Article

Reconstruction of the Cloud-Free Time Series Satellite Observations of Land Surface Temperature (LST) Using Singular Spectrum Analysis (SSA)

Hamid Reza Ghafarian Malamiri[1], Iman Rousta[1, 2,*], Haraldur Olafsson[3], Hadi Zare[4], Hao Zhang[5,*]

[1] Department of Geography, Yazd University, Yazd 8915818411, Iran, hrghafarian@yazd.ac.ir (H.G.M) & irousta@yazd.ac.ir (I.R.)

[2] Senior Researcher, Institute for Meteorological Research, Bustadavegur 9, Reykjavik, Iceland.

[3] Department of Physics, University of Iceland, Icelandic Meteorological Office and Institute for Meteorological Research, Bustadavegur 9, IS-150 Reykjavik, Iceland; haraldur@vedur.is

[4] MSc of Range Management, College of Natural resources and Desert, Yazd University, Yazd 8915818411, Iran, Email: hadi.zarekh@gmail.com

[5] Department of Environmental Science and Engineering Jiangwan campus, Fudan University, 2005 Songhu Road,Yangpu District Shanghai 200438,China, zhokzhok@163.com

* Correspondences: irousta@yazd.ac.ir (I.R); & zhokzhok@163.com Tel.: 00989171902098 (I.R)

Abstract: Land Surface Temperature (LST) is a basic parameter in energy exchange between the land and atmosphere and is frequently used in many sciences such as climatology, hydrology, agriculture, ecology, etc. LST time series data have usually deficient, missing and unacceptable data caused by the presence of clouds in images, presence of dust in atmosphere and sensor failure. In this study, Singular Spectrum Analysis (SSA) algorithm was used to resolve the problem of missing and outlier data caused by cloud cover. The region studied in the present research included an image frame of MODIS with horizontal number 22 and vertical number 05 (h22v05). This image involved a large part of Iran and Turkmenistan and Caspian Sea. In this study, MODIS LST sensor (MOD11A1) was used during 2015 with 1×1 Km spatial resolution and day/night LST data (daily temporal resolution). The results of the data quality showed that cloud cover caused 36.37% of missing data in the studied time series with 730 day/night LST images. Further, the results of SSA algorithm in reconstruction of LST images indicated the Root Mean Square Error (RMSE) of 2.95 K between the original and reconstructed data in LST time series in the study region. In general, the findings showed that SSA algorithm using spatio-temporal interpolation in LST time series can be effectively used to resolve the problem of missing data caused by cloud cover.

Keywords: Gap filling, M-SSA, Monte Carlo test, Time series, LST

1. Introduction:

Land Surface Temperature (LST) is a principal characteristic of land surface for evaluation of energy exchange between the land and atmosphere [1-3]. This parameter is applicable to numerous sciences such as climatology, hydrology, agriculture, ecology, public health and environment [4-12]. One of the methods for measuring LST is to use the satellite data and thermal remote sensing techniques. However, satellite observations of LST time series often have missing data and outliers.

Missing values and/or outliers can be caused by cloud cover, sensor failure, aerosol, and inefficiency of cloud masking algorithms [1,2,13,14]. Atmospheric dust, aerosol, gases and clouds, in particular, can dramatically affect the energy reflected from the surface and expose the reading of optical and thermal sensors to error [13].   Clouds in satellite images are identified as phenomena that have higher reflection in visible bands of electromagnetic spectrum than other most terrestrial phenomena and have lower temperature than the below land surface in the thermal bands [15]. In thermal remote sensing, clouds absorb part of the energy reflected from the earth. They also emit thermal infrared energy at much lower temperature so that they are not correctly diagnosed in cloud masking algorithm, and some negative outliers (lower temperature than other measurements in a time series) will remain in place of missing data in a time series [16].

The missing data in a time series, depending on size dispersion, temporal distribution and their continuity over time, can be very short, dispersed, or have a long gap. According to the missing data in time series, two terms can be defined: missing data with short gap and missing data with long gap. When there are less than 50% of sampled data over time, they are faced with the missing data with short gap/size in time series. However, if more than half of the sampled data are lost over time, this time series will have a large gap [17,18]. The outliers are defined as abnormal values whose deviation from normal changes in dataset is larger [16]. The outliers can be classified into positive and negative e data. For example, values larger than 1 or smaller than -1 in a Normalized Difference Vegetation Index (NDVI) time series. Further, sharp increase or decrease of NDVI data in a time series of NDVI, compared to the data before and after it, can be identified as outliers. Time series can be periodic or non-periodic. Remote sensing LST time series are included in periodic time series because they reflect the seasonal and annual variations of the sun.

Several methods have been developed to identify cloud cover, which can be divided into two groups [13]. The first group, which is a spectrum-based approach, uses all spectral data to determine whether a specific pixel includes cloud. The second group, which is a temporal-based approach, presents an evaluation of the missing values caused by cloud cover through temporal interpolation. For example, Fast Fourier Transform (FFT) algorithm and Harmonic ANalysis of Time Series (HANTS) algorithm used to reconstruct the missing data and remove the outliers in the time series of NDVI are based on the second approach [14,18-22]. HANTS algorithm was first developed to reconstruct the NDVI time series and to extract phonological data. But since this algorithm reconstructs the signal of a time series based on periodic behavior, it was effectively used to reconstruct other time series with periodic behavior like LST [2,16,23,24].   A disadvantage of HANTS is that in these methods only temporal correlation between observations is used to fill the gap of missing data in a time series.

Singular Spectrum Analysis (SSA) is an advanced method to reconstruct the time series with missing data, especially with a large gap, that uses SSA (with only one time series) and Multi-channel Singular Spectrum Analysis (M-SSA for more than a time series). SSA was first developed by Broomhead & King (1986). It was then proposed by Vautard & Ghil (1992) and Kondrashov & Ghil (2006) to fill the gaps of missing data in a time series [25-27]. The value of each pixel in an LST image over time is the result of interaction between radiation energy and ambient turbulent energy. Therefore, this time series has regular (cycle) and irregular (noise) components. Using this

presumption, SSA makes use of Empirical Orthogonal Functions (EOFs) and analysis of major components in temporal domain to extract data from short and noisy time series, without primary knowledge about dynamic processes that affect time series [27]. SSA is an advanced method for analysis of time series that makes use of multivariate principles of statistics and geometry, linear algebra and signal processing [28]. SSA analyses the time series to simpler and interpretable components such as trends, periodic and semi-periodic fluctuations and noises [29]. Thus, several significant components achieve maximum variance and noises achieve minimum variance in observations of a time series.

SSA has been used as a standard tool in analysis of climatic time series, meteorology and geophysics [30-34]. In a research, the hourly LST time series data were reconstructed by SSA and M-SSA. The results showed a mean absolute error of 2.25° K in an hourly time series with 63% missing data caused by cloud cover [1]. In another study, SSA algorithm was used to reconstruct the missing data in Leaf Area Index (LAI) of MODIS data. The findings indicated M-SSA effectively reconstructed this index using spatio-temporal correlation [35]. Moreover, another study showed that SSA could be effective in analysis of hydrologic time series, including reconstruction of time series, extraction of trends, fluctuations and prediction [36].

The current study was aimed to assess the performance of SSA and M-SSA algorithms in removing the clouds and outliers in daily MODIS LST time series. In most applied studies, satellite imageries of LST covered by cloud, has been a limitation in choosing the images. Therefore, the results of this study can be used in most applied studies using LST images according to the pure data generated by SSA algorithm.

2. Data and Methods

2.1. Data

The study region in this research included an image frame in MODIS sensor with horizontal number 22 and vertical number 5 (h22v5) in MODIS sinusoidal tiling system. This image involved a large part of Iran and Turkmenistan and Caspian Sea. Fig. 1 displays an LST image of the studied region by the Universal Transverse Mercator (UTM) coordinate system.
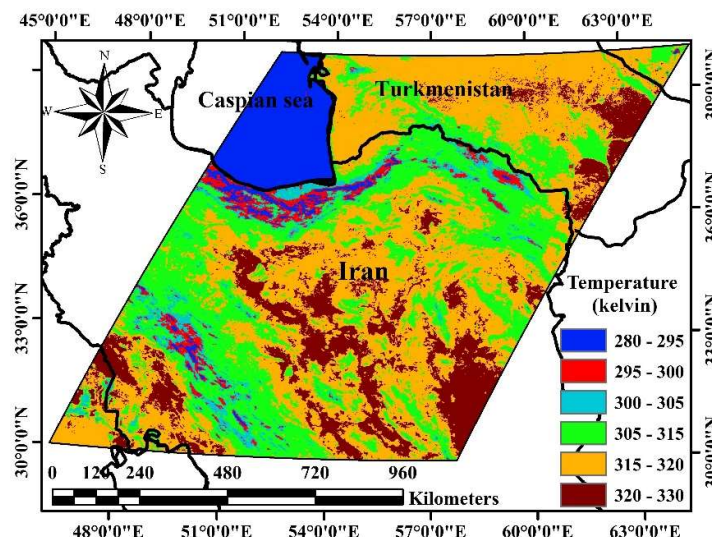
Figure 1. An LST image of the studied region.

2.1.1.   Satellite images and time series studies

In this study, MODIS LST product (MODIS11A1) was used during 2015. This product has a spatial resolution of one kilometer as well as daily temporal resolution and contains day/night LST data. MODIS LST products are obtained from two thermal infrared bands of channels 31 (wavelength range of 10.78-11.28 μm) and 32 (wavelength range of 11.77-12.27 μm) using split-window algorithm 2 [5,37]. Annual LST time series (2015) in this study included 730 images (365 LST images during the day and 365 images at night). The LST images in the present study were considered as day/night sequence during the year. Hence, in this time series with 730 images, the odd numbers showed the LST images during the day and even numbers indicated the LST images at night. Images 1 and 2 are for early January and images 729 and 730 are for late December. First, a time series with 730 LST images along with cloud observations is shown in Fig. 2 A. Fig. 2 B depicts the LST changes along a pixel in this time series. For a sharper display in Fig. 2 C, a part of this time series is magnified.
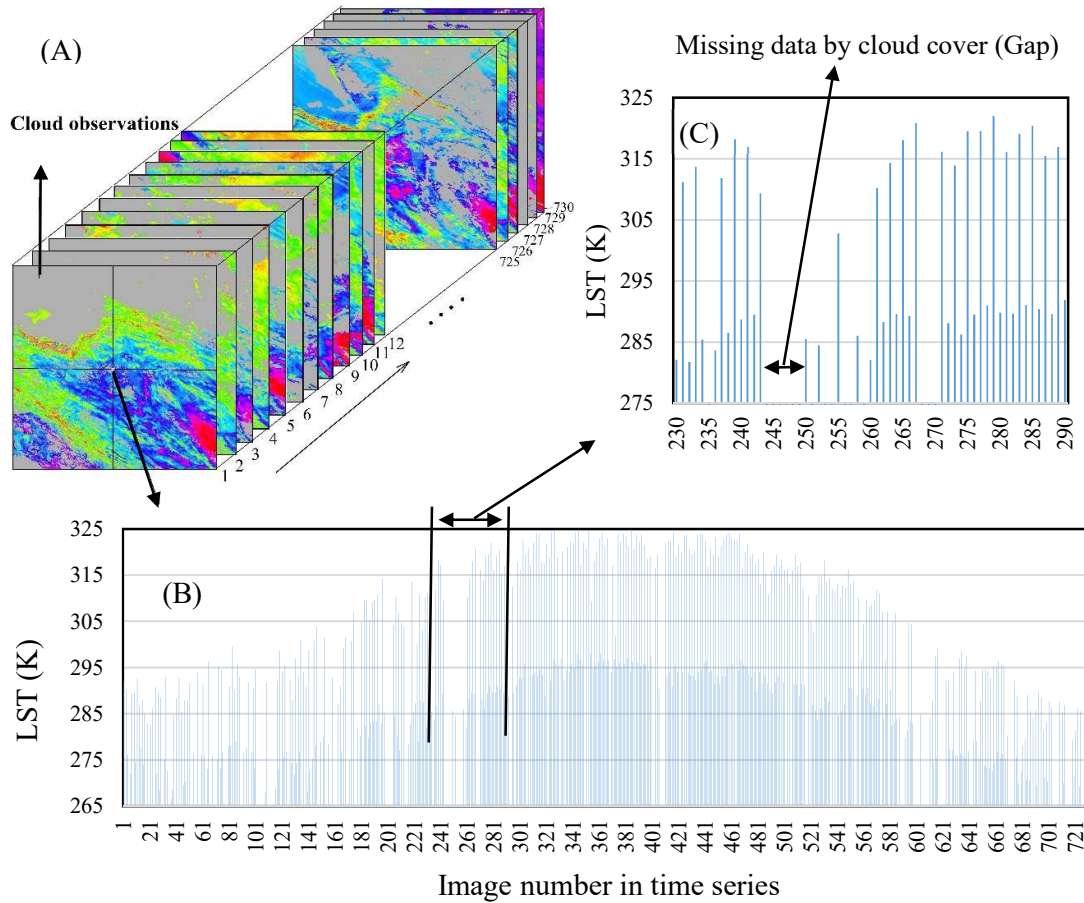
Figure 2. Time series with 730 images (A), trend of temperature changes along a pixel (B) and magnification of a part of time series (C)

2.2. Methods

2.2.1. Singular Spectrum Analysis (SSA)

The whole work flow of SSA gap-filling and smoothing algorithm is illustrated as follows based on [1,38]:

Step1: A single scalar time series x(t); t =1, 2, 3, … , n is embedded into a multidimensional trajectory matrix of lagged vectors

$$x = \left[x_1, ..., x_{n'} \right] \tag{1}$$

Where n'= n-m+1 and each lagged vector is defined as:

$$x_j = \left(x_j, ..., x_{j+m-1} \right)^T \tag{2}$$

This trajectory matrix contains the complete record of patterns presented within a window of size m. By selecting a large number of window size, more information about the basic pattern of the time series will be captured. A small value of window size enhances the statistical confidence at the final

results [39], since the structure of time series will be captured repeatedly [40]. The final form of the trajectory matrix X is a rectangular matrix of the form:

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_k \\ x_2 & x_3 & x_4 & \dots & x_{k+1} \\ x_3 & x_4 & x_5 & \dots & x_{k+2} \\ \dots & \dots & \dots & \dots & \dots \\ x_m & x_{m+1} & x_{m+2} & \dots & x_n \end{pmatrix}$$

(3)

Step2: The next step is decomposition of trajectory matrix X of size m×n' using the Singular Value Decomposition (SVD) method which yields:

$$X = DLE^T$$

(4)

Where D and E are left and right singular vectors of X with m×m and n×n size, respectively, and L is a rectangular diagonal matrix of size m×n. The elements of L, called singular values, are the square roots of the eigenvalues of the lagged-covariance matrix S = XXT of size m×m. The columns of matrix D are the eigenvectors of S also known as Empirical Orthogonal Functions (EOFs). The rows of ET are eigenvectors of matrix XTX. If we plot the singular values in descending order, one can often distinguish between an initial steep slope, representing a signal, and a (more or less) flat floor, representing the noise level [27]. Then any subset of d eigenvalues (EOFs), 1≤d≤ m, for which related eigenvalues are positive provides the best representation of the matrix X as a sum of matrices Xi, i = 1, 2, …, d [29].

Step 3: Partitioning d eigentriples into p distinct subsets. Then, summing all the components inside each subset such that;

$$X = \sum_{n'=1}^{p} X_{I_{n'}} \ \ where, \ \ X_{I_{n'}} = \sum_{i \varepsilon I_{n'}} X_i$$

(5)

The matrices XIn' have the form of a Hankel matrix in an ideal case and consequently fit the trajectory matrices.

Step 4: Since the ideal case described in step 3 is not usually the case, the XIn' matrices should be transformed into the form of a Hankel matrix to fit the trajectory matrices. This step is known as diagonal averaging. In this sense, the original matrix can be reconstructed as the sum of these matrices.

$$x_t = \sum_{n'=1}^{p} x_t^{n'}, \ \ t = 0, \ 1, \ \dots, \ n-1$$

(6)

Where for each p, the series xtn' is the result of the diagonal averaging of the matrix XIn'.

The SSA workflow for gap filling procedure consists of several steps which are explained as follows:

For a given window width (m), the original time series is centered by computing the unbiased value of the mean, and the missing data is set to zero. The first leading EOF is found by an iterative procedure which applies the SSA algorithm on the zeroed and centered set. The missing values are updated based on the reconstructed component of the current EOF. Using this updated set the SSA algorithm is applied again, and the missing values are updated based on the result of this new iteration. The process repeats until convergence is achieved. Then the iteration starts for the second leading EOF (keeping the first one fixed) until convergence has been achieved for the second EOF. This process repeats for the selected number of EOFs, each time keeping the previous ones as fixed. To find the optimal value for the window width and the number of dominant SSA modes to fill the gaps, cross-validation is applied. It means that a portion of the available data (selected as random) is flagged as missing, and the RMSE error from the reconstruction is computed to find the best value for the window size and number of EOFs. The SSA technique can be generalized to be used for multivariate time series and gap-filling of missing values in those time series [41].

In this research, the spatio-temporal dispersion of the missing data in time series is shown. The spatial dispersion is the same missing data caused by cloud cover in each time series image, and temporal distribution is the missing data along 730 images over time. To display the spatial dispersion of the missing data, the image statistics in ENVI 4.8 software were used. Also, to show the temporal distribution of the missing data, the missing data caused by cloud cover along each pixel of this time series were counted and then divided by the total number of data over time. Hence, the map of missing data in this time series was prepared.

SSA-MTM software was sued to reconstruct the LST time series with 730 images during 2015. This software is accessible via http://research.atmos.ucla.edu/tcd//ssa/form.html for free. To reconstruct the time series in SSA software, it is necessary to determine the number of significant components and window size. To determine these parameters, first, a single time series without missing data was selected as a representative of the whole time series. Then, the effects of window size and number of components in reconstruction of this time series were assessed by SSA. Next, the significant components were determined by different tests in SSA software such as Monte Carlo test. This will be explained in details in the results section. Having determined the window size and significant components, the time series of all images were reconstructed by M-SSA using these results. It should be pointed out that due to the lengthy discussion of SSA software tests such as Monte Carlo test as well as testing the null hypothesis of EOFs and theoretical discussion on the trends and periods, more information can be found in the study of Ghafarian Malamiri (2015) [16].

To evaluate the M-SSA performance in reconstructing the images of this time series, Root Mean Square Error (RMSE) was sued. RMSE is obtained from equation 7. In this equation, xi and yi represent original and reconstructed data, respectively.

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

(7)

To calculate RMSE, first the cloud data, the data with zero value, were removed. Then, RMSE was computed between the true data and reconstructed data by SSA algorithm along each pixel of this time series and displayed as an RMSE map.

3. Results

3.1. Spatio-temporal distribution of the missing data

To analyze the performance of SSA algorithm in reconstructing LST images, the distribution style and amount of missing data in LST time series were calculated temporally and spatially. For a better display, the time series with 730 images were depicted in LST time series during the day and night. The percentages of missing data in each LST time series image (spatial dispersion) are shown during the day and at night in figures 3 and 4. In horizontal axis (figure 3), number one is related to early January 2015. Based on figures 3 and 4, the amount of missing data is higher at the start and end of time series. The start and end of times series, depending on the winter and fall seasons, have the highest rate of missing data, which is possibly due to higher cloudiness in these seasons.
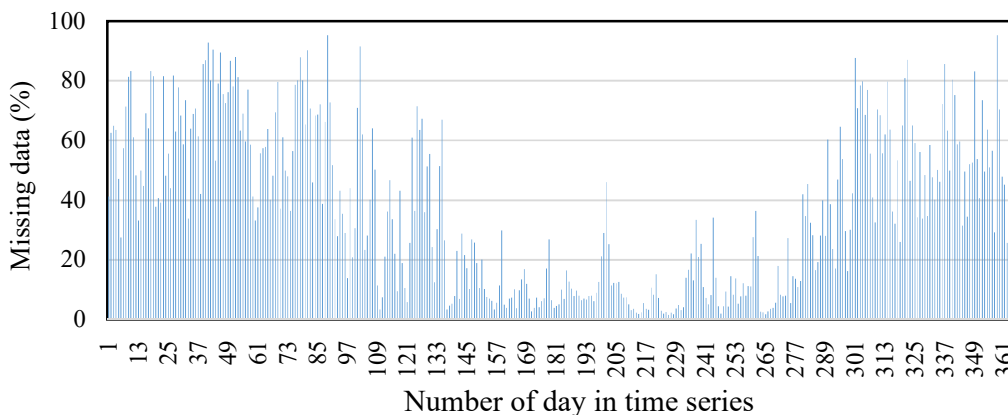

Figure 3. The missing data of each image (Spatial dispersion) in annual LST time series during the day.
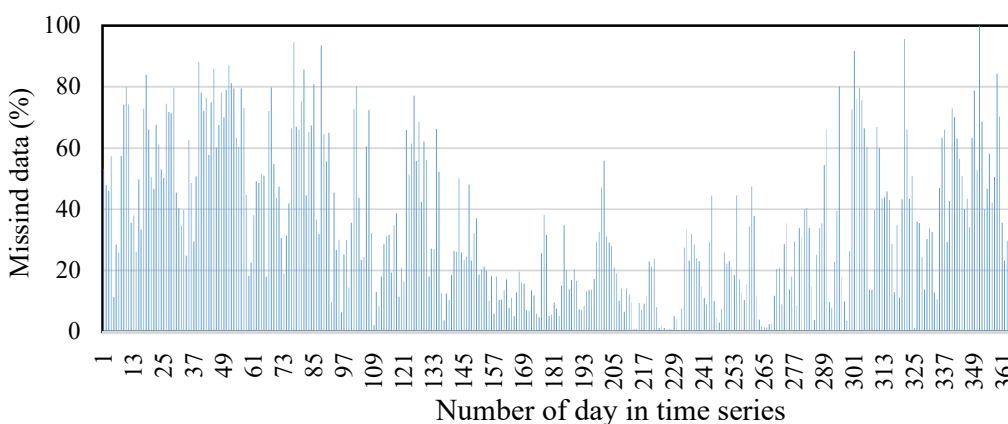

Figure 4. The missing data of each image (Spatial dispersion) in annual LST time series during the night.

The percentage of missing data caused by cloud cover in each pixel during one year (temporal dispersion) is shown in Fig. 5. As shown, the highest rate of missing data is found in the areas around Caspian Sea in Turkmenistan (the northern areas of studied region). The minimum rate of missing data is found in the southern part of study areas (central regions of Iran). In general, the northern and western regions of Iran involve the highest rate of missing data, considering the location of Alborz and Zagros mountain ranges and humidity barrier that causes the formation of cloud and prevents the penetration of humid air into the central regions of Iran. The lowest rate of missing data is observed in the central regions of Iran. Fig. 5 illustrates the histogram of temporal distribution map of the missing data. A large part of this time series has lost 20-60% of LST data as the result of the effects of cloud cover. On average, 36.37% of data have been lost due to cloud cover.
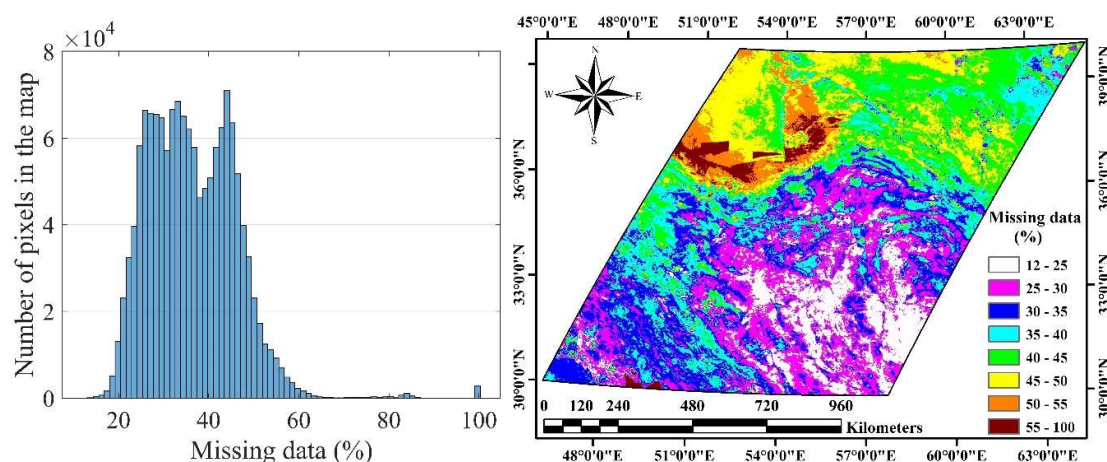

Figure 5. Temporal distribution map of the missing data in annual LST time series with 730 images (left) and histogram of temporal distribution map of missing data (right)

3.2.   Determining the parameters of SSA algorithm

Determining the window size and number of components are two important parameters in reconstructing a time series by SSA algorithm. Therefore, it is necessary to select the window size and number of components optimally. Fig. 6 shows the effects of the number of components in a window size of 60 in reconstructing a time series with 730 images. By increasing the number of components from 1 to 2, the coefficient of determination value (R2) increased significantly between the original data and reconstructed signal by SSA algorithm. Then, by increasing the number of components from 2 to 8, R2 value was not significantly increased. Fig. 6 (right) illustrates the effect of different window size on reconstruction of the same time series with four components. By increasing the window size, the R2 value between the original data and reconstructed data was reduced. Using small window sizes to reconstruct a time series with long-distance missing data has no favorable results, and using large window sizes not only reduces the accuracy of reconstruction but also increases the time of processing. Thus, in the present study window size of 60 was used to process the time series. In the following, the accurate determination of the number of significant components and window sizes by SSA algorithm are fully described.
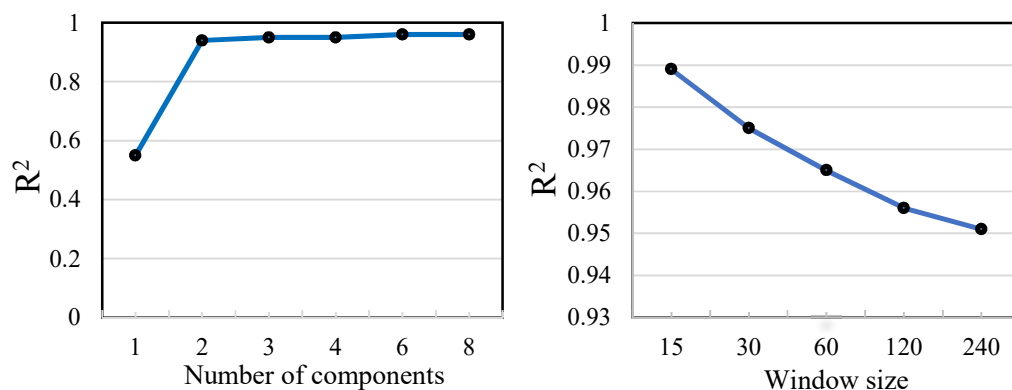
Figure 6. R^2 values between the original data and SSA reconstruction in a time series with 730 data

in different window sizes (right) and different number of components (left)

3.3.   Determining the significant periodic components in SSA algorithm

Figure 7 shows the singular values spectrum of a pixel as representative of data with 60 images as the window size. When we plot the singular values of a time series in descending order, we can usually distinguish between an initial steep slope, representing signals and a more or less flat tail related to noise. In this graph, the horizontal axis shows the number of modes (components) and the vertical axis shows the logarithm of the variance of related mode. Since the window size of 60 was selected, SSA divided this time series into 60 components. Therefore, there are 60 modes in the horizontal axis. The vertical axis shows the variance of components. Based on Fig. 7, components 1, 2 and 3 have the highest variance, respectively. Since components 1, 2 and 3 are located singly in the figure, each of them shows one trend of changes in the signal [31]. When two components are paired equally, their EOFs are transferred in a quarter-phase (phase difference). These components are effectively indicative of fluctuation in time series and are a part of periodic components [31].

The pair of SSA modes that are nearly equal and have the related EOFs which are in quadrature phase shift can efficiently represent an oscillation in time series [31]. This is because a pair of Eigen modes of a data-adaptive method like SSA can capture the periodicity in data having oscillations.
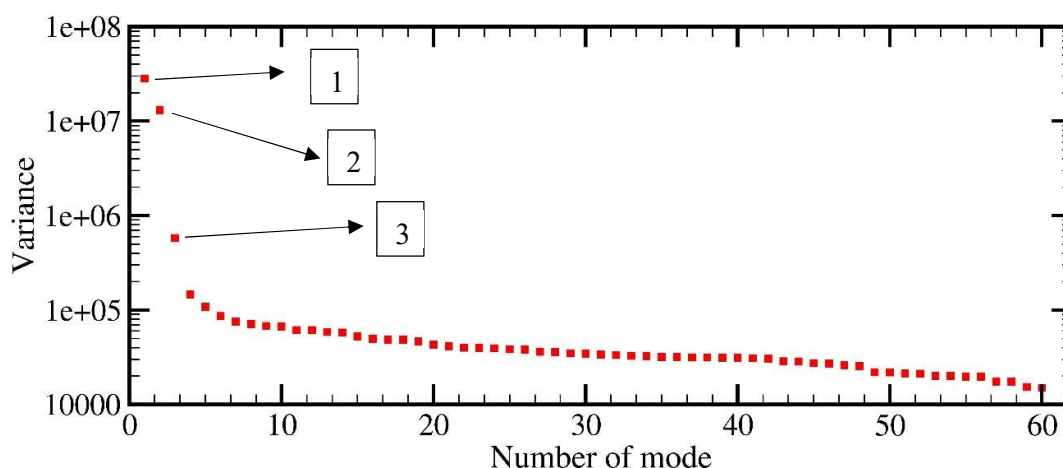
Figure 7. Singular values spectrum of data with window size of 60 image with 3 modes above the flat part of

the spectrum

Fig. 8 presents the normalized curve of specific figures in window sizes 30, 60, 120 and 240. As shown, by selecting the window size of 30 images, components 1 and 2 showed maximum variance in the signal so that a total of 95% of variance in this time series was related to these components (62% for component 1 and 33% for component 2). By increasing the window size to 60 images, almost the same results were obtained. By incasing the window size to 120 and 240 images, the effect of variance of component 3 increased; as in window size 240, components 1, 2 and 3 have 61, 24 and 10% of changes in this time series. In general, this figure showed two important points. Firstly, components 1, 2 and 3 have the highest variance in this time series. On the other hand, failure to select any of these components to reconstruct this time series causes an inaccurate signal. Secondly, the window size change in this time series will not have a significant effect on the results of signal reconstruction because by increasing the window size, components 1, 2 and 3 will still have the highest variance.
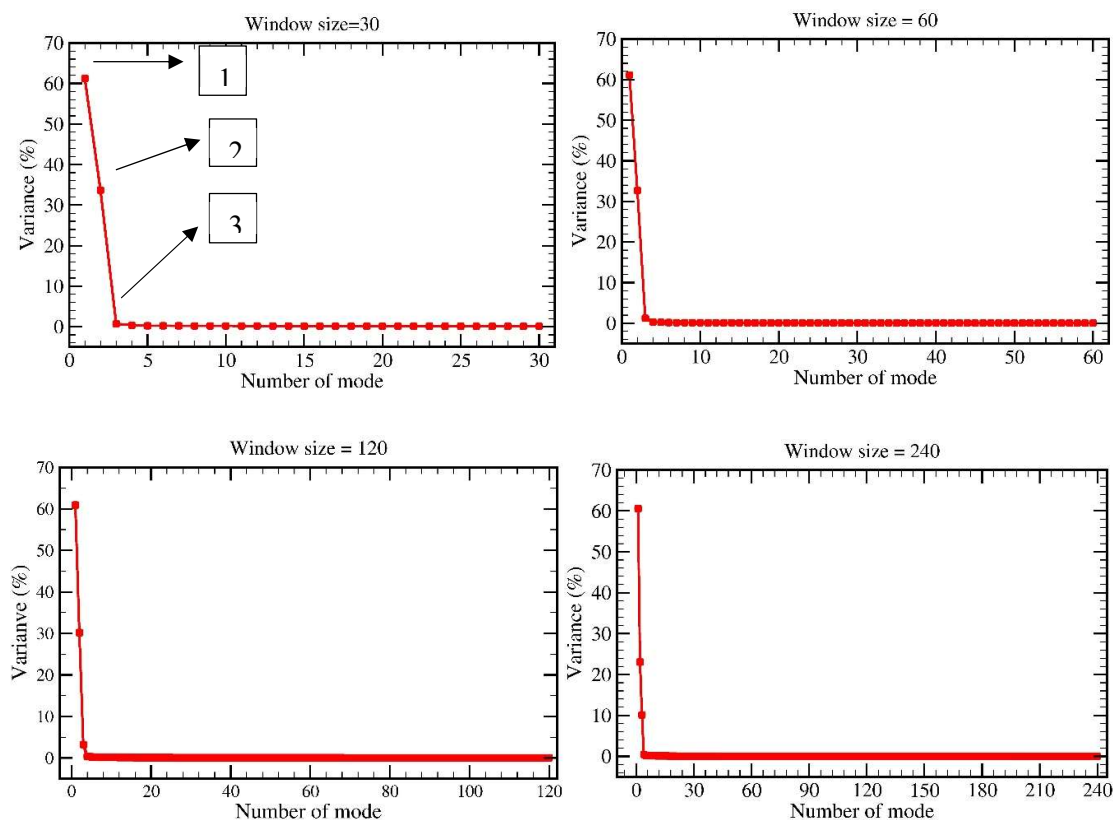

Figure 8. Graph of normalized singular values with 30, 60, 120 and 240 image window size.

3.4. Data-basis and null-hypothesis-basis Monte Carlo test results

The results of data and null-hypothesis EOF-s for pure red noise hypothesis are shown in Figure 9. The graphs show the power in the data eigenvalues and surrogates data bars vs. the frequency associated to EOFs. Obtaining a single frequency for each EOF is not straightforward due to the fact

that the EOFs resulting from SSA for an intrinsically periodic time series are not purely sinusoid unless they are composed of completely purely periodic components (e.g. synthetic periodic time series). To do so, Allen and Smith (1996) associated a frequency to each EOFs by maximizing the R-squared (R2) with a pure sinusoid for better visualization. The top graph of Figure 9 shows the result of the data-based Monte Carlo test.

Fig. 8 illustrates the eigenvalues versus their relevant frequencies. Fig. 9 shows that components 1, 2 and 3 are significantly major components of this time series with specific percentile of 97.5%. In this figure, the components are named according to the larger variance, respectively. The frequencies of components 1, 2 and 3 are 0.449, 0.002 and 0.006 (cycle/day) which belong to periods of 2 (2/2 = 1as daily component), 500 (500/2 = 250 almost one year) and 166 (166/2 = 88 day or almost 3 months as a trend) of the images.

The null-hypothesis EOFs test which is visible as lower graph of Figure 9 confirms the significance of these pairs. The null-hypothesis-based test has the advantage of having a lower probability of selecting a noise component as a signal component, which help to identify significant signal better and thus the result is more reliable [40].
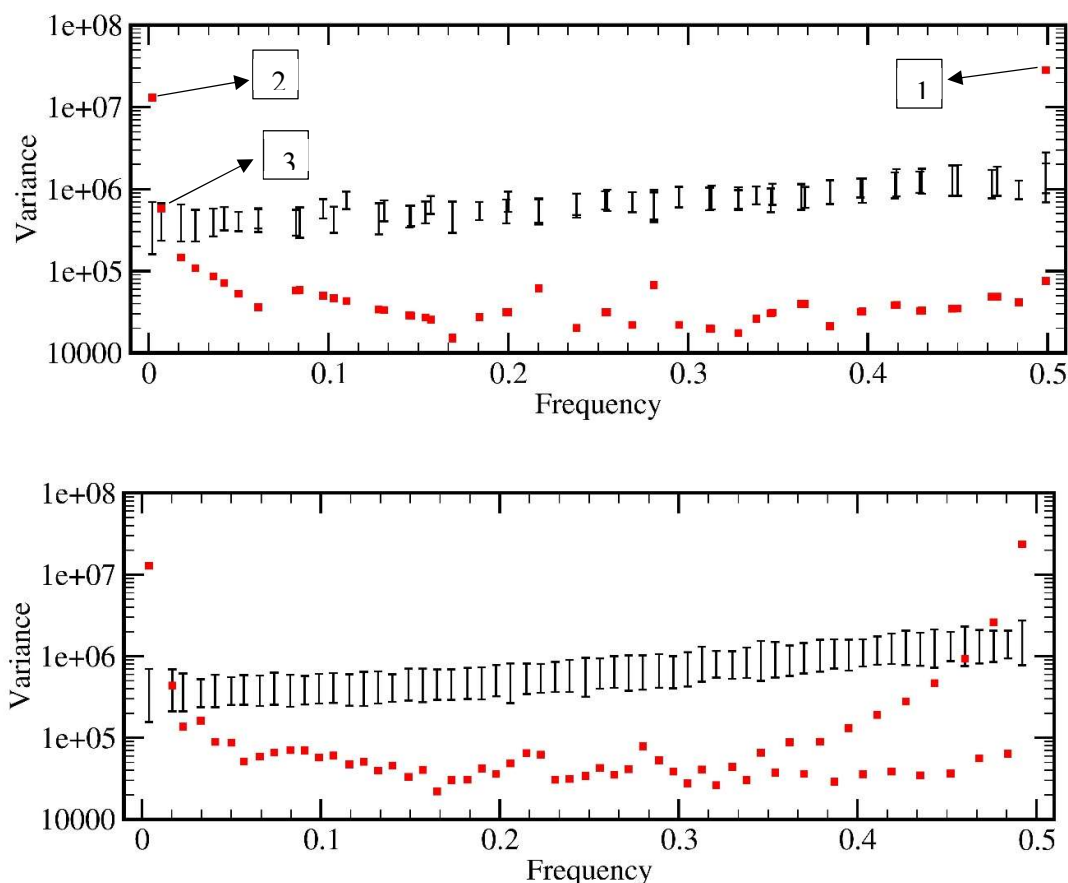

Figure 9. Monte Carlo SSA based on data (top) and null-hypothesis EOFs test (down).

Fig. 10 presents the temporal EOFs of the significant components in this time series based on the results of Monte Carlo test. Component 1 of the image in Monte Carlo test shows the same day/night temperature changes. As shown, the black lines in both image is repeated every day and night. The component 2 of the image indicates the annual temperature changes and component 3 almost shows the seasonal periods in this time series (the components 2 and 3 are shown in red and blue colors in figure 10, up) which belongs to a significant trend. Component 2 is shown as a small part of the annual period and it is periodically repeated in approximately each 730 images. For a better display, component 2 is drawn in window size 730 (Fig. 10, bottom). Component 3 is only part of a periodic signal with 166 images, which is shown as a trend in window size of 60 images (Fig. 10, up).
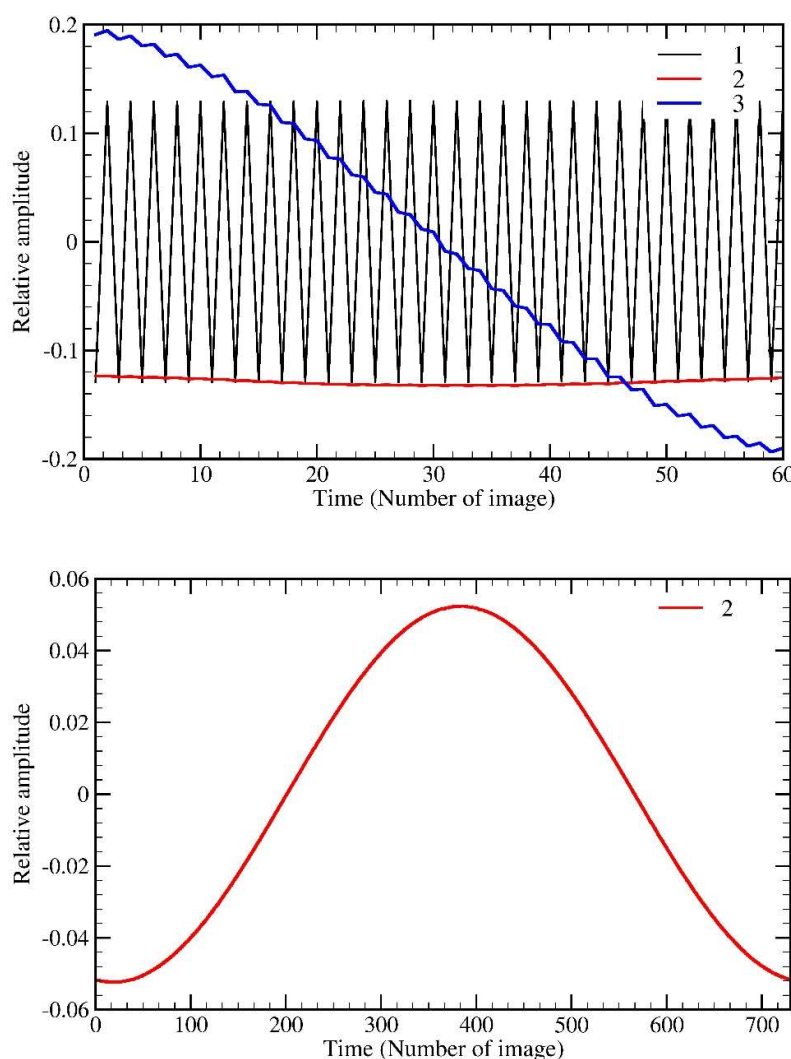

Figure 10. SSA T-EOFs of components 1, 2 and 3 values with 60 image window size (top) and SSA T-EOFs of

components 3 with 730 image window size (down).

Generally, the results of Monte Carlo test and those of other tests showed that window size 60 yielded reliable results for reconstruction of this time series. In window size 60, three significant

components can be differentiated in this time series. Therefore, three components, including 1, 2 and 3 were used to reconstruct the whole time series in window size 60 in M-SSA.

3.5.   Assessment of results

Fig. 11 displays the RMSE map between the original data and reconstructed data by SSA along annual LST time series with 730 images. As shown in Fig. 11 (left), the minimum RMSE is observed on the Caspian Sea. At other points, the RMSE is variable from 2 to 4 kelvin (K). Fig. 11 (right) shows the histogram of RMSE map of LST. Overall, the mean RMSE of the reconstructed LST by SSA algorithm in the studied time series with 730 images is 2.95 K.
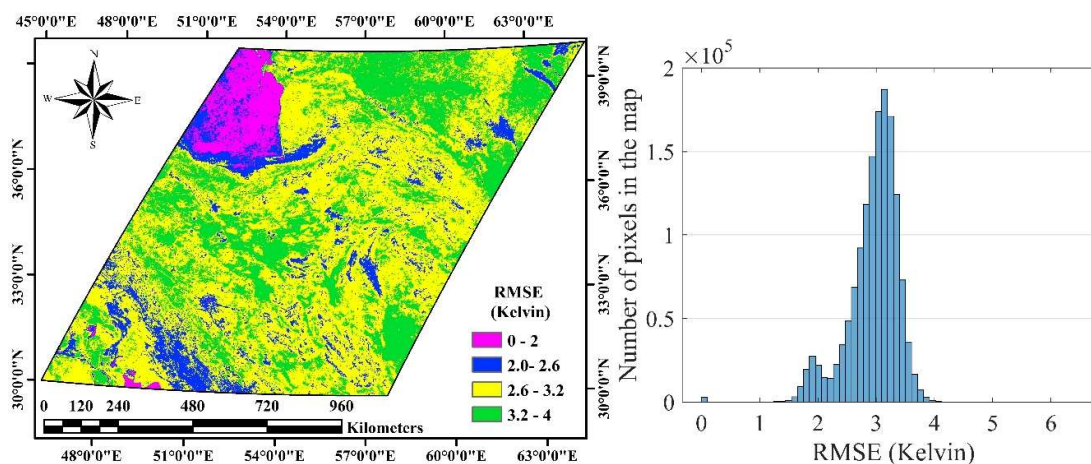

Figure 11. RMSE map between the original data and reconstructed data by SSA algorithm in annual LST time series with 730 images (left) and histogram of RMSE map (right)

Fig. 12 indicates image number 215 as an example in the studied time series before and after reconstruction by SSA algorithm. As shown in Fig. 12 (left), 50.25% of the data of this image are missing due to cloud cover (black areas). It should also be noted that in addition to the missing data in this image, the outlier data exist, which are not detectable spatially but are detectable in a time domain. Hence, the missing data cannot be reconstructed by spatial interpolation techniques. According to Fig. 12 (right), SSA algorithm could reconstruct the missing and outlier data of LST image by spatio-temporal interpolation technique. The reconstructed LST image shows that the missing data of the image that are filled by SSA are similar to the main image pattern. This is indicative of the capability of this algorithm in solving the missing and outlier data problem caused by cloud cover.
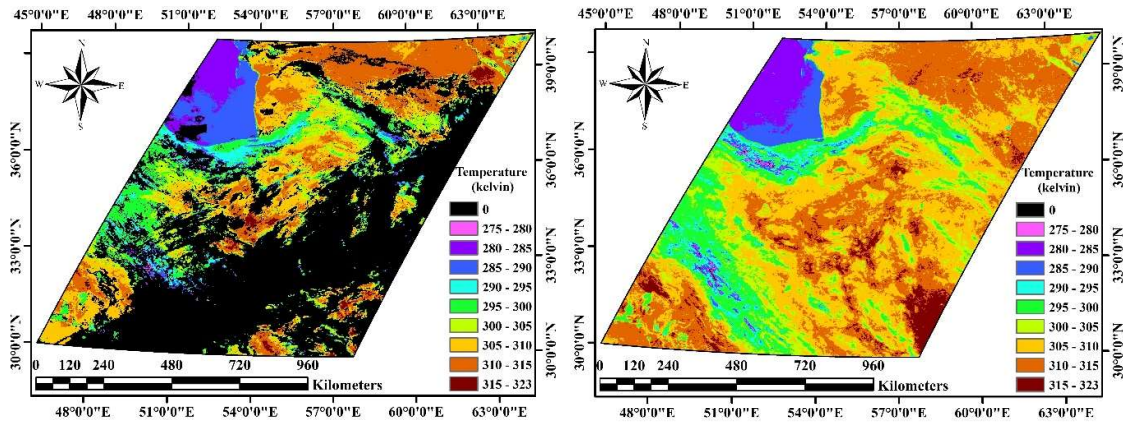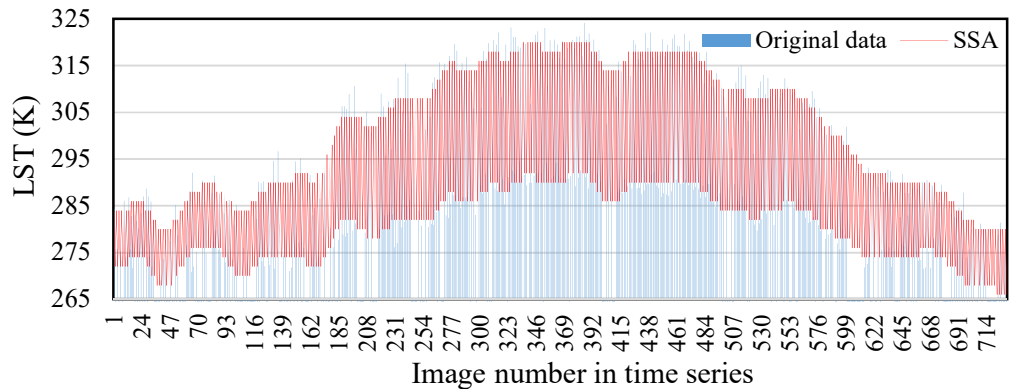
Figure 12. An LST image of the studied region before reconstruction (left) and after reconstruction (right) by

SSA

Fig. 13 indicates a time series along a pixel together with the reconstruction results of SSA algorithm. As shown, SSA algorithm effectively reconstructed the missing and outlier data in this time series. According to Fig. 13 (up), the temperature difference between day and night is lower at the start and end of time series than the middle parts of this time series. SSA algorithm was effective in identification of these changes and trends, and considered these temperature changes in the reconstructed signal. As shown in Fig. 13 (up), there is a growing temperature rise from the start to the middle of time series, followed by a decreasing trend. This trend is the same annual temperature changes being considered as component 2 in determining the significant periodic components. Partial thermal fluctuations are observed during the year, which are detectable as multi-month periods in Fig. 13. These changes are due to considering component 3. Fig. 13 (down) presents a more translucent display of a part of reconstructed signal. The temperature changes during day and night are seen in this figure, which are caused by component 1. Hence, failure to consider any of the mentioned components in reconstruction of this time series causes inaccurate findings.
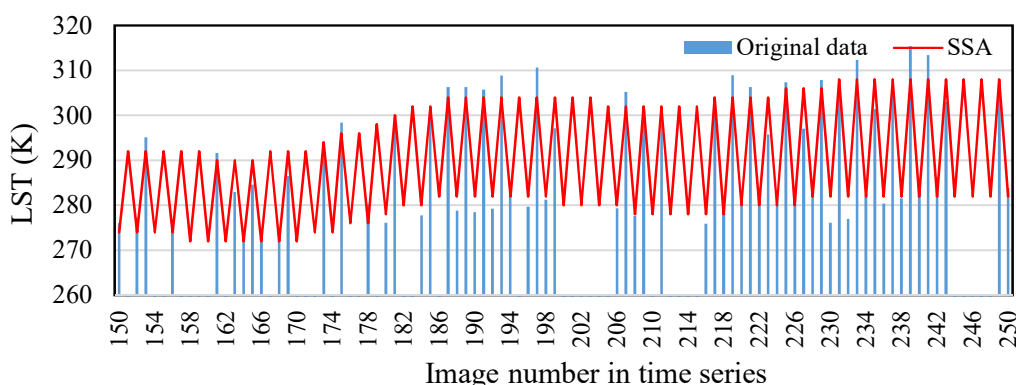
Figure 13. A time series along a pixel together with reconstruction results of SSA (up) and magnification of a part of time series (down)

4.   Conclusion

In the present study, SSA algorithm was used to reconstruct the missing data caused by cloud cover in daily LST time series of MODIS sensor during 2015. According to the results of accuracy evaluation of the SSA algorithm, the mean reconstruction error in the whole studied region was 2.95 °K. These results are promising owing to the amount of missing data caused by cloud cover in the time series with 730 LST images because these figures are in line with the temperature measurement errors (±3 °K) by remote sensing [42]. The findings showed the minimum reconstruction error on the Caspian Sea (<2.6 °K). This is due to the high thermal specific capacity of water which results in little and gradual water temperature changes per day and during the year. Hence, the reconstructed signal is more compatible with the original data. Ghafarian Malamiri and Zare Khormizie (2017) reconstructed a daily MODIS LST time series during a year by HANTS algorithm and reported that reconstruction error of LST time series was lower on the sea [43]. Further, the results of Ghafarian Malamiri and Zare Khormizie (2017) showed that reconstruction error in annual LST time series during the day from MODIS sensor in this study region was 3.87 °K on average [43]. Therefore, the reconstruction accuracy is higher in SSA algorithm than HANTS algorithm. This is because only temporal interpolation is used in HANTS algorithm; whereas, spatio-temporal interpolation is used in SSA algorithm in M-SSA to reconstruct the time series. Since the daily MODIS LST time series includes LST during the day and night, this time series can be processed in two different ways to reconstruct the effects of cloud cover. First, LST images during the day and at night can be reconstructed into two separate time series. Second, LST images can be put together as a day/night sequence and be reconstructed as a time series. In this study, the second method was applied. When daily MODIS LST time series is used in day/night sequence, a significant component (component 1 which indicates the day/night sequence) is created. When this component is used to reconstruct the time series, the accuracy will increase dramatically because the daily missing data caused by cloud cover at certain points may be present during the night and vice versa.

Three significant components were identified in daily MODIS LST time series with day/night sequence. The first component was the same day and night temperature changes, the second

component was annual temperature changes and the third component was seasonal temperature changes. The number of significant components (periods or trends) varies in different time series. Reconstructing the hourly LST time series, Ghafarian Malamiri (2015) reported three significant periodic components 24, 12 and 8 hours were differentiated in the window size of 72 hours [16].

According to the findings, SSA algorithm can effectively resolve the missing data problem in MODIS LST time series. In a research using SSA and M-SSA algorithms, Ghafarian et al. (2012) reconstructed the time series of hourly LST data and showed an absolute mean error of 2.25 °K in an hourly time series with 63% missing data caused by cloud cover [1]. Another study indicated that M-SSA using spatio-temporal correlation effectively reconstructed the MODIS LAI [35]. In general, the results showed that SSA algorithm could be effectively used to resolve the missing data problem caused by cloud cover in daily MODIS LST time series.

References:

1.   Ghafarian, H.R.; Menenti, M.; Jia, L.; den Ouden, H. Reconstruction of cloud-free time series satellite observations of land surface temperature. *EARSel eProc* **2012**, *11*, 123-131.

2.   Xu, Y.; Shen, Y. Reconstruction of the land surface temperature time series using harmonic analysis. *Computers & geosciences* **2013**, *61*, 126-132.

3.   Sobrino, J.; El Kharraz, J.; Li, Z.-L. Surface temperature and water vapour retrieval from modis data. *International Journal of Remote Sensing* **2003**, *24*, 5161-5182.

4.   Running, S.W.; Justice, C.; Salomonson, V.; Hall, D.; Barker, J.; Kaufmann, Y.; Strahler, A.H.; Huete, A.; Muller, J.-P.; Vanderbilt, V. Terrestrial remote sensing science and algorithms planned for eos/modis. *International journal of remote sensing* **1994**, *15*, 3587-3620.

5.   Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z.-l. Validation of the land-surface temperature products retrieved from terra moderate resolution imaging spectroradiometer data. *Remote sensing of Environment* **2002**, *83*, 163-180.

6.   Sun, D.; Pinker, R.T.; Basara, J.B. Land surface temperature estimation from the next generation of geostationary operational environmental satellites: Goes m–q. *Journal of Applied Meteorology* **2004**, *43*, 363-372.

7.   Estes Jr, M.G.; Al-Hamdan, M.Z.; Crosson, W.; Estes, S.M.; Quattrochi, D.; Kent, S.; McClure, L.A. Use of remotely sensed data to evaluate the relationship between living environment and blood pressure. *Environmental health perspectives* **2009**, *117*, 1832.

8.   Tatem, A.J.; Goetz, S.J.; Hay, S.I. Terra and aqua: New data for epidemiology and public health. *International Journal of Applied Earth Observation and Geoinformation* **2004**, *6*, 33-46.

9.   Schmugge, T.; French, A.; Ritchie, J.C.; Rango, A.; Pelgrum, H. Temperature and emissivity separation from multispectral thermal infrared observations. *Remote Sensing of Environment* **2002**, *79*, 189-198.

10.   Rousta, I.; Doostkamian, M.; Haghighi, E.; Mirzakhani, B. Statistical-synoptic analysis of the atmosphere thickness pattern of iran's pervasive frosts. *Climate* **2016**, *4*, 41.

11.   Soltani, M.; Laux, P.; Kunstmann, H.; Stan, K.; Sohrabi, M.M.; Molanejad, M.; Sabziparvar, A.A.; Ranjbar SaadatAbadi, A.; Ranjbar, F.; Rousta, I., *et al.* Assessment of climate variations in temperature and precipitation extreme events over iran. *Theoretical and Applied Climatology* **2016**, *126*, 775-795.

12.   Soltani, M.; Zawar-Reza, P.; Khoshakhlagh, F.; Rousta, I. Mid-latitude cyclones climatology over caspian sea southern coasts–north of iran. In *21st Conference on Applied Climatology, American*

Meteorological Society (AMS), American Meteorological Society: USA, 2014; Vol. 1, pp 1-3 avalable at: , https://ams.confex.com/ams/21Applied17SMOI/webprogram/Paper246601.html.

13.   Julien, Y.; Sobrino, J.A. Comparison of cloud-reconstruction methods for time series of composite ndvi data. *Remote Sensing of Environment* **2010**, *114*, 618-625.

14.   Zhou, J.; Jia, L.; Menenti, M. Reconstruction of global modis ndvi time series: Performance of harmonic analysis of time series (hants). *Remote Sensing of Environment* **2015**, *163*, 217-228.

15.   Ackerman, S.A.; Strabala, K.I.; Menzel, W.P.; Frey, R.A.; Moeller, C.C.; Gumley, L.E. Discriminating clear sky from clouds with modis. *Journal of Geophysical Research: Atmospheres* **1998**, *103*, 32141-32157.

16.   Ghafarian Malamiri, H. Reconstruction of gap-free time series satellite observations of land surface temperature to model spectral soil thermal admittance. doctoral thesis, Delft, Technische Universiteit Delft, The Netherlands., 2015.

17.   Jia, L.; Shang, H.; Hu, G.; Menenti, M. Phenological response of vegetation to upstream river flow in the heihe rive basin by time series analysis of modis data. *Hydrology and Earth System Sciences* **2011**, *15*, 1047-1064.

18.   Roerink, G.; Menenti, M.; Verhoef, W. Reconstructing cloudfree ndvi composites using fourier analysis of time series. *International Journal of Remote Sensing* **2000**, *21*, 1911-1917.

19.   Verhoef, W. Application of harmonic analysis of ndvi time series (hants). *Fourier analysis of temporal NDVI in the Southern African and American continents* **1996**, *108*, 19-24.

20.   Verhoef, W.; Menenti, M.; Azzali, S. Cover a colour composite of noaa-avhrr-ndvi based on time series analysis (1981-1992). *International Journal of Remote Sensing* **1996**, *17*, 231-235.

21.   Menenti, M.; Azzali, S.; Verhoef, W.; Van Swol, R. Mapping agroecological zones and time lag in vegetation growth by means of fourier analysis of time series of ndvi images. *Advances in Space Research* **1993**, *13*, 233-237.

22.   Jiang, X.; Wang, D.; Tang, L.; Hu, J.; Xi, X. Analysing the vegetation cover variation of china from avhrr-ndvi data. *International Journal of Remote Sensing* **2008**, *29*, 5301-5311.

23.   Julien, Y.; Sobrino, J.A.; Verhoef, W. Changes in land surface temperatures and ndvi values over europe between 1982 and 1999. *Remote Sensing of Environment* **2006**, *103*, 43-55.

24.   Alfieri, S.; Lorenzi, F.D.; Menenti, M. Mapping air temperature using time series analysis of lst: The sintesi approach. *Nonlinear Processes in Geophysics* **2013**, *20*, 513-527.

25.   Broomhead, D.S.; King, G.P. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena* **1986**, *20*, 217-236.

26.   Kondrashov, D.; Ghil, M. Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Processes in Geophysics* **2006**, *13*, 151-159.

27.   Vautard, R.; Yiou, P.; Ghil, M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena* **1992**, *58*, 95-126.

28.   Golyandina, N.; Zhigljavsky, A. *Singular spectrum analysis for time series*. Springer Science & Business Media: 2013.

29.   Golyandina., N.; Nekrutkin., V.; Zhigljavsky., A. *Analysis of time series structure: Ssa and related techniques*. CHAPMAN &HALL/CRC: Washington DC, USA, 2001; p 310.

30.   Ghil, M.; Vautard, R. Interdecadal oscillations and the warming trend in global temperature time series. *Nature* **1991**, *350*, 324.

31.   Vautard, R.; Ghil, M. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D: Nonlinear Phenomena* **1989**, *35*, 395-424.

32.   Yiou, P.; Baert, E.; Loutre, M.-F. Spectral analysis of climate data. *Surveys in Geophysics* **1996**, *17*, 619-663.

33.   Yiou, P.; Sornette, D.; Ghil, M. Data-adaptive wavelets and multi-scale singular-spectrum analysis. *Physica D: Nonlinear Phenomena* **2000**, *142*, 254-290.

34.   Kondrashov, D.; Shprits, Y.; Ghil, M. Gap filling of solar wind data by singular spectrum analysis. *Geophysical research letters* **2010**, *37*.

35.   Wang, D.; Liang, S. In *Singular spectrum analysis for filling gaps and reducing uncertainties of modis land products*, Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, 2008; IEEE: pp V-558-V-561.

36.   Marques, C.; Ferreira, J.; Rocha, A.; Castanheira, J.; Melo-Gonçalves, P.; Vaz, N.; Dias, J. Singular spectrum analysis and forecasting of hydrological time series. *Physics and Chemistry of the Earth, Parts A/B/C* **2006**, *31*, 1172-1179.

37.   Wan, Z. Modis land-surface temperature algorithm theoretical basis document (lst atbd). *Institute for Computational Earth System Science, Santa Barbara* **1999**, *75*.

38.   Musial, J.P.; Verstraete, M.M.; Gobron, N. Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. *Atmospheric chemistry and physics* **2011**, *11*, 7905-7923.

39.   Elsner, J.B.; Tsonis, A.A. *Singular spectrum analysis: A new tool in time series analysis*. Springer Science & Business Media: 2013.

40.   Allen, M.R.; Smith, L.A. Monte carlo ssa: Detecting irregular oscillations in the presence of colored noise. *Journal of climate* **1996**, *9*, 3373-3404.

41.   Schoellhamer, D.H. Singular spectrum analysis for time series with missing data. *Geophysical Research Letters* **2001**, *28*, 3187-3190.

42.   Li, Z.-L.; Tang, B.-H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment* **2013**, *131*, 14-37.

43.   Ghafarian Malamiri, H.R.; Zare Khormizie, H. 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* **2017**, *8*, 37-55.