LSWAVE

A MATLAB software for the least-squares wavelet and cross-wavelet analyses

User Manual

written by

Ebrahim Ghaderpour

Canada 2019

Email: ebrahimg2@gmail.com
URL: www.ghader.org
Tel: +1 (647) 879-4343
ORCID: 0000-0002-5165-1773

Acronyms

LSWAVE	Least-Squares Wavelet (MATLAB software)
GUI	Graphical User Interface
TS	Time Series
2TS	Two Time Series
LSS	Least-Squares Spectrum
LSSA	Least-Squares Spectral Analysis
ALLSSA	Antileakage Least-Squares Spectral Analysis
LSWA	Least-Squares Wavelet Analysis
LSWS	Least-Squares Wavelet Spectrogram
LSCSA	Least-Squares Cross-Spectral Analysis
LSCS	Least-Squares Cross-Spectrum
LSCWA	Least-Squares Cross-Wavelet Analysis
LSCWS	Least-Squares Cross-Wavelet Spectrogram
PSD	Power Spectral Density

Installation

- Download the zip file from the internet
- Extract all the files into a directory
- Run LSWAVE_GUI.m in MATLAB to open the GUI

Background

The MATLAB software package LSWAVE includes the LSSA, ALLSSA, LSWA, LSCSA, and LSCWA. These methods are based on Ghaderpour and Pagiatakis (2017) and Ghaderpour et al. (2018a, b) that are extensions of the LSSA proposed by Vaničček P (1969). Users may refer to Ghaderpour (2018) for comprehensive details of these methods.

Introduction to GUI

The LSWAVE is developed for time/data series analysis. It can analyze any type of series with an associated covariance matrix that may have trends and/or datum shifts. The GUI application is user-friendly, and it has a display bar on its top right corner to guide users throughout the process. Users need to follow the steps below for a successful analysis (Fig. 1).

Step 1. File → Open

Files to be opened are series with associated covariance matrices (if they exist). Users may open a file (optional) containing a set of cyclic frequencies $\mathbf{\Omega} = \{\omega_k;\ k = 1, 2, \ldots, \kappa\}$ (cycles per unit time) that has only one column. This set can be any set of positive real numbers based on the scope of analysis. If users do not open such file, then the GUI automatically chooses a set of cyclic frequencies or wavenumbers for the analysis, described in Step 2 below. The extensions of these files can be *.dat or *.txt or *.xlsx.

The file containing the series usually comprises two columns: the first column is for the times, $\mathbf{t}$, or other, and the second column is for the series values, $\mathbf{f}$. If the standard deviations of the series values are available, then the file may have a third column for the standard deviations. If there is a covariance matrix of order n associated with the series of size n, then the file should be formatted so that its third to $n + 2$ columns are the covariance matrix values.

Let M be the average sampling rate. Parameter M determines the Nyquist frequency when series is equally spaced. The GUI calculates M automatically and displays it in the menu bar under the "Window Size Parameters". Since the Nyquist frequency is not explicitly defined for inherently unequally spaced series, users may choose $M/2$ as an upper bound for the frequency band. The sampling rates of two series may not be always identical, i.e., $M_1 \neq M_2$.

For the sake of computational efficiency and display purposes, the GUI subtracts the first entry of the time vector from each of the times. To investigate the coherency and phase differences, the two series must have the same units in the first column (e.g., hours, years, m, km), and these units must be in the same system, e.g., same local time. The GUI subtracts the minimum of the start times of the two series from the first column of each series to avoid the introduction of a phase shift for the sake of numerical efficiency and nicer output display.

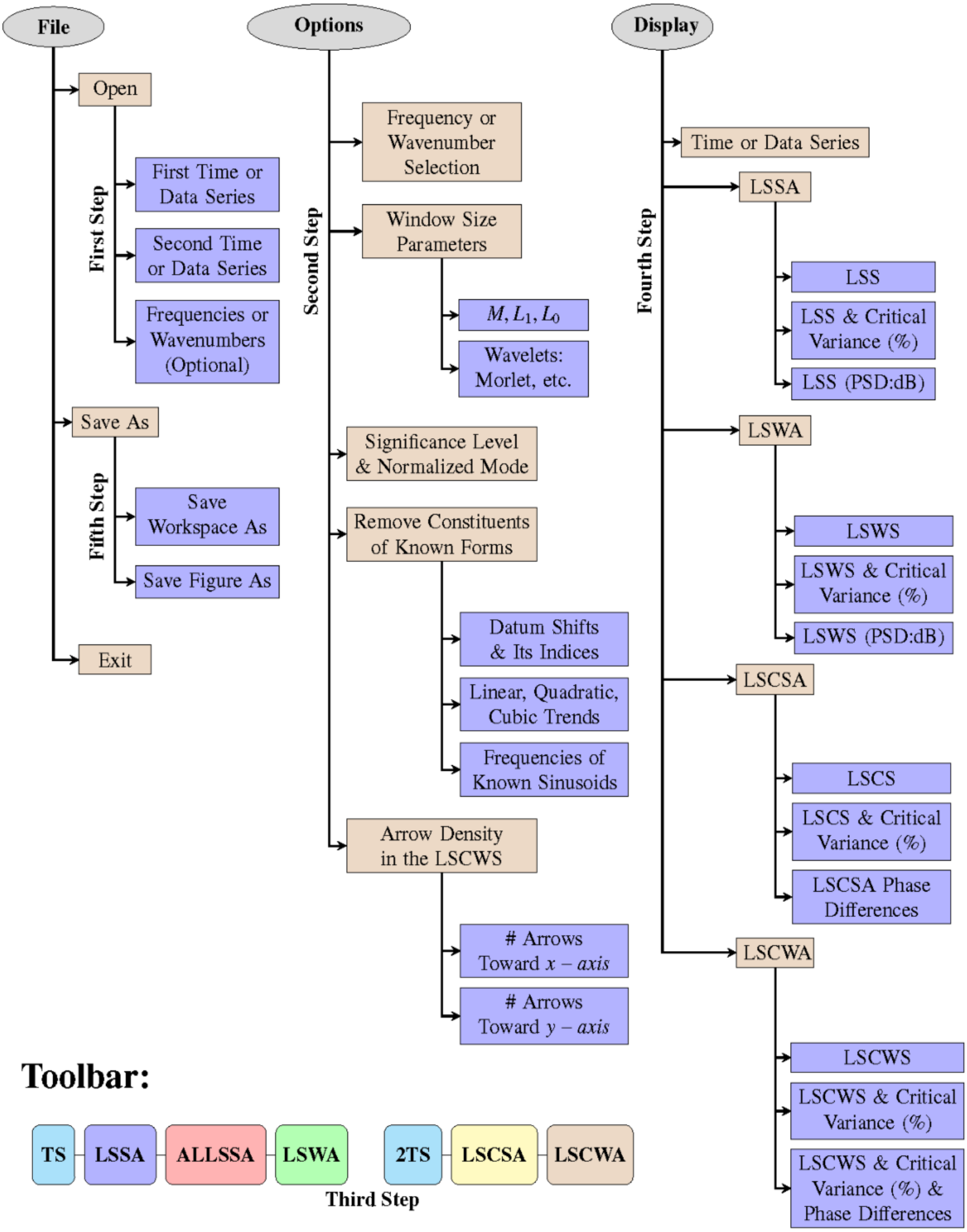

Fig. 1 Partial menu bar of the GUI

Step 2. Options

The GUI has a panel to select a set of cyclic frequencies $\mathbf{\Omega}$ (Fig. 2a). A lower bound for this set is calculated automatically as the inverse of twice of the series length under consideration with upper bound $M/2$. The GUI automatically calculates 200 equally spaced cyclic frequencies from the lower bound to the upper bound; however, users may modify the frequency band, depending on the desired frequency resolution. When there are two series for the coherency analysis, the common lower bound is the maximum of the lower bounds, and upper bound is the minimum of $M_1/2$ and $M_2/2$ as the default settings.

Parameter L_1 is the number of cycles of sinusoids within the translating windows, and L_0 is the additional number of samples within the translating windows in the least-squares fitting process. These parameters determine the segment size that is, $L = \lfloor L_1 M/\omega_k \rfloor + L_0$ (samples), where $\lfloor \cdot \rfloor$ is the floor function. If $L_1 > 0$, the window size decreases when the frequency increases, allowing the detection of short-duration signatures. In unequally spaced series, the sinusoids of frequency ω_k within translating windows may not exactly complete L_1 cycles, and so L_1 is an approximation. The larger the L_0, the higher the frequency resolution but the poorer the time resolution will be in the spectrogram. The selection of parameters L_1 and L_0 depends on whether

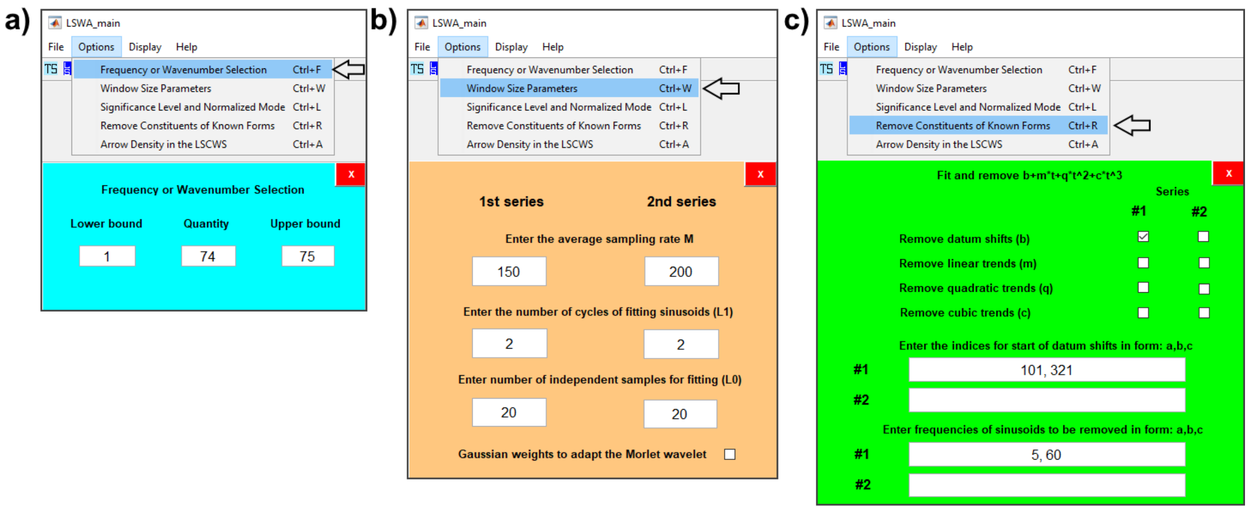

Fig. 2 MATLAB GUI panels for a) frequency or wavenumber selection, b) window size parameters, and c) removing constituents of known forms. Note how the indices for the start of datum shifts and the known frequencies after detection are entered in panel c

the series is weakly or strongly non-stationary. In addition, it depends on the desired time and frequency resolution and the number of constituents being estimated simultaneously within each window whose size should be adequate to avoid singularity of the matrix of normal equations.

For example, if $M = 60$ samples/h and $L_1 = 1$ cycle, then for cyclic frequencies between 20 and 30 cycles/h, we have $L = 2 + L_0$ samples within the translating windows. Now, if there are q constituents to be estimated simultaneously with the cosine and sine basis functions of a cyclic frequency between 20 and 30 cycles/h, then L_0 must be an integer greater than q to avoid the singularity of the matrix of normal equations. The GUI considers window size parameters $L_1 = 2$ and $L_0 = 20$ as the default settings for each series; however, users may modify them (Fig. 2b).

Users can check a box for Morlet wavelet to be used in the LSWA and LSCWA (Fig. 2b). By checking this box, the weight matrix within each translating window will be a diagonal matrix whose diagonal elements are the Gaussian values, adapting the sinusoidal basis functions to the Morlet wavelet in the least-squares sense. Scale 0.0125 is used for the Gaussian weights with $L_1 = 6$ and $L_0 = 0$ as the default settings.

When series values have been derived from populations of random variables following multidimensional normal distributions, users can rigorously identify the statistically significance of peaks in the LSS, LSWS, LSCS, and LSCWS at certain confidence level that is usually 90% or 95% or 99%. The critical percentage values are valid only when the normalized mode for the spectrum and spectrogram is checked. The critical percentage value increases when the window size decreases and vice versa.

There is an option to fit and remove a polynomial of degree three or less from each segment of series, being considered simultaneously with sinusoidal basis functions, with the general form $b + mt + qt^2 + ct^3$ (Fig. 2c). For each series, there are four boxes corresponding to datum shifts ($[\mathbf{1}]$ to estimate b for each segment), linear trends ($[\mathbf{t}]$ to estimate m for each segment), quadratic trends ($[\mathbf{t}^2]$ to estimate q for each segment), and cubic trends ($[\mathbf{t}^3]$ to estimate c for each segment). Series may have sudden jumps or datum shifts, and researchers may approximately find the indices of the start points of these shifts in advance of analysis. The GUI has an option to enter these indices (Fig. 2c). The box corresponding to datum shifts must always be checked to consider these indices. This box is checked as the default setting.

Some of the frequencies may be initially known, and so users may want to remove the sinusoids of those frequencies simultaneously with trends to study the residual segments for any hidden signatures, coherency, and phase differences. These frequencies can also be estimated by analyzing the original series, so they will be known to users for the next round of analysis. There are two boxes in the GUI to enter the known frequencies (Fig. 2c).

Users can choose the number of phase arrows being plotted toward time and frequency axes in the LSCWS. The direction of an arrow shows how much the constituent of the second series segment lags or leads the constituent of the first series segment. It lags if the arrow is in the first or second quadrant, and it leads if the arrow is in the third or fourth quadrant of the trigonometric circle.

Step 3. Toolbar

There are seven run buttons in the GUI toolbar as shown in Fig. 1. Choosing buttons “TS” and “2TS” will plot the series, and selecting buttons “LSSA”, “ALLSSA”, “LSWA”, “LSCSA” and “LSCWA” will run the MATLAB scripts of the respective methods. The frequency resolutions of these methods depend on the input frequencies, $\mathbf{\Omega}$.

By selecting the “ALLSSA” button, all the statistically significant spectral peaks will be computed after one round of iterations. The ALLSSA spectrum is a frequency-amplitude spectrum. In the ALLSSA, since the sinusoidal basis functions will be fitted to the entire series, its spectrum may not show peaks corresponding to short duration signatures or components with variability of amplitude and frequency over time. For non-stationary series, users may save the residual series and then select the “ALLSSA” button on the residual series.

The frequency-partitioning in the ALLSSA depends on the input set of frequencies. More precisely, in an iteration step, suppose that $\omega_k\,(1 < k < \kappa)$ in $\mathbf{\Omega}$ is selected whose corresponding peak has maximum percentage variance in the spectrum. Then the ALLSSA searches through 100 equally spaced frequencies from $(\omega_{k-1} + \omega_k)/2$ to $(\omega_k + \omega_{k+1})/2$ to find a more accurate frequency. Therefore, the denser the input set of frequencies is, the smaller the norm of the residual series may become.

Step 4. Display

The GUI "Display" in the menu bar includes (Fig. 1):

- "Time or Data series" with four options for displaying the series: display only the first series with/without error bars or display both series with/without error bars.
- "LSSA" with three options: display the spectrum, display the spectrum and critical percentage variance that is a horizontal line parallel to the frequency axis, or display the spectrum in PSD, decibel (dB).
- "LSWA" with three options: display the spectrogram, display the spectrogram and stochastic surface (critical percentage variance), or display the spectrogram in PSD (dB).
- "LSCSA" with three options: display the cross-spectrum, display the cross-spectrum and critical percentage variance (a horizontal line parallel to the frequency axis), or display the phase differences.
- "LSCWA" with three options: display the cross-spectrogram, display the cross-spectrogram and stochastic surface with/without the phase differences (arrows).

The spectrograms and their stochastic surfaces will be displayed using MATLAB *mesh* function that is useful for unequally spaced series or *surf* function. These can be achieved by selecting the displaying options in "LSWA" and "LSCWA" or using the accelerators: Ctrl+V, Ctrl+C, Ctrl+P for the LSWA, and Ctrl+X, Ctrl+Y, Ctrl+Z for the LSCWA. Users may use the same accelerator repeatedly to switch displays between *mesh* and *surf*. The 'freezeColors' tool developed by Iversen enables simultaneous display of the spectrogram and stochastic surface in jet and gray colormaps, respectively.

Step 5. File → Save As

Depending on which buttons in the GUI toolbar are selected and ran successfully, users may save the work as *.dat or *.txt or *.xlsx. For each method, a separate window will pop up to save. Furthermore, they may save the GUI figure as *.jpeg or *.png or *.tiff in high resolution.

In the LSSA, users may save the estimated coefficients of the constituents of known forms and their estimated covariance matrix, computed by the covariance law. The diagonal entries of this covariance matrix from top left to bottom right are the variances of estimated coefficients for datum shifts, trends, and cosine and sine functions of known frequencies/wavenumbers in ascending order and if any of them are entered by users. Similarly, users may save the coefficients of constituents estimated by the ALLSSA and their estimated covariance matrix in two separate files. The amplitudes and phases calculated from the estimated coefficients of cosine and sine basis functions with their errors will also be saved.

Examples

There are two synthetic time series “SyntheticTS1.dat” and “SyntheticTS2.dat” and a GPS height time series from PRDS “GPS_PRDS_Height_TS.dat” that come with the LSWAVE, and users may open them in the GUI and analyze them. The detailed analyses of these time series can be found in the toolbox paper.

For example, users may open “SyntheticTS1.dat” in the GUI as the first time series with the same input values shown in Fig. 2. Then select the “LSSA” and “LSWA” buttons in the GUI toolbar. From the “Display” in the GUI menu bar, if they select LSS with critical percentage variance and LSWS with stochastic surface, then the result should appear as shown in Fig. 3, after rotating the 3D spectrogram and inserting the color bars using the additional buttons in the toolbar.

The LSWAVE also includes “Output_ALLSSA_GPS_PRDS_Height_TS.dat” that is the ALLSSA output result of the GPS PRDS height time series. To create this output, users need to enter 0.2 c/y and 8.2 c/y as the lower and upper bounds for the set of frequencies with quantity 40 in the GUI panel shown in Fig. 2a. Also, they should enter the indices 304, 427, 1588, and 3737 for the starts of datum shifts in the GUI panel shown in Fig. 2c and check the boxes for datum shifts and linear trend. Then select the “ALLSSA” button in the GUI toolbar and save the result.

For the LSWA of the GPS series, if users check the box in Fig. 2b, then the weight matrices within the segments of series will be diagonal matrices whose diagonal entries are the inverse of variances of measurements multiplied by the Gaussian weights, adapting the sinusoids to the Morlet wavelet in the lest-squares sense. Applying the Gaussian weights smooths the spectral peaks with an optimal time-frequency resolution in LSWS. Checking this box will reset the window size parameters to $L_1 = 6$ and $L_0 = 0$.

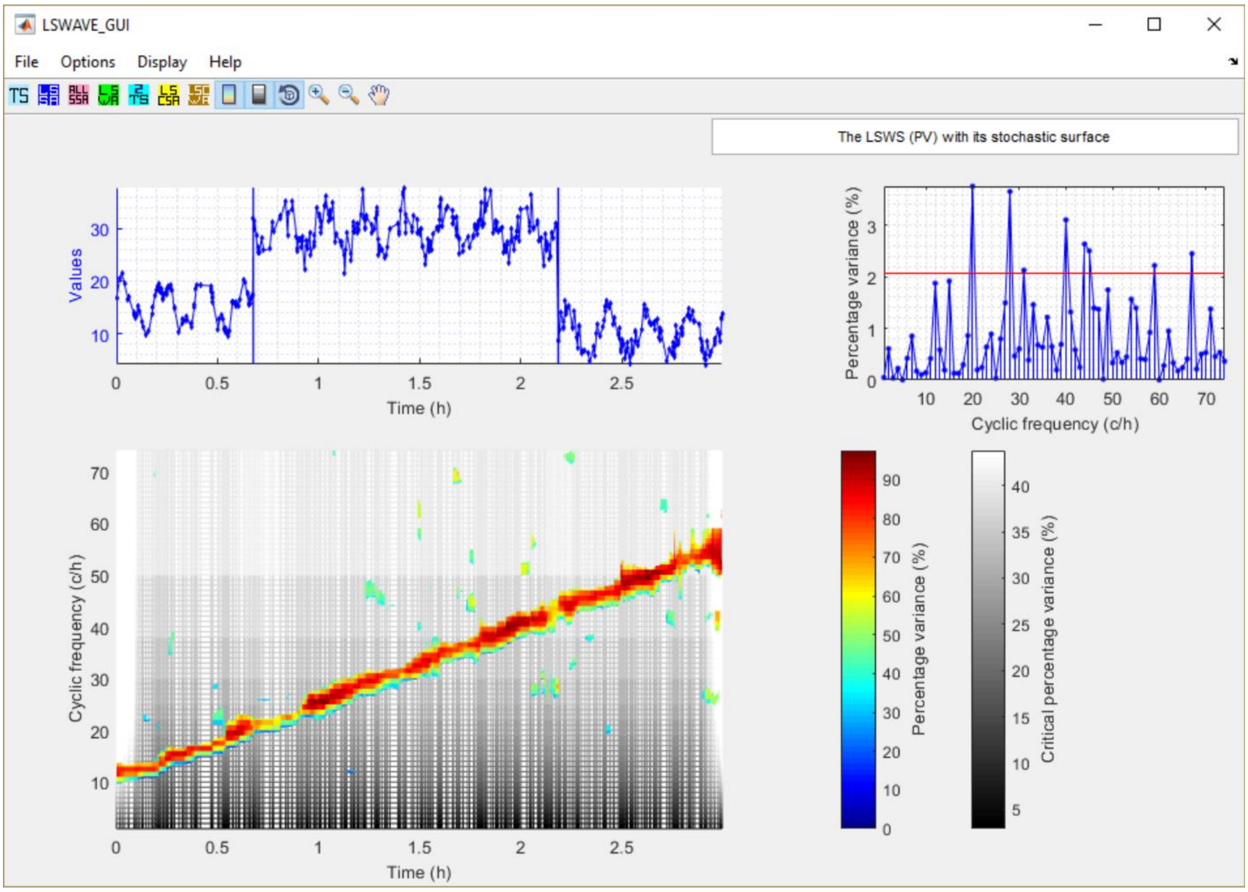

Fig. 3 A GUI screenshot showing the original synthetic time series with two datum shifts shown in the top left panel, and the LSSA and LSWA results of its residual after removing the constituents of known forms shown in the top right and bottom panels, respectively

References

Ghaderpour E (2018) Least-squares wavelet analysis and its applications in geodesy and geophysics. PhD dissertation, York University

Ghaderpour E, Ince ES, Pagiatakis SD (2018a) Least-squares cross-wavelet analysis and its applications in geophysical time series. J Geod 92(10): 1223–1236

Ghaderpour E, Liao W, Lamoureux MP (2018b) Antileakage least-squares spectral analysis for seismic data regularization and random noise attenuation. Geophysics 83(3): V157–V170

Ghaderpour E, Pagiatakis SD (2017) Least-squares wavelet analysis of unequally spaced and non-stationary time series and its applications. Math Geosci 49(7): 819–844

Vaníček P (1969) Approximate spectral analysis by least-squares fit. Astrophys Space Sci 4, 387–391