MATEMATIKA, MJIM, Volume 37, Number 3, 121–134

Analysing Trends and Forecasting of COVID-19 Pandemic in Malaysia using Singular Spectrum Analysis

[1]Nurul A. Othman, [1]Ahmad A. Abdul Aziz, [1]Noor A. Ahmad*, [1]Mohd H. Mohd and [2]Syafiqah I. Mohd Adam

[1]School of Mathematical Sciences, Universiti Sains Malaysia
11800 USM Pulau Pinang, Malaysia

[2]Dept. of Mathematical Sciences, Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor, Malaysia

*Corresponding author: nooratinah@usm.my

Article history
Received: 29 August 2021
Received in revised form: 26 October 2021
Accepted: 8 November 2021
Published online: 15 December 2021

Abstract The Singular Spectrum Analysis (SSA) is a powerful non-parametric time series analysis that has demonstrated its capability in forecasting different time series in various disciplines. SSA falls in the framework of data-driven modelling of dynamical system which does not rely on any underlying assumption except the inherent dynamics which are captured over time. The capabilities of SSA are mainly afforded by its direct connection to the singular value decomposition (SVD). It is generally accepted that SVD-based methods are very affective for the noise reduction in deterministic time series and consequently for forecasting, as well as for extracting trends and structures. Despite its strength, several shortcomings of SSA in the analysis of COVID-19 time series have been reported in the literature. The aim of this paper is to determine the scope of this limitation and we confine our investigation in the analysis and forecasting of COVID-19 Pandemic in Malaysia. We scrutinize the results from the SSA analysis of the number of daily confirmed cases to gain further insight into the intrinsic trends of the pandemic. Groupings of the singular spectra that contributes to different features of the pandemic time series are identified using analysis of the singular value spectrum, periodogram analysis and analysis of the weighted correlation matrix. It was revealed that under stationary conditions, the principal eigentriple is sufficient to produce reliable forecast. However, in non-stationary conditions, for example during a movement control order, it is useful to also study the minor eigentriples which could contain transient dynamics that may persist.

Keywords Singular spectrum analysis; data-driven modelling; COVID-19; Malaysia; machine learning.

Mathematics Subject Classification 37M10, 94A16.

1 Introduction

The ability to predict the future of the COVID-19 pandemic is important to help gain better understanding of the current situation. Many policymakers have relied on mathematical models to guide timely, well-informed responses. Although there are a lot of debate on how much faith one should put on the validity of epidemic forecasting and mathematical models [1–3], it is inevitable that mathematical models shall continue to be one of the main tools to provide insights and possible solutions. Conventional epidemiological models such as the family of susceptible-infected-recovered (SIR) compartmental models [4–6] are among the popular mathematical models used for prediction, mainly due to their sound theoretical basis and a history of useful applications. Epidemiological models are driven by prior assumptions which are translated into a set of assumed parameterized mathematical equations. However, it has been reported that COVID-19 can behave in unexpected ways; for example, asymptomatic cases that can be infectious agents for several weeks [7, 8], the emergence of new variants such as Alpha (B.1.17) (UK), Beta (B.1.351) (South Africa) and Delta (B.1.617.2) (India) with the potential of faster transmission and the possibility of being immune to existing vaccines. These challenges lead to the possibility of unknown dynamics and can limit the ability of conventional models in capturing certain intrinsic trends of the spread.

The Singular Spectrum Analysis (SSA) is a linear approach to analysis and prediction of time series. The data-adaptive nature of the basis functions used in SSA makes it suitable for analysis of some nonlinear dynamics. It was introduced into nonlinear dynamics by Broomhead and King [9] and later by Vautard [10]. The key advantage of SSA is its data-driven nature which does not rely on any prior assumptions except the inherent dynamics which are observed over time. Until recent years, SSA has enabled significant contributions in the study of dynamical system with denoising problems from various fields, for instance fluid dynamics [11], software system [12], climate change [13], mineral processing [14], neuroimmune system [15], image processing [16], and epidemiological studies [17, 18]. In epidemiological studies, SSA has been shown to be useful in analysing rotavirus seasonality by [19]. The recurrent SSA (R-SSA) was used in the predictive modelling of COVID-19 cases in Malaysia [20, 21] and it was noticed in [21] that RF-SSA is unable to detect sudden drop in COVID-19 cases due to change in intervention strategies. The potential of SSA in the forecasting of COVID-19 pandemic times series is also investigated in [22, 23]. Although both studies noted the potential of SSA as a powerful tool for forecasting, it was pointed out in [22] that SSA can produce negative point forecasts which may render the results meaningless.

The shortcomings of SSA in the analysis of COVID-19 time series as reported in [21, 22] is consistent with various evidence found in the literatures (see for example [24, 25]) about the limitations of SSA in separating non-stationary components of a time series. The objective of our paper is to investigate further the scope of the limitation when analysing trends and forecasting of COVID-19 Pandemic in Malaysia. The rest of the paper is organized as follows: In Section 2, basic concepts and methodology of SSA is described. In Section 3 we explore several techniques to determine the statistical dimension of the Malaysian COVID-19 time series. Experiments in the forecasting of future trends of the the time series are presented in Section 4.

2 Materials and Methods

Discrete-time evolution of a dynamical system is specified by discrete maps

$$\mathbf{x}_{n+1} = \mathcal{F}(\mathbf{x}_n, \theta). \tag{1}$$

Here, n denotes the discrete time step and the system is assumed to be sampled every Δt in time such that the nth state vector $\mathbf{x}_n = \mathbf{x}(t_0 + n\Delta t)$. Suppose we have a set of scalar observations $s(t_0 + n\Delta t) = s_n$ of equally sampled data from one of the state variable of the system. Taken's delay embedding theorem [26] states that the geometric structure of the state space dynamics can be reconstructed from vectors of the form:

$$\mathbf{y}_n = [s_n, s_{n+1}, \ldots, s_{n+L-1}]^T \in W,$$

where L is the embedding dimension. It is assumed that the Euclidean L-dimensional space W is related to the original space of the state vector $\mathbf{x}_n$ by smooth, differentiable transformations such that $\mathbf{y}_n$ defines the coordinates of the phase space that will approximate the dynamics of the system from which the time series was sampled. Given the time series $s_1, s_2, \ldots, s_m$ with $K = m - L + 1$, the phase space reconstruction is represented by the matrix of 'snapshots' of the time series $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K\}$ so that the functional form for $\mathcal{F}(\mathbf{x}_n, \theta)$ in the embedding space W is given by

$$\begin{bmatrix} \mathbf{y}_1^T\mathbf{w} \\ \mathbf{y}_2^T\mathbf{w} \\ \vdots \\ \mathbf{y}_K^T\mathbf{w} \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & \cdots & s_L \\ s_2 & s_3 & \cdots & s_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_K & s_{K+1} & \cdots & s_m \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \mathbf{X}\mathbf{w}, \tag{2}$$

for some vector $\mathbf{w} \in R^L$. Matrix $\mathbf{X}$ is the so-called trajectory matrix and it contains the complete record of patterns that have occurred within a window of size L. Notice that the trajectory matrix is a Hankel matrix of size $K \times L$.

2.1 Optimal Basis and Singular Value Decomposition

The phase space reconstruction in (2) assumes the columns of $\mathbf{X}$ as the basis for the embedding space W which may not be the most optimal. Particularly, if L is greater than the statistical dimension of X, we expect some of the columns of $\mathbf{X}$ to be linearly dependent or close to be linearly dependent. For maximum separation of distinctive patterns that have been captured in the trajectory matrix, we seek an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_L\}$ of the embedding space W and this problem can be formulated as a subspace optimization problem of the form

$$\max_{\mathbf{W} \in R^L} Tr(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}), \tag{3}$$

where the optimizer is a matrix $\mathbf{V}$ whose columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_L$ are the eigenvectors of $\mathbf{X}^T\mathbf{X}$ and $Tr(\cdot)$ is the trace function. The scalar value $Tr(\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V})$ gives the sum of the eigenvalues of $\mathbf{X}^T\mathbf{X}$ which is equivalent to the total variance in the time series. Hence by choosing the orthonormal eigenvectors of $\mathbf{X}^T\mathbf{X}$ as the basis for W, the most informative patterns of the time series will be captured in the reconstruction.

The patterns that are contained in the trajectory matrix can be investigated by analysing the patterns that arise from the singular value decomposition (SVD) of $\mathbf{X}$. The SVD is defined by $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where the columns of $\mathbf{U} \in R^{K\times L}$ are the left singular vectors of $\mathbf{X}$ (with $K = m - L + 1$), the columns of $\mathbf{V} \in R^{L\times L}$ are the right singular vectors of $\mathbf{X}$ (which are also the orthonormal eigenvectors of $\mathbf{X}^T\mathbf{X}$) and $\mathbf{\Sigma} \in R^{L\times L}$ is a diagonal matrix whose diagonal entries are the ordered singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_L$ of $\mathbf{X}$. Whenever $\sigma_i \neq 0$, the left singular vectors $\mathbf{u}_i$'s are related to the right singular vectors $\mathbf{v}_i$'s by the equation

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{X}\mathbf{v}_i,$$

for $i = 1, 2, \ldots, L$. By writing the SVD as a sum of rank-one matrices, we may express X as

$$X = X_1 + X_2 + \cdots + X_L, \tag{4}$$

where for each $i = 1, 2, \ldots, L$, $X_i = \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the so-called eigentriple associated with the ith singular value. Given an integer $r \leq L$, the partial sum $\mathbf{X}^{(r)} = \sum_{i=1}^{r} \mathbf{X}_i = \mathbf{U}^{(r)}\mathbf{\Sigma}^{(r)}\mathbf{V}^{(r)T}$ provides the best rank-r approximation to the trajectory matrix $\mathbf{X}$ such that $\|\mathbf{X} - \mathbf{X}^{(r)}\|_F^2$ is minimized among all rank r matrices.

2.2 Singular Spectrum Analysis

The Singular Spectrum Analysis (SSA) is the formal procedural method for analysing a time series via spectral decomposition or SVD. Complete description of the method, can be found in [27, 28]. The technique consists of two stages known as decomposition and reconstruction which are summarized below:

Stage 1. Decomposition

With a choice of embedding dimension L, the trajectory matrix $\mathbf{X}$ is constructed. Next, the SVD of $\mathbf{X}$ is computed. Each eigentriple $\mathbf{X}_i$ is expected to represent a distinctive pattern in the time series.

Stage 2. Reconstruction

The reconstruction stage involves an analysis of the spectrum of singular values in order to identify and differentiate between defining patterns of the time series and noise. The objective is to produce a reconstruction of a less noisy time series which can be used to forecast future data points. An important parameter to be determined at this stage is the statistical dimension r, i.e., the maximum number of eigenvalues/singular values that contribute to the principal subspace (noiseless part) of the time series. Separating the principal subspace from the noise subspace requires some idea about the desired components.There are basically three main components that need to be identified and grouped accordingly [29]; trend, harmonic component and noise. Several strategies can be adopted to determine r, for example, we have the option of analysing the periodogram and the singular spectrum of $\mathbf{X}$. Once r is chosen, the effectiveness of this separability can be assessed using the weighted correlation (w-correlation) statistic. The w-correlation measures the dependence between any two time series and if the separability is sound the two time series will report zero w-correlation. On the other hand, a large w-correlation indicates that the components should be considered as one group. Once the value of r is decided the trajectory matrix can now be treated as the sum $\mathbf{X}^{(r)} + \mathbf{X}^{(L-r)}$ where $\mathbf{X}^{(r)}$ is the principal subspace that contains noise reduced components and $\mathbf{X}^{(L-r)}$ is a subspace which predominantly contains noise.The process of reconstructing $\mathbf{X}^{(r)}$ into a time series is called *diagonal averaging*. Basically this is the process of transforming $\mathbf{X}^{(r)}$ into a Hankel matrix.

3 Statistical Dimension of the Malaysian COVID-19 Pandemic Time Series

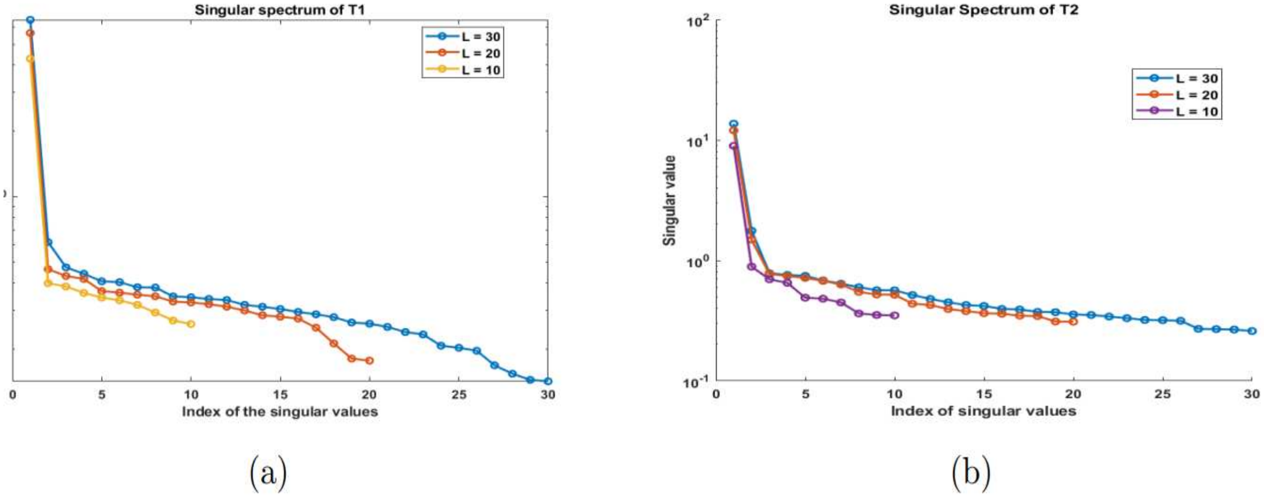
In this section, we demonstrate the process of determining the statistical dimension r of the COVID-19 pandemic time series for Malaysia. The data set used in the experiments are taken from `https://www.worldometers.info/coronavirus/` for the dates 15 February, 2020 to 27 June 27, 2021 and we use the time series for the number of daily cases. To set the stage for the analysis in Section 4, we perform the investigation on two portions of time series:

1. Time series T1 (number of daily cases from 15 February, 2020 to 10 January, 2021): This portion covers the beginning of the pandemic until three days before the second Movement Control Order (MCO2);
2. Time series T2 (number of daily cases from 15 February, 2020 to 19 February, 2021): This portion covers the beginning of the pandemic until the fifth week of the second Movement Control Order (MCO2).

To determine the statistical dimensions for T1 and T2, we analysed i) the spectrum of singular values, ii) the periodogram of the right singular vectors, and iii) the weighted correlation matrices.

3.1 The singular spectrum

Figure 1 depicts the singular value spectrum for T1 and T2 for three different values of the embedding dimension, namely $L = 10$, $L = 20$ and $L = 30$. For both time series we observe the largest difference in magnitudes of the singular values occurs for the first singular value. For T1, the second singular value appears to separate slightly from the smaller singular values when $L = 30$. A similar observation is found in the spectrum for T2. Major trends in a time series are often found in the few leading eigenvalues therefore from the singular value spectrums of T1 and T2, we find choosing $L = 30$ and $r = 2$ will lead to the dominant dynamics in the time series.



(a)

(b)

Figure 1: (a) Singular value spectrum of T1. (b) Singular value spectrum of T2

3.2 Periodogram Analysis

In the periodogram analysis, we compare the frequency content of T1 and T2 and the frequency contents of their right singular eigenvectors. Eigenvectors whose frequencies coincide with the frequencies of the original series indicate the importance of that eigenvector because it most likely contains the same content as in the original time series. Therefore in the reconstruction process, it is necessary to include that eigenvector.

Figure 2 shows the periodogram of T1 and T2, together with the first three principal right singular vectors. We can see from the periodogram of time series T1 in Figure 2(a) that T1 contains mostly low frequency contents and these frequencies, almost all of them, appear in the periodogram of its first right singular vector, $\mathbf{v}_1$. We also see the same low frequencies in the periodogram of $\mathbf{v}_2$. In the periodogram of $\mathbf{v}_3$, the higher frequencies begin to creep in and we are seeing less of the low frequency contents which are present in the periodogram of T1. So Figure 2(a) gives us a good indication of, on one hand, the strong correlation between T1 and the first two right singular vectors, and on the other hand, the clear separation of $\mathbf{v}_3$ from the first two eigenvectors.

In Figure 2(b), the first right singular vector of T2, $\mathbf{v}_1$, is shown to contain most of the low frequencies appearing in the periodogram of T2. The periodogram of $\mathbf{v}_2$ shows quite a different profile compared to the periodogram of $\mathbf{v}_1$, yet still contain the low frequencies appearing in the periodogram of T2. The eigenvector $\mathbf{v}_2$ is seen to contain frequencies which are dominant in $\mathbf{v}_3$. Figure 2(b) indicates that $\mathbf{v}_1$ and $\mathbf{v}_2$ may represent two different dynamics in the time series. The separation between $\mathbf{v}_3$ and the first two eigenvectors is also not as well defined as in Figure 2(a). This observation gives us a hint that T2 may contain non-stationary components that possibly spread out over several different eigentriples.

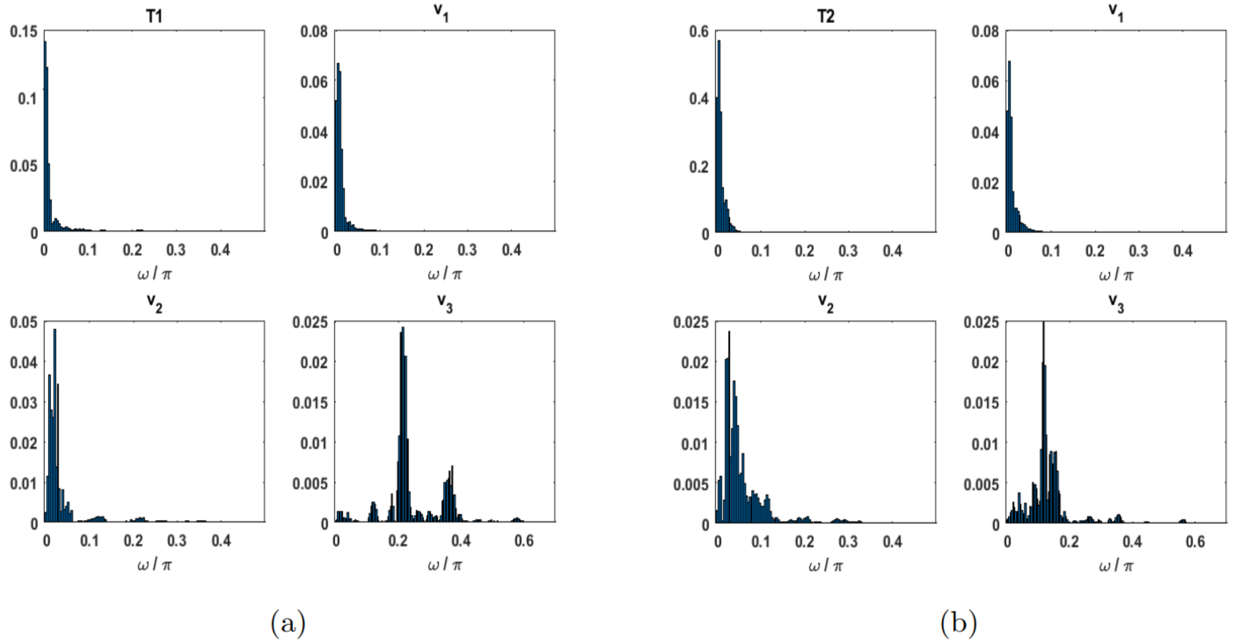


Figure 2: (a) Periodogram of T1 and its first three right singular vectors. (b) Periodogram of T2 and its first three right singular vectors of T2

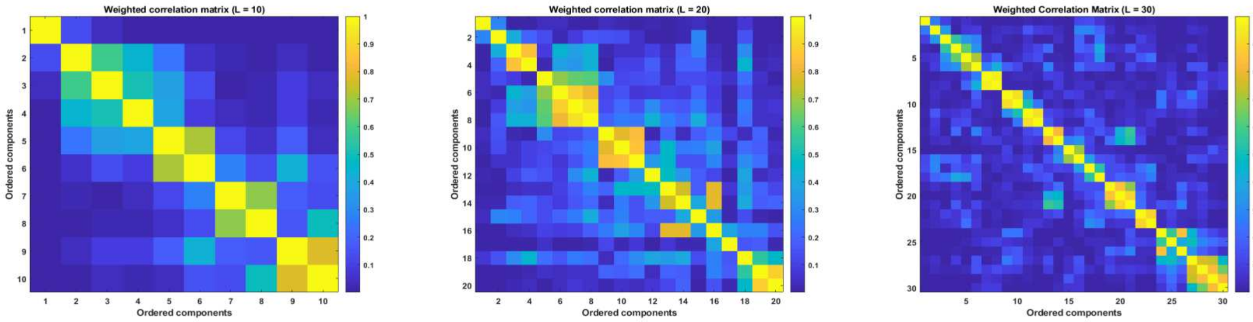
3.3 Weighted Correlation Matrices

Let $\mathbf{Y}_i$ be the time series reconstruction from the ith eigentriple. The measure of separability between $\mathbf{Y}_i$ and $\mathbf{Y}_j$ is given by the value of the weighted correlation (or w-correlation) defined by [28]

$$\rho_{ij}^{(w)} = \frac{(\mathbf{Y}_i, \mathbf{Y}_j)_w}{\|\mathbf{Y}_i\|_w \|\mathbf{Y}_j\|_w},$$

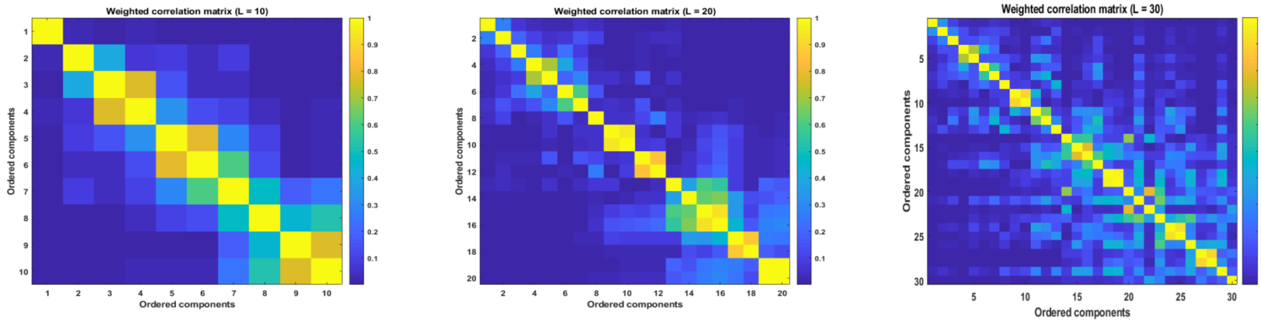
where $\|\mathbf{Y}_i\|_w = \sqrt{(\mathbf{Y}_i, \mathbf{Y}_j)_w}$, $(\mathbf{Y}_i, \mathbf{Y}_j)_w = \sum_{k=1}^{m} w_k y_k^{(i)} y_k^{(j)}$, $w_k = \min\{k, L, m-k\}$ (assuming $L \leq T/2$).

In Figures 3 and 4, heatmaps of the w-correlation matrix for time series T1 and T2 are presented. The yellow region indicate high w-correlation between respective eigentriples and blue region indicate low w-correlation. Low w-correlation implies high separability between the eigentriples. Both Figure 3(a) and Figure 4(a), demonstrate a high separability between the first eigentriple and the rest of the eigentriples, however, the separability between the second eigentriple and the third, fourth and so on is quite poor. As L is increased to 20 and 30, the second eigentriple begins to separate from the minor eigentriples although the separation still appears rather weak. This goes to show that there are certain components in the second eigentriple that is spread out in the minor eigentriples and can be quite tricky to isolate.



(a) (b) (c)

Figure 3: Weighted correlation matrix of T1. (a) $L = 10$, (b) $L = 20$, (c) $L = 30$



(a) (b) (c)

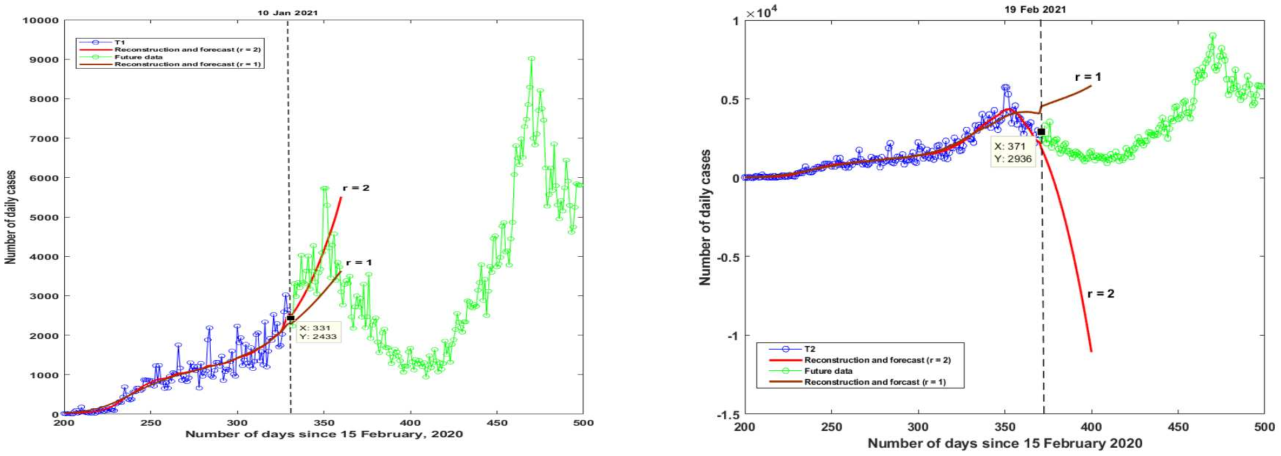
Figure 4: Weighted correlation matrix of T2. (a) $L = 10$, (b) $L = 20$, (c) $L = 30$

4 Forecasting Future Trends of COVID-19 in Malaysia

Time series T1 and T2 are chosen specifically to analyse the trends of COVID-19 pandemic in Malaysia before and after the second nationwide Movement Control Order (MCO2) which was imposed on 13 January 2021. Time series T1 consists of the number of daily cases during which most of travel restrictions in the country have been lifted, whereas the time series T2 covers the five weeks duration of MCO2. Our investigations in this section are performed using $L = 30$ and $r = 1, 2$.

4.1 Forecasting Trends of COVID-19 in Malaysia Before and After MCO2 and Its Limitations

Figure 5(a) shows the reconstructed time series and forecasts of T1 using statistical dimensions $r = 1$ and $r = 2$ respectively.The data used for the reconstruction are marked in blue circles and data marked in green circles are used for testing the accuracy of forecasts. It can be seen that for both $r = 1$ and $r = 2$, the same increasing trend is captured in the forecasts. The forecast with $r = 2$ provides a slightly better accuracy compared to that of $r = 1$. This result agrees well with our periodogram analysis in Section 3 that suggests strong correlation between time series T1 and its two principal eigenvectors. In Figure 5(b) we show the reconstructed time series and forecasts of T2, again using statistical dimensions $r = 1$ and $r = 2$ respectively. The forecasts with $r = 1$ shows a very different trend compared to the forecast with $r = 2$. While the forecasts with $r = 1$ shows an increasing trend, the forecasts with $r = 2$ shows numbers dipping in a downward trend. Most interestingly, the numbers go into the negative region after a short while.



(a) (b)

Figure 5: (a) Forecast one day before MCO2., (b) Forecast three days before MCO2

In order to explain these contrasting observations, we plot the time series reconsructed from the first and second eigentriple of T1 and T2 in Figures 6 and 7 respectively. Notice that in both Figures 6 and 7, the leading eigentriples capture the major trend in their respective time series. The second eigentriple of T1 captures the more recent increase in the number of daily cases which results in an improvement in the forecasting result in Figure 5(a). The second eigentriple

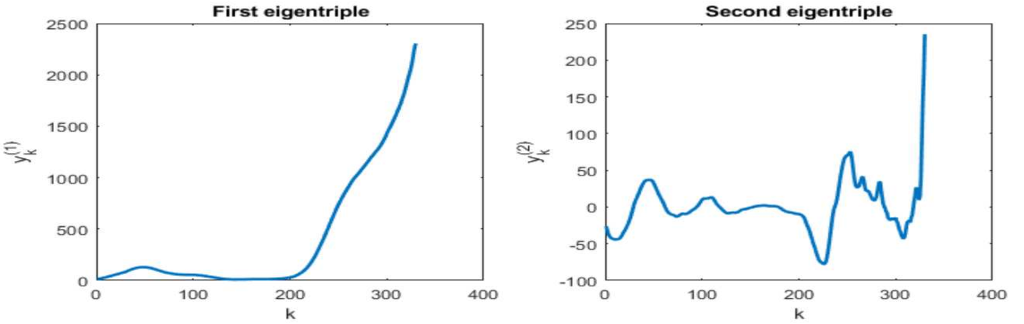


Figure 6: Forecast three days before MCO2

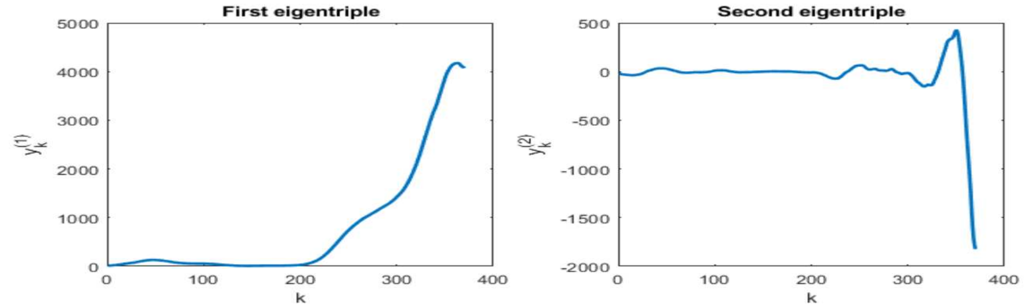


Figure 7: Forecast after 5^{th} week of MCO2

of T2 however, captures the more recent decrease in the number of daily cases which is most likely the cause for the contrasting forecasting result in Figure 5(b) for $r = 1$ and $r = 2$. The change in dynamics due to the intervention strategies introduced during MCO2 is evident in the second eigentriple. However the statistical dimension of the most recent trend cannot be resolved completely which is why the forecasts become negative. We suspect that the recent decreasing trend is a transient dynamic which may or may not persist. To determine whether it will persist, more data is needed.

To obtain a more complete picture on how the trend evolves with additional data, we produce Figure 8. The value m refers to the number of days since 15 February 2020, which is also the number of data used to construct the model for prediction. It can be seen clearly in Figure 8 that, as the value of m increases (i.e. with additional data), the forecast for both $r = 1$ and $r = 2$ begins to flatten. An interesting observation is that as m increase the trend observed for $r = 2$ matches the trend of the forecasts for $r = 1$. The problem of negative forecasts observed earlier for $r = 2$ also slowly diminishes as more data is added.

4.2 Effects of Non-stationarity

Evidence shown in Section 4.1 reveals the limitation of SSA when there exists transient component in the time series. Transient component suggests the presence of non-stationary dynamics in the time series. For non-stationary time series, the following three possible scenarios can affect separability [25]:

1. a signal component may be spread out over several subspaces;
2. mixing or overlap between different signal components;
3. an appearance or vanishing of a component.

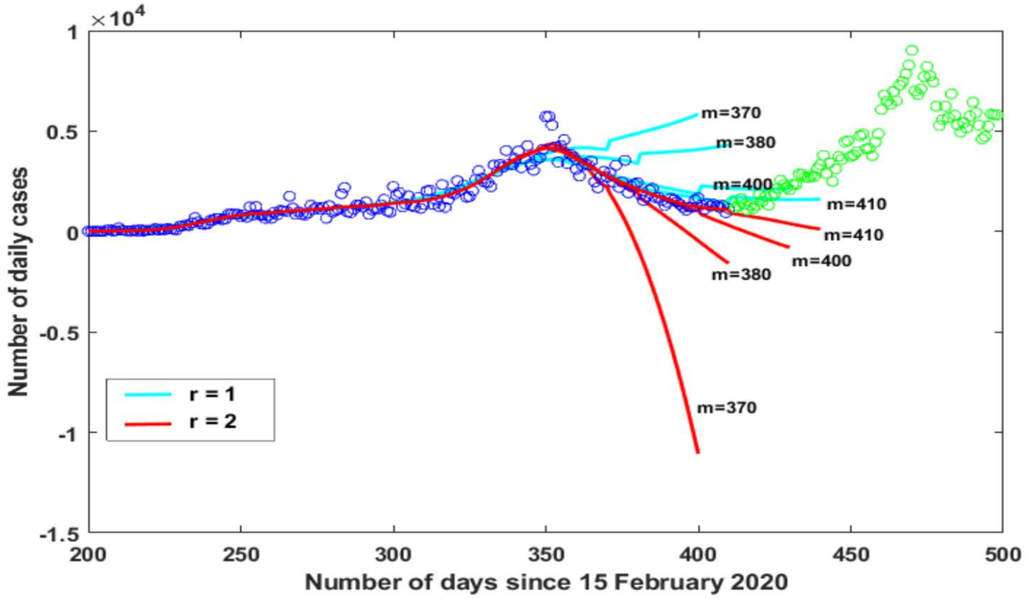


Figure 8: The number of days, $m = 370$ gives the forecast in 5^{th} week of MCO2

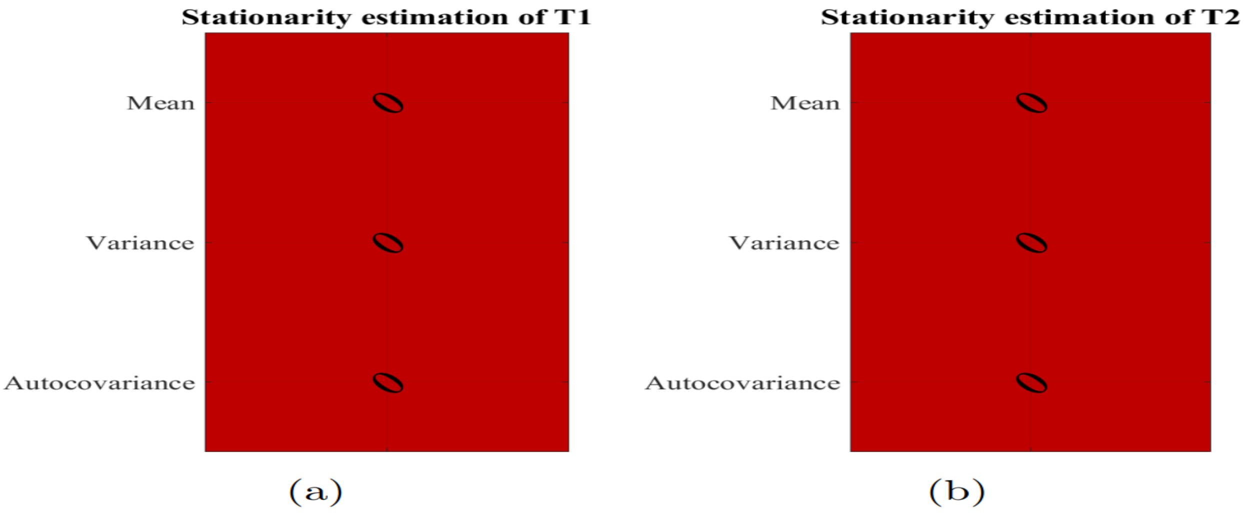
A dynamical system that operates under transient (i.e. non-stationary) conditions (for example, a pandemic under the influence of a movement control order) gives rise to non-stationary time series with time-changing statistics (i.e. having varying mean and variances). We apply a quick test for non-stationarity using methods proposed in [30]. Three statistical properties are checked for stationarity, namely the mean, variance and autocovariance. If these properties are approximately equal or similar, then the initial signal is estimated as stationary. The comparisons are conducted via Wilcoxon rank-sum test for equality of means, BrownForsythe test for equality of variances and similarity check based on Euclidean distances is used to compare autocovariances. The results are shown in Figure 9. For both time series T1 and T2, all three statistical properties are found to be non-stationary. Based on the assumptions of the method, the results also imply the non-stationarity of the two time series.

According to [31], an infinite time series $\{s_1, s_2, s_3, \ldots\}$ is called stationary if, for any integer n there exists a limit

$$\mathbf{T}_n = \lim_{m \to \inf} \mathbf{T}_{m,n}, \tag{5}$$

with $(\mathbf{T}_n)_{ij} = R(i-j), i, j \geq 0$ (the function $R : \mathbb{Z} \to \mathbb{R}$ is called the covariance function), $\mathbf{T}_{m,n} = \frac{1}{m}\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}$ is the trajectory matrix. As such, the eigenvalues and eigenvectors of $\mathbf{T}_n$ are tied up with the singular values and the singular vectors of $\mathbf{X}$. Under non-stationary conditions, each row of the trajectory matrix (which itself is a time series) experience simultaneous drifting [32] and this is registered as correlations of the rows leading to temporal dependence in the variances and autocovariances. From the perspective of the signal subspace, temporal dependence of singular values and singular vectors are observed, which results in transient eigentriples.

The transient dynamics in time series T2 that is captured by SSA is illustrated in Figure 10. The decreasing trend is in fact, part of an oscillatory wave that evolves into a damped oscillation as the time series begins to plateau. When limited data is available, the transient component only captures part of the power spectrum of the signal (time series) while the rest



(a)

(b)

Figure 9: (a) Test for stationarity of time series T1, (b) Test for stationarity of time series T2. The value ’0’ indicates non-stationary

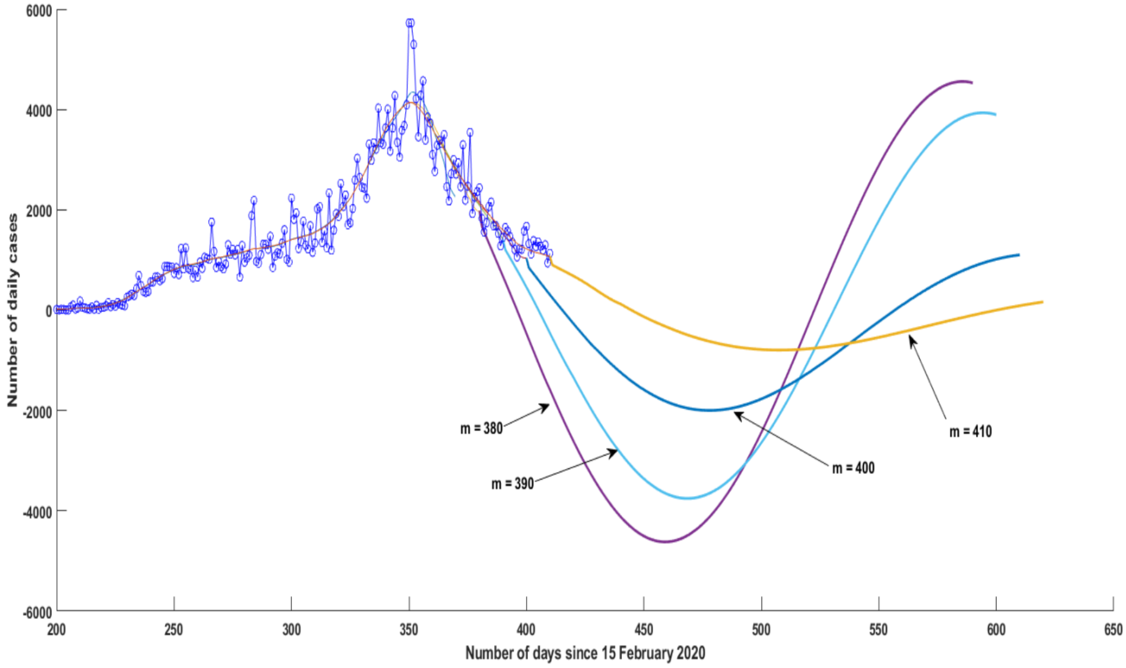


Figure 10: Long-term dynamics exhibited by the transient component of T2

of the spectrum is spread out in the other components. Furthermore, the transient component may assume negative values due to Gibbs effects [33]. As a result, the reconstructed time series can have negative values, rendering the forecasts meaningless. With the availability of more data, the Gibbs effect begins to reduce and so do the negative point forecasts. As we have already seen in Figure 8, more reliable forecast is achieved when similar forecast is obtained for both statistical dimensions $r = 1$ and $r = 2$.

5 Conclusion

Our investigation on the potential of SSA to provide meaningful insights into COVID-19 pandemic in Malaysia has revealed the following:

1. Most of the major trend of the pandemic can be explained in the principle eigentriple. However when the time series is heavily influenced by non-stationary events (for example during MCO), certain limitations were found such as negative point forecasts;
2. The second eigentriple was found to consist some of the most recent trend, however separability is hindered due to the high correlation with other components and the component associated with the recent trend is most likely to have spread out over several components.
3. Temporal dependent of singular values and singular vectors of the trajectory matrix of a non-stationary time series can lead to negative values in the restructured time series and this is mostly due to Gibbs effect. Sufficient data is needed to reduce the Gibbs effect and achieve more reliable forecast that will provide useful insights into future trend;

We have demonstrated the applicability of SSA in the analysis of COVID-19, particularly for the Malaysian scenario. Generalization of our results will require validation using COVID-19 time series from other countries. Just like any other data-driven methods, the quality of results from SSA can only be as reliable as the data itself. The capability of SSA to identify unknown dynamics is quite evident, but good quality data is crucial to realize its true potential.

Acknowledgments

The authors would like to thank the Ministry of Higher Education Malaysia for the financial support under the FRGS grant with Project Code: FRGS/1/2020/ICT06/USM/02/1. We also thank the reviewers for their helpful comments.

References

[1] James, L. P., Salomon, J. A., Buckee, C. O. and Menzies, N. A. The use and misuse of mathematical modeling for infectious disease policymaking: Lessons for the covid-19 pandemic. *Medical Decision Making*. 2021. 41(4): 379–385. doi:10.1177/0272989X21990391.

[2] Ioannidis JPA, T. M., Cripps S. Forecasting for covid-19 has failed. *International Journal of Forecasting*. 2020. ISSN 0169-2070. doi:https://doi.org/10.1016/j.ijforecast.2020.08.004.

[3] Holmdahl, I. and Buckee, C. Wrong but useful what covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine*. 2020. 383(4): 303–305. doi:10.1056/NEJMp2016822.

[4] Kermack, W. O., McKendrick, A. G. and Walker, G. T. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1927. 115(772): 700–721. doi:10.1098/rspa.1927.0118.

[5] Tolles, J. and Luong, T. Modeling Epidemics With Compartmental Models. *JAMA*. 2020. 323(24): 2515–2516.

[6] Khan, A., Zarin, R., Hussain, G., Ahmad, N. A., Mohd, M. H. and Yusuf, A. Stability analysis and optimal control of covid-19 with convex incidence rate in khyber pakhtunkhawa (pakistan). *Results in Physics*. 2021. 20: 103703.

[7] Kronbichler, A., Kresse, D., Yoon, S., Lee, K. H., Effenberger, M. and Shin, J. I. Asymptomatic patients as a source of covid-19 infections: A systematic review and meta-analysis. *International Journal of Infectious Diseases*. 2020. 98: 180–186. doi: 10.1016/j.ijid.2020.06.052.

[8] Peirlinck, M., Linka, K., Sahli Costabal, F., Bhattacharya, J., Bendavid, E., Ioannidis, J. P. and Kuhl, E. Visualizing the invisible: The effect of asymptomatic transmission on the outbreak dynamics of covid-19. *Computer Methods in Applied Mechanics and Engineering*. 2020. 372: 113410. doi:10.1016/j.cma.2020.113410.

[9] Broomhead, D. S. and King, G. P. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*. 1986. 20(2-3): 217–236.

[10] Vautard, R., Yiou, P. and Ghil, M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena*. 1992. 58(1-4): 95–126.

[11] Schmidt, O. T., Towne, A., Rigas, G., Colonius, T. and Brès, G. A. Spectral analysis of jet turbulence. *Journal of Fluid Mechanics*. 2018. 855: 953–982.

[12] Golyandina, N., Korobeynikov, A. and Zhigljavsky, A. *Singular spectrum analysis with R*. Springer. 2018.

[13] Vautard, R. and Ghil, M. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D: Nonlinear Phenomena*. 1989. 35(3): 395–424.

[14] Jemwa, G. T. and Aldrich, C. Classification of process dynamics with monte carlo singular spectrum analysis. *Computers & chemical engineering*. 2006. 30(5): 816–831.

[15] Alharbi, N. A novel approach for noise removal and distinction of eeg recordings. *Biomedical signal processing and control*. 2018. 39: 23–33.

[16] Rodriguez-Aragon, L. J. and Zhigljavsky, A. Singular spectrum analysis for image processing. *Statistics and Its Interface*. 2010. 3(3): 419–426.

[17] Bilancia, M., Stea, G. *et al.* Timescale effect estimation in time-series studies of air pollution and health: A singular spectrum analysis approach. *Electronic Journal of Statistics*. 2008. 2: 432–453.

[18] Ghodsi, M., Hassani, H. and Sanei, S. Extracting fetal heart signal from noisy maternal ecg by singular spectrum analysis. *Journal of Statistics and its Interface, Special Issue on the Application of SSA*. 2010. 3(3): 399–411.

[19] Alsova, O. K., Loktev, V. B. and Naumova, E. N. Rotavirus seasonality: an application of singular spectrum analysis and polyharmonic modeling. *International journal of environmental research and public health*. 2019. 16(22): 4309.

[20] Shaharudin, S. M., Ismail, S., Tan, M. L., Mohamed, N. S. and AininaFilzaSulaiman, N. Predictive modelling of covid-19 cases in malaysia based on recurrent forecasting-singular spectrum analysis approach. *International Journal of Advanced Trends in Computer Science and Engineering*. 2020. 9(1.4 Special Issue).

[21] Shaharudin, S. M., Ismail, S., Samsudin, M. S., Azid, A., Tan, M. L. and Basri, M. A. A. Prediction of epidemic trends in covid-19 with mann-kendall and recurrent forecasting-singular spectrum analysis. *Sains Malaysiana*. 2021. 50(4): 1131–1142.

[22] Kalantari, M. Forecasting covid-19 pandemic using optimal singular spectrum analysis. *Chaos, Solitons & Fractals*. 2021. 142: 110547. doi:10.1016/j.chaos.2020.110547.

[23] Alharbi, N. Predicting covid-19 pandemic in saudi arabia using modified singular spectrum analysis. *medRxiv*. 2020. doi:10.1101/2020.05.24.20111872.

[24] Leles, M., Sanso, J., Mozelli, L. and Guimares, H. A new algorithm in singular spectrum analysis framework:the overlap-ssa (ov-ssa). *SoftwareX*. 2018. 8: 26–32. Digital Signal Processing & SoftwareX - Joint Special Issue on Reproducible Research in Signal Processing.

[25] Harmouche, J., Fourer, D., Auger, F., Borgnat, P. and Flandrin, P. The Sliding Singular Spectrum Analysis: a Data-Driven Non-Stationary Signal Decomposition Tool. *IEEE Transactions on Signal Processing*. 2017. doi:10.1109/TSP.2017.2752720. URL `https://hal.archives-ouvertes.fr/hal-01589464`.

[26] Takens, F. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*. 1981. 898: 366381.

[27] Golyandina, N. and Zhigljavsky, A. *Singular Spectrum Analysis for Time Series*. Springer-Verlag Berlin Heidelberg. 2013.

[28] Sanei, S. and Hassani, H. *Singular Spectrum Analysis of Biomedical Signals (1st ed.)*. CRC Press. 2015.

[29] Hassani, H. and Zhigljavsky, A. Singular spectrum analysis: methodology and application to economics data. *J Syst Sci Complex*. 2009. 22: 372394. doi:10.1007/s11424-009-9171-9.

[30] Zhivomirov, H. and Nedelchev, I. A method for signal stationarity estimation. *Romanian Journal of Acoustics and Vibration*. 2020. 17(2): 149–155.

[31] Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. A. Analysis of time series structure - ssa and related techniques. In *Monographs on statistics and applied probability*. 2001.

[32] Lansangan, J. and Barrios, E. Principal components analysis of nonstationary time series data. *Statistics and Computing*. 2009. 19: 173–187. doi:10.1007/s11222-008-9082-y.

[33] Bozzo, E., Carniel, R. and Fasino, D. Relationship between singular spectrum analysis and fourier analysis: Theory and application to the monitoring of volcanic activity. *Computers & Mathematics with Applications*. 2010. 60(3): 812–820.