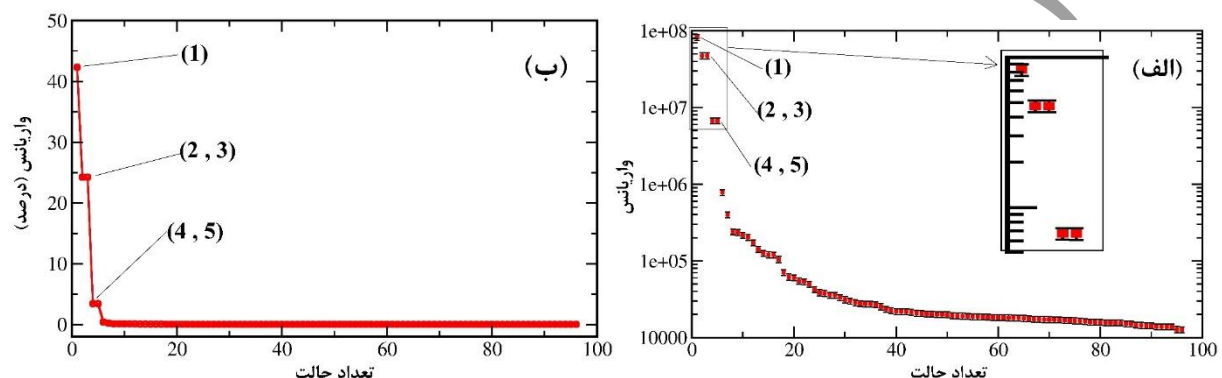


شکل ۳. نقشه مکانی درصد داده‌های از دست رفته در سری زمانی ساعتی دمای سطح زمین در طول یک سال

۳-۲- بررسی روند و دوره‌ها در سری زمانی ساعتی دمای سطح زمین

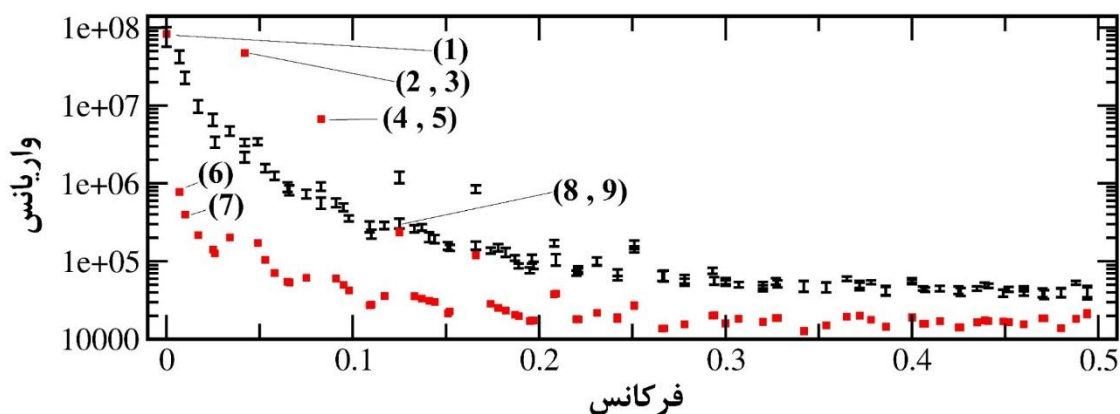
همان طور که در بخش مواد و روش‌ها توضیح داده شده به منظور پر کردن فاصله داده‌های از دست رفته یا گم شده در سری‌های زمانی با استفاده از الگوریتم SSA در ابتدا لازم است اندازه پنجره و سپس تعداد مولفه‌های معنی‌دار سری‌های زمانی در آن اندازه پنجره تعیین شود. منحنی اعداد ویژه با اندازه پنجره ۹۶ ساعت در طول یک سری زمانی LST ساعتی سالانه (۸۷۶۰ داده) در شکل ۴ الف نشان داده شده است. با توجه به اینکه اندازه پنجره ۹۶ ساعت انتخاب کردیم در محور عمودی شکل ۴ الف و ب، ۹۶ حالت یا مولفه داریم. هنگامی که منحنی اعداد تکین یک سری زمانی به ترتیب صعودی به نزولی مرتب شود؛ یک نمودار از سمت چپ به راست با شیب زیاد به مناطق با شیب کم به دست می‌آید. سمت راست نمودار منحنی اعداد تکین یا مناطق با شیب زیاد، مولفه‌های اصلی یک سیگنال و مناطق با شیب کم، نویز در سیگنال را نشان می‌دهد. به عبارت دیگر سمت راست منحنی اعداد تکین تعداد مولفه‌های معنی‌دار سری زمانی را نمایش می‌دهد. محور عمودی شکل ۴ الف اثر واریانس هر کدام از

مولفه‌ها را نشان می‌دهد. با توجه به شکل ۴ ب، مولفه‌های اول تا نهم به طور معنی‌دار بیشترین میزان واریانس را تحت کنترل دارند (مولفه‌های ۱ تا ۵ در شکل ۴ الف با بزرگ‌نمایی مشخص شده است). مولفه‌های ۲ و ۳ یا ۴ و ۵ یا ۸ و ۹ با توجه به اینکه به صورت جفت در کنار هم قرار گرفته‌اند؛ توابع متعامد تجربی آنها در ربع فاز انتقال می‌یابد (اختلاف فاز) که این مولفه‌ها به طور موثر نشان دهنده مولفه‌های دوره‌ای می‌باشند (ووتارد و گیل، ۱۹۸۹). منحنی اعداد تکین نرمال شده در اندازه پنجره ۹۶ در شکل ۴ ب نشان داده شده است. با توجه به شکل ۴ ب، مولفه‌های ۱ در این سری زمانی تقریباً ۴۲/۵ درصد، مولفه ۲ و ۳ هر کدام ۲۴ درصد و مولفه ۴ و ۵ هر کدام تقریباً ۳/۵ درصد تغییرات را تحت کنترل دارند. جمعاً مولفه‌های ۱ تا ۵، ۹۷/۵ درصد تغییرات این سری زمانی را تحت کنترل دارد. به عبارتی عدم انتخاب هر کدام از این مولفه‌ها در بازسازی این سری زمانی باعث ایجاد یک سیگنال غیر قابل قبول یا ناقص خواهد شد.

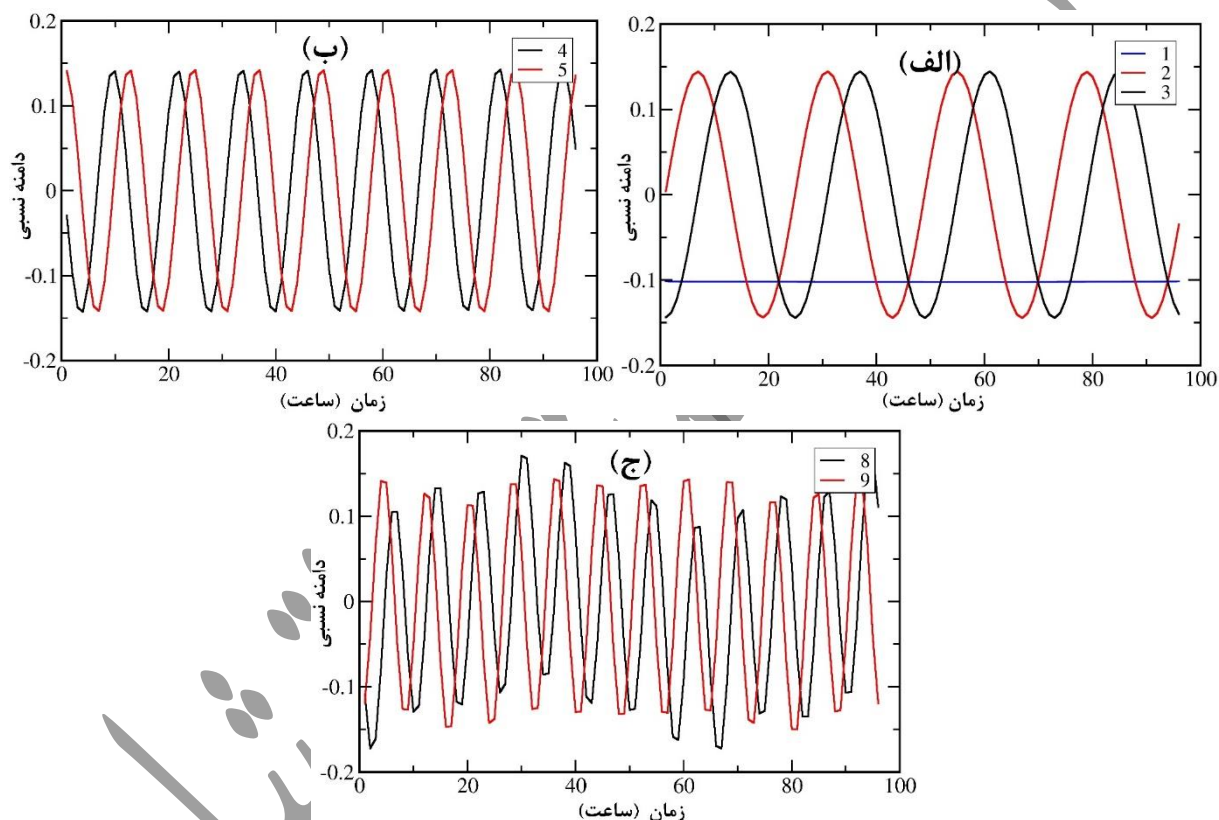


شکل ۴. منحنی اعداد تکین در اندازه پنجره ۹۶ ساعت (راست) و منحنی اعداد تکین نرمال شده (تغییرات واریانس نسبت به هر مولفه) در اندازه پنجره ۹۶ ساعت (چپ)

خروجی آزمون مونت کارلو SSA با داده‌های ساعتی LST در اندازه پنجره ۹۶ ساعت در شکل ۵ نشان داده شده است. شکل ۵ در واقع مقادیر اعداد ویژه را در مقابل فرکانس مربوطه نشان می‌دهد. با توجه به شکل نامبرده مولفه‌های ۱ تا ۵ مولفه‌های اصلی و معنی‌دار این سری زمانی می‌باشند (این مولفه‌ها در شکل با شماره مشخص شده‌اند). در شکل ۵ ترتیب نام گذاری مولفه‌ها بر اساس میزان واریانس بزرگتر می‌باشد. مولفه‌های ۲ با ۳، ۴ با ۵ و ۸ با ۹ در شکل ۵ بر روی یکدیگر قرار گرفته‌اند. فرکانس مربوط به مولفه‌های اول، دوم با سوم، چهارم با پنجم و هشتم با نهم به ترتیب ۰، ۰/۰۴۲، ۰/۰۸۳ و ۰/۱۲۵ سیکل بر تصویر می‌باشد. از آنجایی که معکوس فرکانس دوره را نشان می‌دهد. بنابراین به ترتیب مولفه اول نشان دهنده روند سالانه تغییرات دما (شکل ۶ الف، خط آبی)، مولفه دوم با سوم نشان دهنده تغییرات دوره‌ای ۲۴ ساعته یا روزانه دما (شکل ۶ الف)، مولفه چهارم با پنجم نشان دهنده تغییرات دوره‌ای ۱۲ ساعته دما (شکل ۶ ب) و مولفه هشتم با نهم نشان دهنده تغییرات دوره‌ای ۸ ساعته دما می‌باشد (شکل ۶ ج). در مجموع بر اساس نتایج این بخش از تعداد ۹ مولفه در اندازه پنجره ۹۶ ساعت برای بازسازی سری‌های زمانی ماهواره Meteosat-۹ استفاده شد.



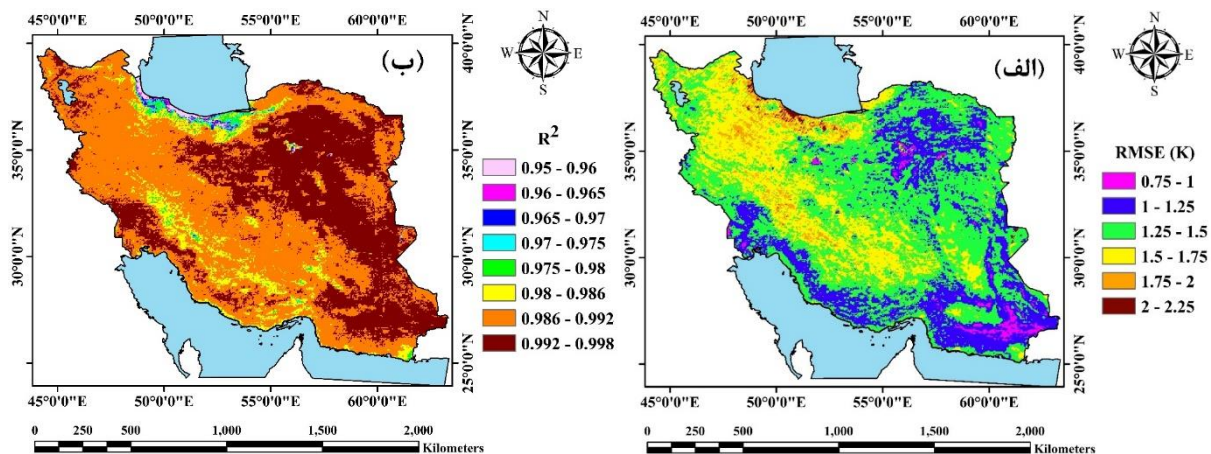
شکل ۵. آزمون مونت کارلو SSA با داده سالانه سری زمانی دمایی سطح زمین (اندازه پنجره ۹۶ ساعت)



شکل ۶. توابع متعامد تجربی زمانی برای مولفه‌های ۱، ۲ و ۳ (الف)، مولفه‌های ۴ و ۵ (ب) و مولفه‌های ۸ و ۹ (ج)

۳-۳- ارزیابی عملکرد M-SSA در شرایط آسمان صاف

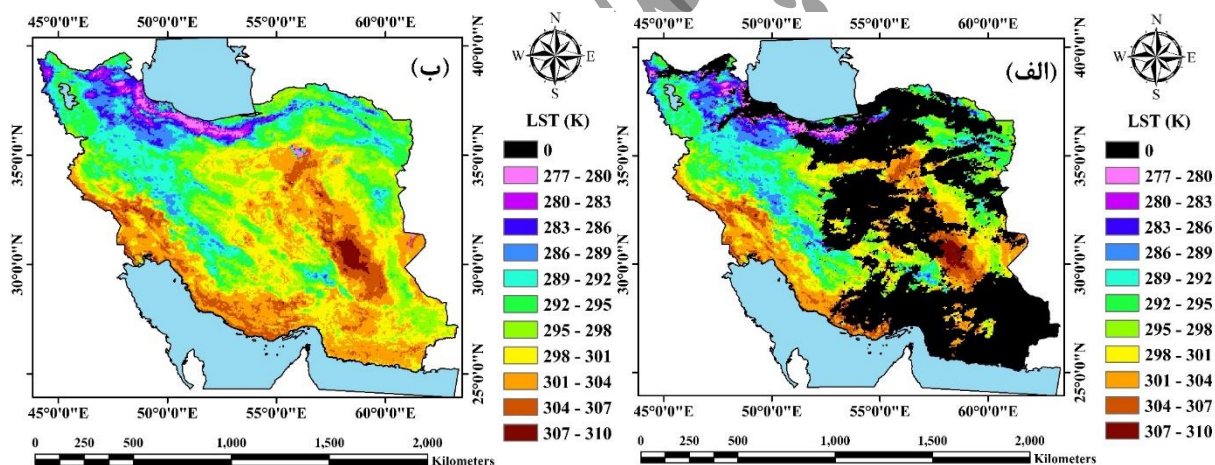
نقشه خطای RMSE و ضریب تبیین (R^2) بین داده‌های موجود سری زمانی LST ساعتی سالانه (داده‌های آسمان صاف) و داده‌های بازسازی شده توسط الگوریتم M-SSA در شکل ۷ نشان داده شده است. با توجه به شکل ۷ الف، حداقل خطای RMSE، ۰/۷۵ و حداکثر آن در نواحی غربی حاشیه دریای خزر با ۲/۲۵ کلون مشاهده شد. همچنین با توجه به نقشه ضریب تبیین، کمترین میزان همبستگی در سواحل دریای خزر (۰/۹۵) مشاهده شد. به طور میانگین در سطح کل ایران، خطای RMSE و ضریب تبیین بین داده‌های اصلی موجود و داده‌های بازسازی شده توسط الگوریتم M-SSA، به ترتیب ۱/۳۸ کلون و ۰/۹۹ بود.



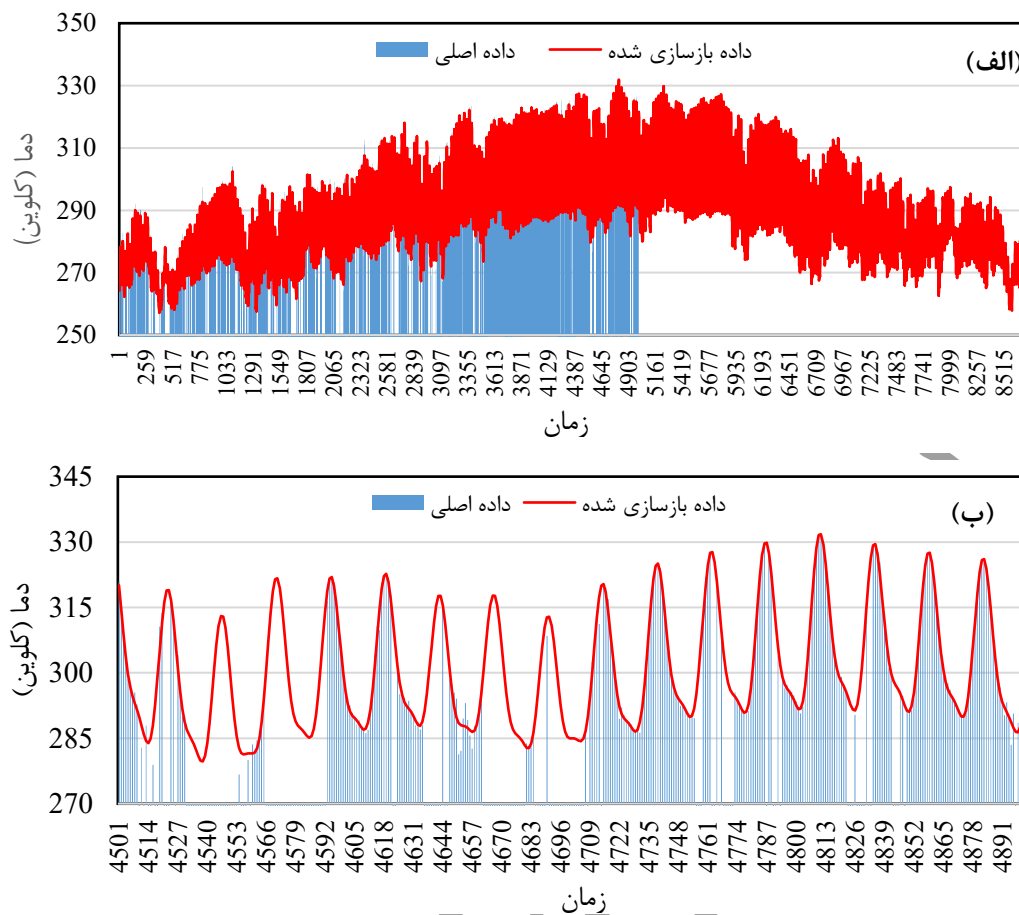
شکل ۷. نقشه خطای RMSE (الف) و ضریب تبیین (ب) بین داده های موجود سری زمانی LST ساعتی سالانه و داده های

بازسازی شده توسط الگوریتم M-SSA

یک نمونه تصویر LST قبل و بعد از بازسازی (تصویر تاریخ ۱۴۰۱/۰۴/۲۲ در ساعت ۲۱) در شکل ۸ نشان داده شده است. با توجه شکل ۸ مشاهده می شود نواحی با داده های از دست رفته توسط پوشش ابر (نواحی با رنگ سیاه) به خوبی توسط الگوریتم M-SSA بازسازی شده است. زیرا با توجه به شکل ۸ الگوها و طبقات تغییرات دما ثابت مانده و نواحی بدون داده پر شده است (نمایش مکانی). یک نمونه سری زمانی LST ساعتی سالانه (۸۷۶۰ داده) با سری زمانی بازسازی شده توسط الگوریتم M-SSA در شکل ۹ الف نشان داده شده است. با توجه به تراکم داده های زیاد، در شکل ۹ ب بخشی از این سری زمانی نمایش داده شده است. با توجه به شکل ۹ الگوریتم M-SSA به خوبی توانسته است نواحی با داده های از دست رفته را بازسازی کند (نمایش زمانی).



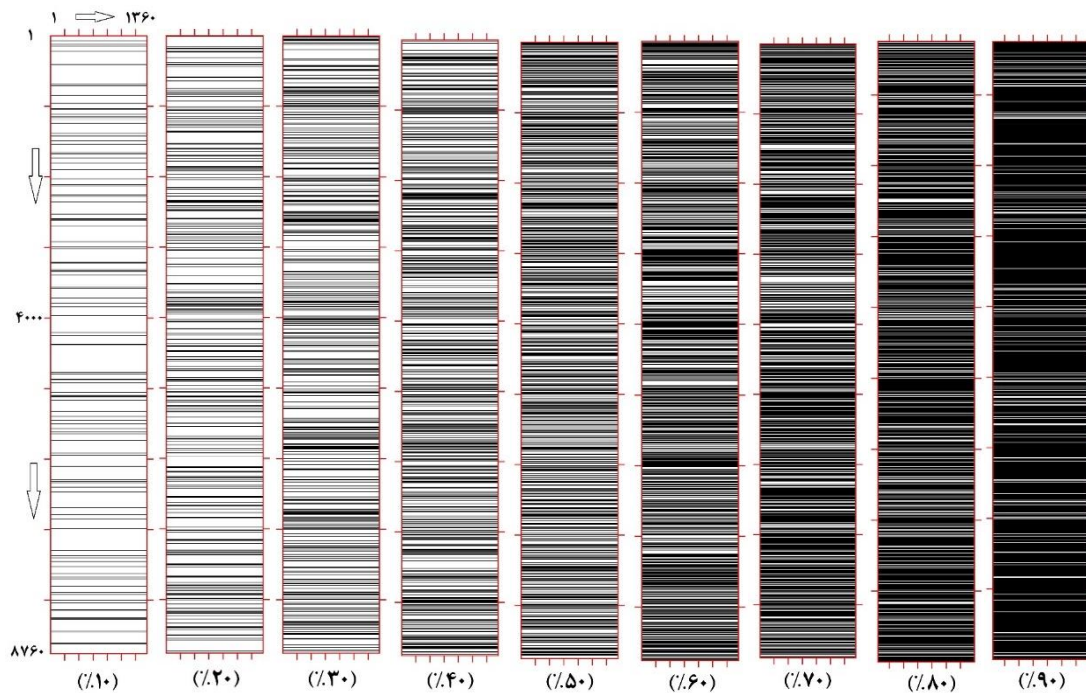
شکل ۸. یک نمونه تصویر با داده های از دست رفته (الف) و تصویر نظیر بازسازی توسط الگوریتم M-SSA (ب)



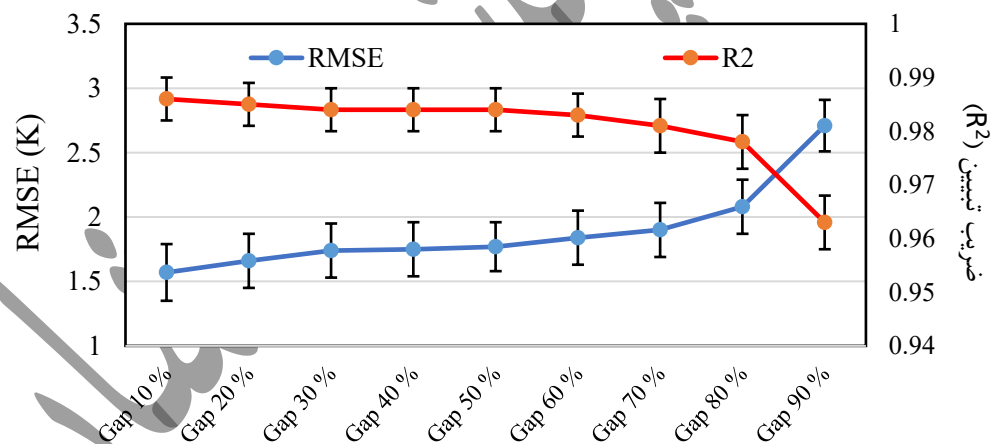
شکل ۹. یک نمونه سری زمانی LST ساعتی سالانه (میللهای آبی) با سری زمانی بازسازی شده (خط قرمز) توسط الگوریتم M-SSA (الف) و بزرگنمایی بخشی از سری زمانی (ب)

۳-۴- ارزیابی عملکرد M-SSA در شرایط آسمان ابری

همان‌طور که در بخش مواد و روش‌ها توضیح داده شده برای ارزیابی خطای پرکردن نقاط بدون داده در یک سری زمانی ماتریسی با ۱۳۶۰ سری زمانی LST ساعتی سالانه با طول ۸۷۶۰ داده، ابر مصنوعی یا گپ ایجاد شد. نقشه داده‌های از دست رفته مصنوعی در سری‌های زمانی ماتریسی با ۱۳۶۰ سری زمانی LST ساعتی در طول یکسال (۸۷۶۰ داده) در شکل ۱۰ نشان داده شده است. خطای RMSE و ضریب تبیین (R^2) در سری‌های زمانی LST ساعتی سالانه با شدت‌های مختلف داده‌های از دست رفته در شکل ۱۱ نشان داده شده است. با توجه به شکل ۱۱ مشاهده می‌شود با افزایش شدت داده‌های از دست رفته به صورت تصادفی، خطای RMSE افزایش و ضریب تبیین کاهش می‌یابد. اما این افزایش خطای RMSE تا سطح ۸۰ درصد داده از دست رفته در سری زمانی اندک بود. در سطح ۹۰ درصد داده از دست رفته در سری زمانی، خطای RMSE و ضریب تبیین بین داده‌های حذف شده به صورت مصنوعی و بازسازی شده توسط الگوریتم M-SSA به ترتیب، ۲/۷ کلون و ۰/۹۶ بود، که نسبت به سطح ۸۰ درصد داده از دست رفته، افزایش معنی‌دار یافت. با این حال این نتایج بسیار مطلوب و امیدوار کننده است زیرا الگوریتم M-SSA در یک سری زمانی با ۹۰ درصد داده از دست رفته به خوبی عمل کرد و خطای RMSE نسبت به سری‌های زمانی با سطح ۱۰ درصد داده از دست رفته فقط ۱/۱ کلون افزایش یافت.

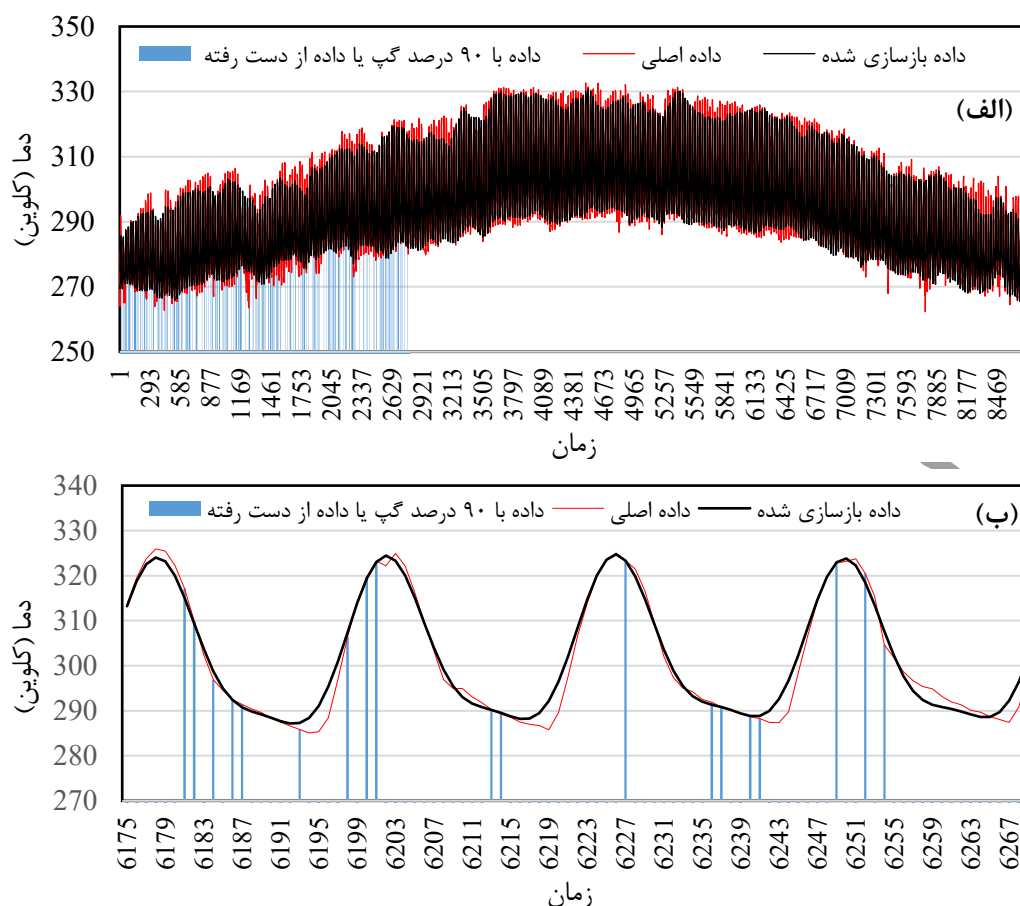


شکل ۱۰. نقشه داده‌های از دست رفته مصنوعی (گپ) در سری‌های زمانی ماتریسی با ۱۳۶۰ سری زمانی LST ساعتی در طول یکسال (۸۷۶۰ داده). از سمت چپ به راست داده‌های از دست رفته افزایش می‌یابد (خطوط سیاه داده‌های از دست رفته به صورت تصادفی)



شکل ۱۱. خطای RMSE و ضریب تبیین در سری‌های زمانی LST ساعتی سالانه با شدت‌های مختلف داده‌های از دست رفته

یک نمونه سری زمانی LST ساعتی سالانه با داده‌های اصلی و داده‌های از دست رفته به صورت مصنوعی به همراه داده‌های بازسازی شده در شکل ۱۲ الف نشان داده شده است. در شکل ۱۲ ب بخش کوچکی از این سری زمانی برای وضوح نشان داده شده است. با توجه به شکل ۱۲ ب داده‌های بازسازی شده در سری زمانی با ۹۰ درصد داده از دست رفته به داده‌های اصلی و اولیه بسیار نزدیک است.



شکل ۱۲. یک نمونه سری زمانی LST ساعتی سالانه (خطوط قرمز) به همراه داده‌های از دست رفته مصنوعی با شدت ۹۰ درصد (خطوط آبی) و سری زمانی بازسازی شده توسط الگوریتم M-SSA (الف) و بزرگ‌نمایی بخشی از سری زمانی (ب)

۵- بحث و نتیجه گیری

کلید اصلی بازسازی سری‌های زمانی با رفتار دوره‌ای، شناسایی مولفه‌های معنی‌دار دوره‌ای می‌باشد. در پژوهش حاضر با توجه به فرکانس ساعتی داده‌های دمای سطح زمین، سه مولفه دوره‌ای معنی‌دار ۲۴، ۱۲ و ۸ ساعته در سری‌های زمانی شناسایی شد. در بازسازی LST ساعتی ماهواره زمین ثابت ۲C-Fengyun نیز مولفه‌های دوره‌ای معنی‌دار ۸، ۱۲ و ۲۴ ساعته در اندازه پنجره ۷۲ ساعت شناسایی شد (Ghafari et al., ۲۰۱۲). در بازسازی اثر پوشش ابر تصاویر دمای سطح زمین سنجنده MODIS با توالی شب و روز سه مولفه معنی‌دار که مولفه اول نشان دهنده تغییرات دمای شبانه روز، مولفه دوم تغییرات دمای سالانه و مولفه سوم تغییرات دمایی فصلی بود، شناسایی شد (Ghafari et al., ۲۰۱۸). در بازسازی سری زمانی سالانه دمای سطح زمین ماهواره لندست ۸ با توالی ۱۶ روزه تنها دوره‌های معنی‌دار سالانه شناسایی شد (Zare Khormizi et al., ۲۰۲۳). به طور کلی تغییرات دوره‌ای و مولفه‌های دوره‌ای در سری‌های زمانی دمای سطح زمین ناشی از چرخش زمین به دور خود و خورشید و انحراف محور آن شکل می‌گیرند. بنابراین این مولفه‌ها عموماً برای بازسازی دمای سطح زمین در هر کجای کره زمین یکسان هستند.

بر اساس نتایج، خطای RMSE و ضریب تبیین (R^2) بین داده‌های اولیه موجود و داده‌های بازسازی شده توسط الگوریتم M-SSA در شرایط آسمان صاف به ترتیب ۱/۳۸ کلون و ۰/۹۹ بود. همچنین بر اساس نتایج بررسی عملکرد M-SSA در شرایط آسمان ابری خطای RMSE تا سطح ۸۰ درصد داده از دست رفته به صورت تصادفی همواره کمتر از ۲/۱ کلون بود. به

طور کلی، خطاهای بازسازی در شرایط آسمان صاف و ابری در محدوده قابل قبولی قرار گرفتند زیرا با خطاهای تخمین دمای سطح زمین با سنجش از دور ($1 \pm$ تا $3 \pm$ کلین) همسو می‌باشد (Li et al., ۲۰۱۳, Chakraborty et al., ۲۰۱۵). نتایج سایر پژوهش‌ها با نتایج پژوهش حاضر همسو می‌باشد. در پژوهشی خطای RMSE بین داده‌های غیر ابری و بازسازی شده در سری زمانی روزانه سنجنده MODIS با ۷۳۰ داده در سال ۲/۹۵ کلین بدست آمد (Ghafarian et al., ۲۰۱۸). در مطالعه‌ای دیگر نشان داده شد خطای میانگین مطلق (MAE) در بازسازی یک سری زمانی ساعتی با میانگین ۶۳ درصد داده از دست رفته توسط پوشش ابر، ۲/۲۵ کلین است (Ghafarian et al., ۲۰۱۲). خطای بازسازی داده از طریق روش درون‌یابی داده‌ها با توابع متعامد تجربی در تصاویر دمای سطح زمین سنجنده MODIS در زمان روز و شب به ترتیب ۴/۵ و ۲/۳ کلین بود (Zhou et al., ۲۰۱۷). تغییرات سری‌های زمانی دمای سطح زمین به شدت تحت تاثیر مولفه‌های دوره‌ای می‌باشد که این امر به بازسازی دقیق تصاویر دمای سطح زمین با خطای کم کمک شایانی می‌کند. با این حال دقت بازسازی سری‌های زمانی به سه فاکتور، اندازه بزرگترین فاصله داده‌های از دست رفته (بزرگ‌ترین شکاف)، میزان داده‌های از دست رفته و تعداد مکان‌های داده از دست رفته (تعداد شکاف‌ها) بستگی دارد (Zhou et al., ۲۰۱۲). همچنین اوج‌ها و پایه‌ها یا نقاط عطف تغییرات سری زمانی در دقت بازسازی بسیار تاثیر دارند (Zhou et al., ۲۰۱۲). با این حال الگوریتم M-SSA با توجه به در نظر گرفتن مولفه‌های دوره‌ای و روندها و همچنین استفاده از همبستگی زمانی و مکانی می‌تواند در بازسازی داده‌های از دست رفته با فاصله بزرگ موثر باشد. یکی از موارد قابل توجه در بازسازی اثر پوشش ابر در پژوهش حاضر و بسیاری از پژوهش‌های دیگر، بازسازی دمای سطح زمین با شرط آسمان صاف است. به عبارت دیگر بسیاری از روش‌های درون‌یابی زمانی، LST زیر پوشش ابر را بدست نمی‌آورند. با این حال با توجه به فرکانس ساعتی داده‌های سطح زمین در پژوهش حاضر الگوریتم SSA به طور دقیق مولفه‌های دوره‌ای و روندهای تغییرات را استخراج می‌کند و فرکانس بالای داده‌ها منجر به برازش دقیق سیگنال بر داده‌های موجود می‌شود. این امر باعث می‌شود دمای بازسازی شده به LST زیر پوشش ابر نزدیک باشد. به طور کلی بازسازی LST زیر پوشش ابر از پیچیدگی بیشتری برخوردار است (Sun et al., ۲۰۱۷). در برخی از پژوهش‌ها از معادلات رگرسیون چند متغیره و بیان انرژی سطح برای تخمین LST زیر پوشش ابر استفاده شده است (Lu et al., ۲۰۱۱, Sun et al., ۲۰۱۷, Zhao & Duan, ۲۰۲۰). ابرها یکی از مهم‌ترین عوامل ایجاد داده‌های گمشده (شکاف) و نامعتبر در سری‌های زمانی دمای سطح زمین هستند. به طور کلی در پژوهش حاضر شکاف‌های ایجاد شده توسط پوشش ابر به طور موثر و با خطای کم توسط الگوریتم M-SSA پر شدند. تخمین دقیق LST در زیر پوشش ابرها به عوامل زیادی از جمله ساعت ابرناکی (روز یا شب)، مدت زمان ابرناکی، فصل، توپوگرافی، رطوبت خاک، نوع پوشش زمین، خواص حرارتی خاک و اینرسی حرارتی خاک و ... بستگی دارد. بنابراین بازسازی دمای سطح زمین زیر پوشش ابر می‌تواند چالش و پیشنهادی برای مطالعات بیشتر در آینده باشد.

۶- منابع

1. Broomhead, D. S., & King, G. P., ۱۹۸۶, **Extracting qualitative dynamics from experimental data**. Physica D: Nonlinear Phenomena, ۲۰, PP. ۲۱۷-۲۳۶. [https://doi.org/10.1016/0167-2789\(86\)90031-X](https://doi.org/10.1016/0167-2789(86)90031-X)
2. Chakraborty, S. D., Kant, Y., & Mitra, D., ۲۰۱۵, **Assessment of land surface temperature and heat fluxes over Delhi using remote sensing data**. Journal of environmental management, ۱۴۸, PP. ۱۴۳-۱۵۲. <https://doi.org/10.1016/j.jenvman.2013.11.034>
3. Chen, D., Zhuang, Q., Zhu, L., & Zhang, W., ۲۰۲۲, **Comparison of methods for reconstructing MODIS land surface temperature under cloudy conditions**. Applied Sciences, ۱۲(۱۲), PP. ۶۰۶۸. <https://doi.org/10.3390/app12126068>

۴. Cui, J., Zhang, M., Song, D., Shan, X., & Wang, B., ۲۰۲۲, **MODIS Land Surface Temperature Product Reconstruction Based on the SSA-BiLSTM Model**. Remote Sensing, ۱۴(۴), PP. ۹۰۸. <https://doi.org/10.3390/rs14040908>
۵. Freitas, S. C., Trigo, I. F., Macedo, J., Barroso, C., Silva, R., & Perdigão, R., ۲۰۱۳, **Land surface temperature from multiple geostationary satellites**. International Journal of Remote Sensing, ۳۴(۹-۱۰), PP. ۳۰۵۱-۳۰۶۸. <https://doi.org/10.1080/01431161.2012.717920>
۶. Geng, L., Ma, M., Wang, X., Yu, W., Jia, S., & Wang, H., ۲۰۱۴, **Comparison of eight techniques for reconstructing multi-satellite sensor time-series NDVI data sets in the Heihe river basin, China**. Remote Sensing, ۶(۳), PP. ۲۰۲۴-۲۰۴۹. <https://doi.org/10.3390/rs6030249>
۷. Ghafarian Malamiri, H. R., & Zare Khormizie, H., ۲۰۱۷, **Reconstruction of cloud-free time series satellite observations of land surface temperature (LST) using harmonic analysis of time series algorithm (HANTS)**. Journal of RS and GIS for Natural Resources, ۸(۳), PP. ۳۷-۵۰. <http://dori.net/dor/20100112676708213968330>
۸. Ghafarian Malamiri, H. R., ۲۰۱۵, **Reconstruction of gap-free time series satellite observations of land surface temperature to model spectral soil thermal admittance** (Doctoral dissertation), Technische Universiteit Delft, The Netherlands. <https://doi.org/10.4233/uuid:63dc3402-9fd6-4094-a00e-7aa0ae2001aa>
۹. Ghafarian Malamiri, H. R., Rousta, I., Olafsson, H., Zare, H., & Zhang, H., ۲۰۱۸, **Gap-filling of MODIS time series land surface temperature (LST) products using singular spectrum analysis (SSA)**. Atmosphere, ۹(۹), PP. ۳۳۴. <https://doi.org/10.3390/atmos9090334>
۱۰. Ghafarian, H. R., Menenti, M., Jia, L. H., & den Ouden, R., ۲۰۱۲, **Reconstruction of cloud-free time series satellite observations of land surface temperature**. In, EARSel eProceedings (pp. ۱۲۱-۱۳۱)
۱۱. Golyandina, N., Nekrutkin, V., & Zhigljavsky, A., ۲۰۰۱, **Analysis of Time Series Structure: SSA and Related Techniques**. Washington DC, USA: CHAPMAN & HALL/CRC
۱۲. Helman, D., Lensky, I. M., Yakir, D., & Osem, Y., ۲۰۱۷, **Forests growing under dry conditions have higher hydrological resilience to drought than do more humid forests**. Global Change Biology, ۲۳(۷), PP. ۲۸۰۱-۲۸۱۷. <https://doi.org/10.1111/gcb.13001>
۱۳. Julien, Y., & Sobrino, J. A., ۲۰۱۰, **Comparison of cloud-reconstruction methods for time series of composite NDVI data**. Remote Sensing of Environment, ۱۱۴(۳), PP. ۶۱۸-۶۲۵. <https://doi.org/10.1016/j.rse.2009.11.001>
۱۴. Ke, L., Ding, X., & Song, C., ۲۰۱۳, **Reconstruction of time-series MODIS LST in Central Qinghai-Tibet Plateau using geostatistical approach**. IEEE Geoscience and Remote Sensing Letters, ۱۰(۶), PP. ۱۶۰۲-۱۶۰۶. <https://doi.org/10.1109/LGRS.2013.2263003>
۱۵. Kleipool, Q. L., Dobber, M. R., de Haan, J., & Levelt, P. F., ۲۰۰۸, **Earth surface reflectance climatology from ۳ years of OMI data**. Journal of Geophysical Research: Atmospheres, ۱۱۳(D۱۸). <https://doi.org/10.1029/2008JD010290>
۱۶. Kondrashov, D., & Ghil, M., ۲۰۰۶, **Spatio-temporal filling of missing points in geophysical data sets. Nonlinear Processes in Geophysics**, ۱۲(۲), PP. ۱۵۱-۱۵۹. <https://doi.org/10.5194/npg-12-151-2006>
۱۷. Kondrashov, D., Shprits, Y., & Ghil, M., ۲۰۱۰, **Gap filling of solar wind data by singular spectrum analysis**. Geophysical research letters, ۳۷(۱۵). <https://doi.org/10.1029/2010GL044138>
۱۸. Li, J., Carlson, B. E., & Lacis, A. A., ۲۰۱۳, **Application of spectral analysis techniques in the intercomparison of aerosol data: ۱. An EOF approach to analyze the spatial-temporal**

- variability of aerosol optical depth using multiple remote sensing data sets.** *Journal of Geophysical Research: Atmospheres*, 118(10), PP. 8640–8648. <https://doi.org/10.1029/jgrd.201606>
19. Lu, L., Venus, V., Skidmore, A., Wang, T., & Luo, G., 2011, **Estimating land-surface temperature under clouds using MSG/SEVIRI observations.** *International Journal of Applied Earth Observation and Geoinformation*, 13(2), PP. 260–276. <https://doi.org/10.1016/j.jag.2010.12.007>
 20. Martins, J. P., Coelho e Freitas, S., Trigo, I. F., Barroso, C., & Macedo, J., 2023, **Copernicus Global Land Operations-Lot I “Vegetation and Energy” Algorithm Theoretical Basis Document.** Land Surface Temperature—LST, 1. V2.1. ISSUE 12,23.
 21. Mukherjee, S., Joshi, P. K., & Garg, R. D., 2014, **A comparison of different regression models for downscaling Landsat and MODIS land surface temperature images over heterogeneous landscape.** *Advances in Space Research*, 54(4), PP. 600–619. <https://doi.org/10.1016/j.asr.2014.04.013>
 22. Musial, J. P., Verstraete, M. M., & Gobron, N., 2011, **Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series.** *Atmospheric chemistry and physics*, 11(10), PP. 7905–7923. <https://doi.org/10.5194/acp-11-7905-2011>
 23. Pede, T., & Mountrakis, G., 2018, **An empirical comparison of interpolation methods for MODIS 8-day land surface temperature composites across the conterminous United States.** *ISPRS journal of photogrammetry and remote sensing*, 142, PP. 137–150. <https://doi.org/10.1016/j.isprsjprs.2018.06.003>
 24. Shiff, S., Helman, D., & Lensky, I. M., 2021, **Worldwide continuous gap-filled MODIS land surface temperature dataset.** *Scientific Data*, 8(1), PP. 74. <https://doi.org/10.1038/s41597-021-00861-7>
 25. Sun, L., Chen, Z., Gao, F., Anderson, M., Song, L., Wang, L., ... & Yang, Y., 2017, **Reconstructing daily clear-sky land surface temperature for cloudy regions from MODIS data.** *Computers & Geosciences*, 100, PP. 1–20. <https://doi.org/10.1016/j.cageo.2017.04.007>
 26. Vautard, R., & Ghil, M., 1999, **Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series.** *Physica D: Nonlinear Phenomena*, 130, PP. 390–424. [https://doi.org/10.1016/S0167-2789\(99\)00077-8](https://doi.org/10.1016/S0167-2789(99)00077-8)
 27. Vautard, R., Yiou, P., & Ghil, M., 1992, **Singular-spectrum analysis: A toolkit for short, noisy chaotic signals.** *Physica D: Nonlinear Phenomena*, 58(1–2), PP. 95–126. [https://doi.org/10.1016/0167-2789\(92\)90103-T](https://doi.org/10.1016/0167-2789(92)90103-T)
 28. Verhoef, W., Menenti, M., & Azzali, S., 1996, **Cover A colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981–1992).** *International Journal of Remote Sensing*, 17(2), PP. 231–239. <https://doi.org/10.1080/01431169608949001>
 29. Wang, H., Mao, K., Yuan, Z., Shi, J., Cao, M., Qin, Z., ... & Tang, B., 2021, **A method for land surface temperature retrieval based on model-data-knowledge-driven and deep learning.** *Remote Sensing of Environment*, 260, PP. 112660. <https://doi.org/10.1016/j.rse.2021.112660>
 30. Xu, Y., & Shen, Y., 2013, **Reconstruction of the land surface temperature time series using harmonic analysis.** *Computers & geosciences*, 61, PP. 126–132. <https://doi.org/10.1016/j.cageo.2013.08.009>
 31. Yu, P., Zhao, T., Shi, J., Ran, Y., Jia, L., Ji, D., & Xue, H., 2022, **Global spatiotemporally continuous MODIS land surface temperature dataset.** *Scientific Data*, 9(1), PP. 143. <https://doi.org/10.1038/s41597-022-01214-8>

۳۲. Zare Khormizi, H., Ghafarian Malamiri, H. R., & Alian, S, ۲۰۲۳, **Calculation of Land Surface Temperature Using a Generalized Split-Window Algorithm and the Reconstruction of its Lost Data by Cloud Cover Through a Singular Spectral Analysis (SAA)-algorithm**. Desert, ۲۸(۱), PP. ۲۷-۴۸. <https://doi.org/10.22059/jdesert.2023.93537>
۳۳. Zare Khormizi, H., Tavili, A., & Ghafarian Malamiri, H. R., ۲۰۲۱, **Estimation of Actual Evapotranspiration Using SEBAL Algorithm and Comparison with Improved FAO ۵۶ Standard Evapotranspiration with KC-NDVI Relationship**. Iranian Journal of Remote Sensing & GIS, ۱۳(۳), PP. ۷۳-۹۲. <https://doi.org/10.52547/gisj.13.3.73>
۳۴. Zhang, C., Li, W., & Travis, D. J., ۲۰۰۹, **Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging**. International Journal of Remote Sensing, ۳۰(۹), ۲۱۷۳-۲۱۹۵. <https://doi.org/10.1080/01431160802549294>
۳۵. Zhao, W., & Duan, S. B., ۲۰۲۰, **Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data**. Remote Sensing of Environment, ۲۴۷, PP. ۱۱۱۹۳۱. <https://doi.org/10.1016/j.rse.2020.111931>
۳۶. Zhou, J., Jia, L., & Menenti, M., ۲۰۱۵, **Reconstruction of global MODIS NDVI time series: Performance of Harmonic ANalysis of Time Series (HANTS)**. Remote Sensing of Environment, ۱۶۳, PP. ۲۱۷-۲۲۸. <https://doi.org/10.1016/j.rse.2015.03.018>
۳۷. Zhou, J., Jia, L., Hu, G., & Menenti, M., ۲۰۱۲, **Evaluation of Harmonic Analysis of Time Series (HANTS): Impact of gaps on time series reconstruction**. In ۲۰۱۲ Second International Workshop on Earth Observation and Remote Sensing Applications (pp. ۳۱-۳۵). IEEE. <https://doi.org/10.1109/EORSA.2012.6261129>
۳۸. Zhou, W., Peng, B., & Shi, J., ۲۰۱۷, **Reconstructing spatial-temporal continuous MODIS land surface temperature using the DINEOF method**. Journal of Applied Remote Sensing, ۱۱(۴), PP. ۰۴۶۰۱۶-۰۴۶۰۱۶. <https://doi.org/10.1117/1.JRS.11.046016>

ارزیابی عملکرد الگوریتم تجزیه تحلیل طیفی منفرد در بازسازی داده‌های از دست رفته با شدت‌های مختلف در
سری زمانی ساعتی دمای سطح زمین

چکیده

تولید داده‌های دمای سطح زمین (LST) با پیوستگی زمانی و مکانی برای مطالعات هیدرولوژی، هواشناسی، اکولوژی، زیست محیطی و غیره تقاضای زیادی دارد. با این حال ابرها با تاثیر بر باندهای مادون قرمز حرارتی باعث ایجاد داده‌های از دست رفته و نامعتبر در سری‌های زمانی LST می‌شوند. در پژوهش حاضر برای پر کردن و بازسازی فاصله‌های داده از دست رفته در سری زمانی سالانه LST ساعتی ماهواره ۹-Meteosat در سطح کل ایران از الگوریتم تجزیه تحلیل طیفی منفرد-چند کاناله (M-SSA) استفاده شد. خطای بازسازی در شرایط آسمان صاف با داده‌های موجود سری زمانی و خطای بازسازی در شرایط آسمان ابری با ایجاد داده‌های از دست رفته مصنوعی (ابر مصنوعی) با شدت‌های ۱۰، ۲۰، ۳۰، ...، ۹۰ درصد در سری زمانی اندازه‌گیری شد. بر اساس نتایج، خطای جذر میانگین مربعات (RMSE) و ضریب تبیین (R^2) بین داده‌های موجود و داده‌های بازسازی

شده توسط الگوریتم M-SSA در شرایط آسمان صاف به ترتیب ۱/۳۸ کلوین و ۰/۹۹ بود. همچنین بر اساس نتایج، در شرایط آسمان ابری خطای RMSE تا سطح ۸۰ درصد داده از دست رفته به صورت تصادفی همواره کمتر از ۲/۱ کلوین بود. الگوریتم M-SSA با استفاده از همبستگی زمانی و مکانی می‌تواند در بازسازی داده‌های از دست رفته با فاصله بزرگ در سری‌های زمانی LST موثر باشد.

کلمات کلیدی: پوشش ابر، تجزیه تحلیل طیفی منفرد، توابع متعامد تجربی، درون‌یابی زمانی- مکانی، سری زمانی

Evaluation of the performance of the Singular Spectrum Analysis (SSA) algorithm in reconstructing missing data with different intensities in the hourly Land Surface Temperature time series

Abstract

Generating land surface temperature (LST) data with continuous temporal and spatial continuity is in great demand for hydrology, meteorology, ecology, environment, etc. studies. However, clouds affect thermal infrared bands and cause missing and invalid data in the LST time series. In the present study, the Multi-channel Singular Spectrum Analysis (M-SSA) was used to fill and reconstruct the missing data (gaps) in the annual time series of hourly LST of Meteosat-9 satellite in the whole of Iran. Reconstruction errors in clear sky conditions with available time series data and reconstruction error in cloudy sky conditions were examined by creating artificial missing data (artificial cloud) with intensities of ۱۰, ۲۰, ۳۰, ..., ۹۰% in the time series. Based on the results, the Root Mean Square Error (RMSE) and the Coefficient of determination (R^2) between the original and the reconstructed data in clear sky conditions were ۱,۳۸ and ۰,۹۹ Kelvin, respectively. Also, based on the results, in cloudy sky conditions, the RMSE error up to the level of ۸۰% of randomly lost data was always less than ۲,۱ Kelvin. Therefore, as, M-SSA algorithm uses both temporal and spatial correlation in reconstructing lost data with large continuous gaps in LST time series, it is an effective algorithm for gap-filling time series satellite observations.

keywords: Cloud cover, Empirical Orthogonal Functions, Singular Spectrum Analysis, Temporal-spatial interpolation, Time series