A New Signal Processing Approach for Discrimination of EEG Recordings

Hossein Hassani [1,*] [ID] **, Mohammad Reza Yeganegi** [2] **and Emmanuel Sirimal Silva** [3]

[1] Research Institute for Energy Management and Planning of University of Tehran, No. 9, Ghods St., Enghelab St., Tehran 1417466191, Iran

[2] Department of Accounting, Islamic Azad University, Central Tehran Branch, Tehran 1955847781, Iran; m.yeganegi@iauctb.ac.ir

[3] Fashion Business School, London College of Fashion, University of the Arts London, 272 High Holborn, London WC1V 7EY, UK; e.silva@fashion.arts.ac.uk

[*] Correspondence: hassani.stat@gmail.com

Abstract: Classifying brain activities based on electroencephalogram (EEG) signals is one of the important applications of time series discriminant analysis for diagnosing brain disorders. In this paper, we introduce a new method based on the Singular Spectrum Analysis (SSA) technique for classifying brain activity based on EEG signals via an application into a benchmark dataset for epileptic study with five categories, consisting of 100 EEG recordings per category. The results from the SSA based approach are compared with those from discrete wavelet transform before proposing a hybrid SSA and principal component analysis based approach for improving accuracy levels further.

Keywords: EEG; singular spectrum analysis; time series discriminant analysis; kernel; principal component analysis

1. Introduction

Discriminating electroencephalogram, EEG, signals are of great interest in both biology and statistics [1–4]. Identifying a discrimination (classification) rule to classify different sets of EEG recordings can provide a tool to classify brain activities as well as diagnose brain diseases. In statistics, the problem is related to time series discriminant analysis and has been tackled with different approaches [5]. The basic idea in time series discrimination is to use a similarity measure and build a classification rule. Optimality approaches use difference or divergence measures to maximize the difference between the probability distributions or spectral densities corresponding to groups of time series. These approaches consider certain assumptions (model formulation and/or distribution) to make a discrimination function. For instance, [6–8] used differences between spectral densities to discriminate stationary time series. While applying these approaches is straightforward, in time series with time-varying dynamics they do not provide good accuracy [5]. Shumway [9] and Sakiyama and Taniguchi [10] used time varying spectra and the Kullback-Leibler divergence measure to discriminate locally stationary time series (Time series with a slowly time-varying structure are called locally stationary because the small parts of the series show a stationary structure; see [11] and [12] for more details.). Huang et al. [13] used the smooth localized complex exponential model to discriminate non-stationary time series. These techniques have strong theoretical foundations and can be applied in non-stationary time series. However, they are generally too complicated and have their own restrictions [14]. For instance, in methods based on time-varying spectra, the appropriate choice of bandwidth and window length or some parametric model for spectral densities are required [5].

Another class of techniques extract some features from each time series so discrimination can be carried out by classifying the features instead of the time series. Feature extraction techniques are mostly heuristic, that is, in each problem the researcher finds the quantities which best visualize the difference between time series groups. Maharaj and Alonso [15], Fryzlewicz and Ombao [16], Maharaj and Alonso [17] and Yeganegi et al. [14] used different features based on wavelet transform to discriminate time series models. Wavelet transform provides a simple and effective solution to many time series discrimination problems. However, the choice of base function (mother wavelet) could affect the resulting features and consequently the discrimination error rate.

Classifying brain activities based on EEG signals is one of the important applications of time series discriminant analysis for diagnosing brain disorders such as epilepsy. Andrzejak et al. [1] introduced a benchmark dataset for epileptic study. The dataset contains five categories of EEG recordings. Each category consists of 100 EEG recordings from participants. EEG signals have been recorded from brain activities for 23.6 s (4096 observations). Category A contains EEG recordings from healthy subjects while their eyes are open. The second category, category B, contains EEG recordings from healthy subjects with their eyes closed. Categories C and D represent the EEG recordings from epileptic patients. Category D contains EEG signals recorded from the seizure generating area and category C contains signals recorded from the opposite hemisphere of the brain. The last category, category E, consists of EEG signal recordings during seizure activity [1].

Figure 1 shows a typical EEG signal from each category. There are different approaches in the literature for discriminating between these five classes. Nigam and Graupe [18] used a large memory storage and retrieval (LAMSTAR) neural network for classification between category A and E. Kannathal et al. [19] tackled the same problem using four entropy measures and an adaptive neuro-fuzzy system. Guleret et al. [20] used Lyapunov exponents and a recurrent neural network to classify category A, D and E patterns. Maharaj and Alonso [15] used maximum overlap discrete wavelet transform to discriminate between the five categories. Their use of wavelets for feature extraction has become the primary method underlying the majority of subsequent research. Ubeyli [21] used features extracted from a discrete wavelet transform and a mixture of expert classifiers to discriminate category A, D and E patterns. Ghosh-Dastidar et al. [22] extracted EEG features using the mixed-band wavelet-chaos methodology, and applied principal component analysis (PCA) and a cosine radial basis function neural network to classify categories A, C and D. Orhan et al. [23] used features extracted from discrete wavelet transform sub-bands (six sub-bands) and applied k-means clustering and a multilayer perceptron neural network to classify different patterns from category E. Wang et al. [24] used the same approach, although they applied a preprocessing step for selecting the best base in the wavelet transform. Guo et al. [25] used a genetic programming scheme to improve the feature database and applied a k-nearest neighbor analysis to discriminate between categories A, D and E. This research also extracted features from a discrete wavelet transform. Bajaj et al. [26] used time-frequency images of EEG signals and a variety of classifiers to discriminate between category E and other categories' patterns. All previous research, except that of [15], focuses on discrimination between two or three categories at a time (mostly discrimination between epileptic and healthy categories). This research follows the work of [15] and focuses on the five categories problem. The discrimination results (misclassification and accuracy) of [15] are given in Table 1.

Table 1. Classification results for five patterns based on discrete wavelet transform (DWT) (multi-block approach with four blocks) [15].

Predicted Category → Real Category ↓	A	B	C	D	E
A	91	4	1	0	4
B	8	90	0	0	2
C	2	1	86	11	0
D	1	0	13	86	0
E	0	1	0	1	98

The majority of previous research studies extracted the signal's features from wavelet transform or dynamic Fourier transform of the signal. The theory behind extracting features from wavelet and Fourier transforms comes from the fact that different brain activities produce signals with different frequencies over time.
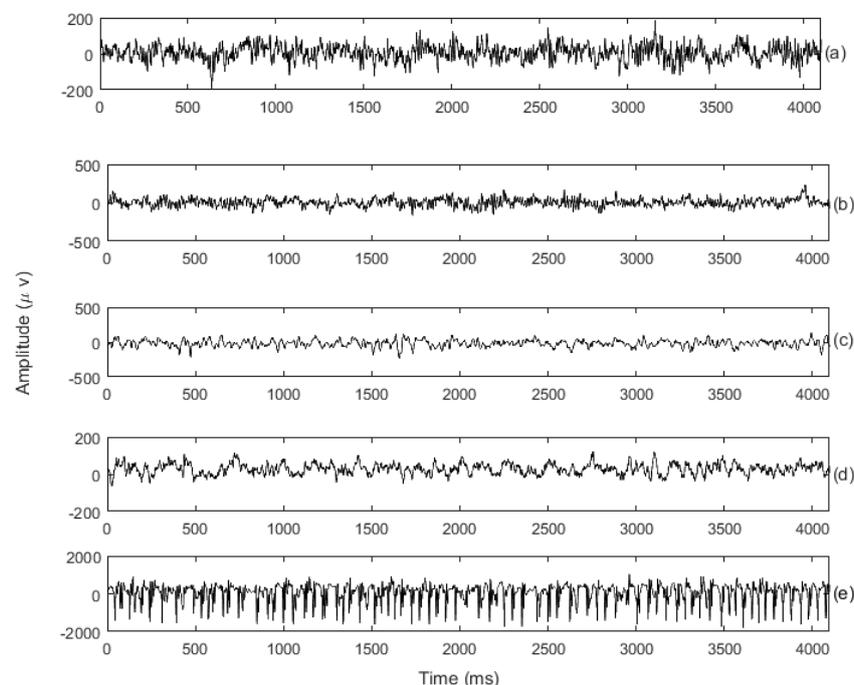
Figure 1. Typical electroencephalogram(EEG) signals: (**a**) EEG recording from healthy subject with eyes open, (**b**) EEG recording from healthy subject with eyes closed, (**c**) EEG recording from non-epileptogenic area of epileptic subject, (**d**) EEG recording from epileptogenic area of epileptic subject, (**e**) EEG recording from seizure activity.

In this paper, we consider introducing a new method based on the singular spectrum analysis (SSA) technique for classifying brain activity based on EEG signals. The proposed method is applied to the Andrzejak et al. dataset [1] to discriminate between five categories. The SSA technique is a nonparametric filtering technique [27–29] with varying applications in forecasting [30,31] signal extraction [32,33] and denoising [34,35]. Given its nonparametric nature, the SSA technique is not restricted by the parametric assumptions of normality, linearity or stationarity [36], which enables the modeling of data without the need for transformations. At the outset, we find it pertinent to point out some additional features of SSA in relation to classical time series analysis techniques and the second machine age. Firstly, the SSA technique has some similarities with the popular autoregressive integrated moving average (ARIMA) technique, as explained in detail in [37] and [38]. Secondly, the emergence of big data has transformed machine learning and deep learning into increasingly popular data analysis techniques. However, as explained

in [39] and [40], the increasing signal-to-noise ratios in big data can make it difficult for machine and deep learning techniques to capture the underlying signal. The filtering capabilities within the SSA technique can be extremely useful under such scenarios to enable machine learning and deep learning techniques to fit better on big data.

The remainder of the paper is organized as follows: Section two introduces the SSA technique; Section 3 presents a statistical summary of the data used in this study; the SSA based discrimination method is introduced in Section 4; Section 5 reports the SSA based extractions; and the paper ends with some conclusions in Section 6.

2. Singular Spectrum Analysis

The SSA technique has its roots in the work by [27,28]. In what follows, we present a concise summary of the SSA process, and in doing so we mainly follow [36] and [29].

The entire SSA process is dependent upon the selection of window length *L* and the number of eigenvalues *r*. A concise discussion around the selection of *L* is presented in the following section. Figure 2 presents a graphical representation of the SSA process, which includes two stages known as decomposition and reconstruction, each with two steps of their own. Consider a one-dimensional time series Y_N of length *N*. In the decomposition stage, *L*, an integer such that $2 \leq L \leq N-1$ is utilized within the embedding step to transfer the one-dimensional time series Y_N into a multidimensional series $X_1, \ldots, X_K$ with vectors

$$X_i = [y_i, \ y_{i+1}, y_{i+2}, \ldots, \ y_{i+L-1}]^{\mathrm{T}}, \tag{1}$$

where $i = 1, 2, \ldots, K$, *T* denotes transposition and $K = N - L + 1$. The output from this embedding step is the trajectory matrix **X**, which is a Hankel matrix.

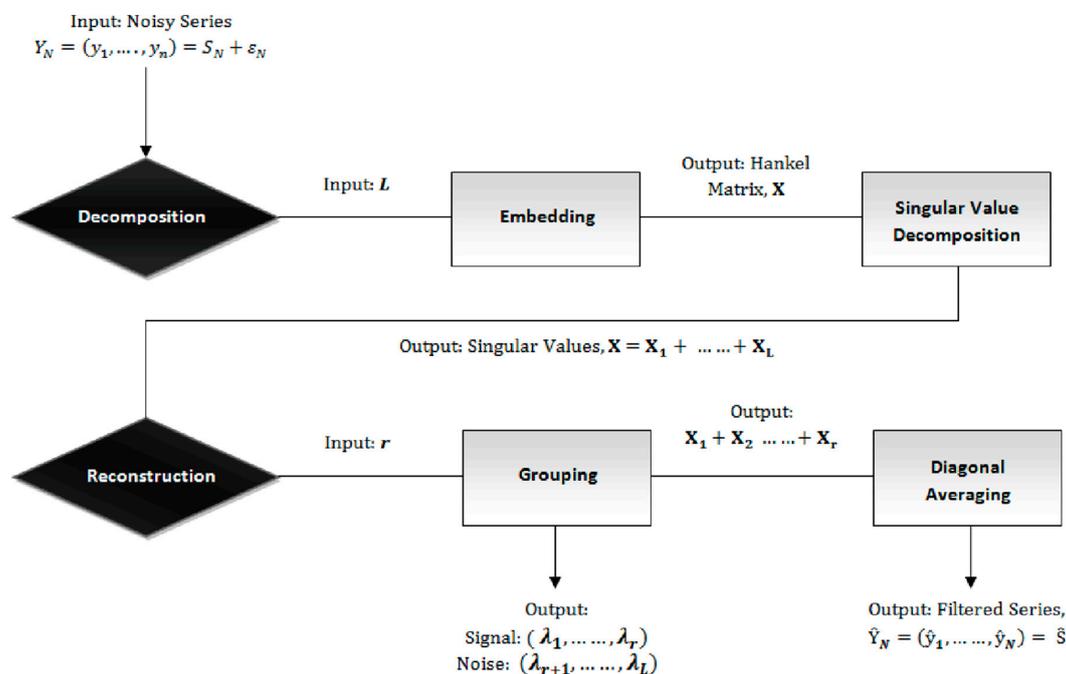**Figure 2.** The basic singular spectrum analysis (SSA) process [36].

$$X = [X_1, \ldots, X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{pmatrix}. \tag{2}$$

The singular value decomposition (SVD) step is purely concerned with the extraction of singular values from the trajectory matrix X. Interestingly, these singular values (eigenvalues) capture all information contained within Y_N. Taking XX^T provides us with positive eigenvalues $\lambda_1, \ldots, \lambda_L$ in decreasing order of magnitude. Then, the SVD of X can be written as

$$X = X_1 + \ldots + X_L, \tag{3}$$

where $X_i = \sqrt{\lambda_i} U_i V_i^T$ ($i = 1, \ldots, d$). Thereafter, the grouping step consists of splitting the elementary matrices into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_p\}$ denote a group of indices $i_1, \ldots, i_p$. Then the matrix X_I corresponding to the group I can be defined as $X_I = X_{i_1} + \cdots + X_{i_p}$. The split of the set of indices $J = 1, \ldots, d$ into the disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation $X = X_{I_1} + \cdots + X_{I_m}$. The procedure of choosing the sets $I_1, \ldots, I_m$ is called theeigentriple grouping. For a given group I, the contribution of the component X_I is measured by the share of the corresponding eigenvalues: $\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$. Finally, in the diagonal averaging step, we transform each matrix I into a time series, which is an additive component of the initial series Y_N. For example, suppose z_{ij} stands for an element of a matrix Z, then the k-th term of the resulting series is obtained by averaging z_{ij} over all i, j, such that $i + j = k + 2$. This procedure is also known as Hankelization of the matrix Z. The output of the Hankelization of matrix Z is the Hankel matrix $\mathcal{H}Z$, which is the trajectory matrix corresponding to the series obtained as aresult of diagonal averaging.

As the gist of this research is focused around the decomposition and signal extraction capabilities of SSA, as opposed to its forecasting capabilities, we do not go in to a discussion of the forecasting phases. Those interested are referred to [36].

3. Features Extracted Based on SSA

As a usual technique in feature extraction, one may use a filter to decompose the original signal into its components. Then, characteristics of components which show the most difference between categories will be extracted as statistical features. For instance, let us consider the SSA based decomposition of a typical EEG recording (as shown in Figure 1). Here, we use SSA with $L = 20$ to decompose each recording. Figures 3 and 4 show first five components in each case.
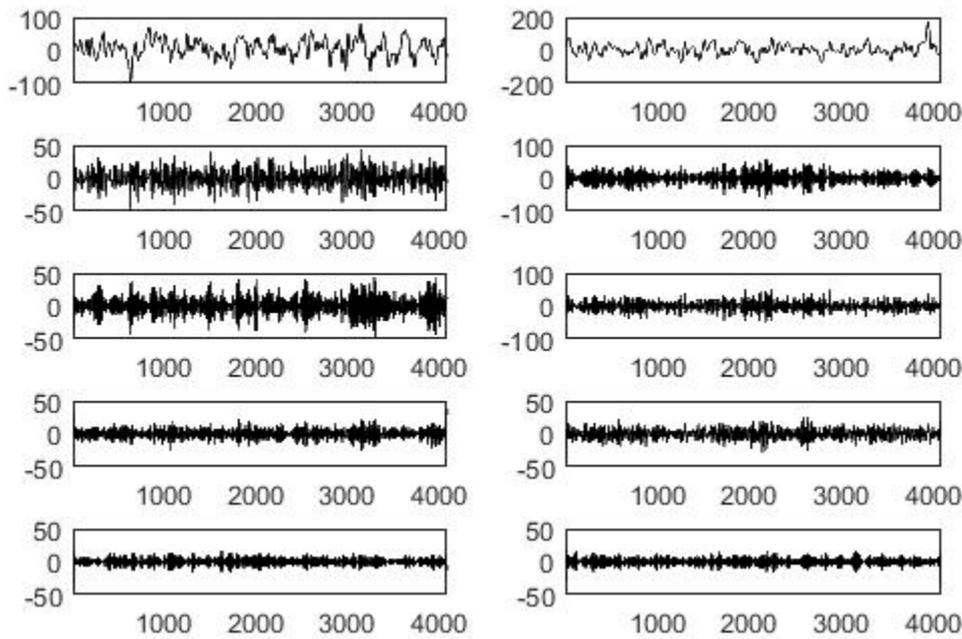
Figure 3. First five components (first component at the top and fifth component at the bottom) of typical EEG recordings from Figure 1: Recordings from categories A (**left**) and B (**right**).
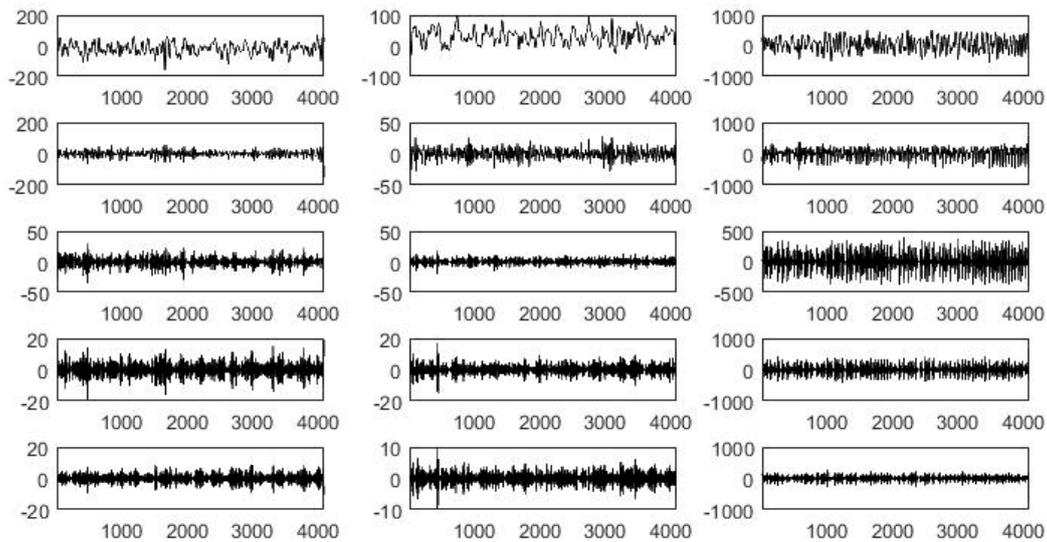
Figure 4. First five components (first component at the top and fifth component at the bottom) of typical EEG recordings from Figure 1: Recordings from categories C (**left**), D (**middle**) and E (**right**).

According to Figures 3 and 4, most differences between the five recordings are related to the amplitude and dispersion of their components. This behavior motivates the use of eigenvalues, variances and minimum and maximum values of components, as features to discriminate between categories.

Accordingly, we consider box plots for the logarithm of the above mentioned values in Figure 5. It can be seen from left to right (from category A to category E) that the logarithm of eigenvalues and variances grow larger (relatively). This shows a clear difference between the logarithm of eigenvalues and variances corresponding to categories.
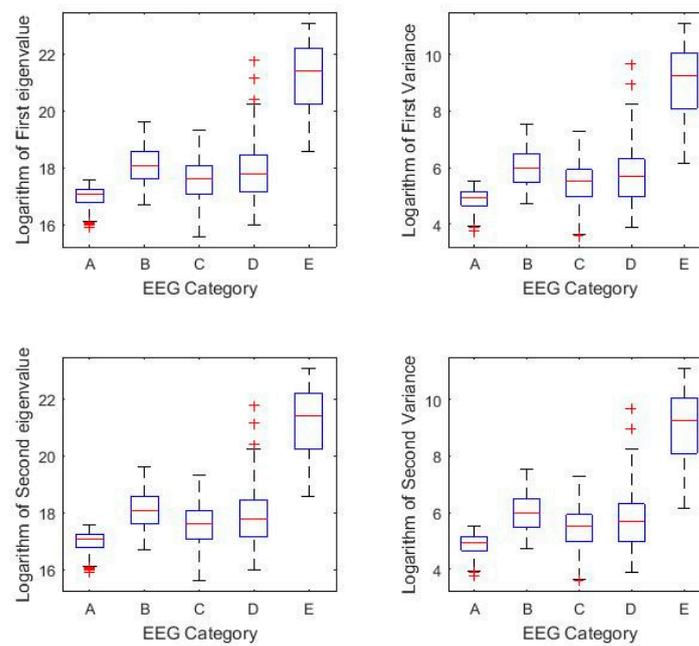
Figure 5. Categories" Box plots for logarithm of eigenvalues and variance of first two components of EEG recordings.

It should be noted that these results are based on SSA with $L = 20$. Although our results can show the difference between these signals, the window length should be chosen carefully. On one hand, a larger window length gives smoother components [36], which may remove the difference between variances. On the other hand, a smaller window length would not lead to a sound separation of noise from signal. There are several approaches for choosing the best L. For example, in [41] and [42], the authors suggested that L is chosen to be equal to the correlation length when the sample autocorrelation function crosses the confidence interval of 95%. [43] consider the selection of L based on the separability between signal and noise, whilst [44] introduced the minimization of a loss function as a means of determining the best L for forecasting purposes. More recently, [45] further investigated the signal extraction problem and constructed a minimum description length criterion that can be employed to identify the dimension (rank) of the signal component. In contrast to the popular binary approach for eigenvalue selection in SSA, [46] introduced a new approach based on Colonial Theory, which appreciates that for more complex time series, a binary approach to reconstruction would not suffice. Ghodsi et al. [47] considered an extensive study of the characteristics underlying 100 real data sets and its influence on the selection of SSA choices when decomposition and reconstruction are based on the [44] algorithm. They found the distribution of data, stationarity, frequencies and series length to be factors enabling differentiation between the best SSA forecasting approaches. Given the importance of the window length to the overall SSA process, we conduct a robustness analysis aimed at determining the effect of window length on eigenvalues. For this purpose, the average eigenvalues are calculated for different window lengths. Figure 6 shows the logarithm of the average eigenvalues corresponding to EEG patterns A and E for different window lengths.
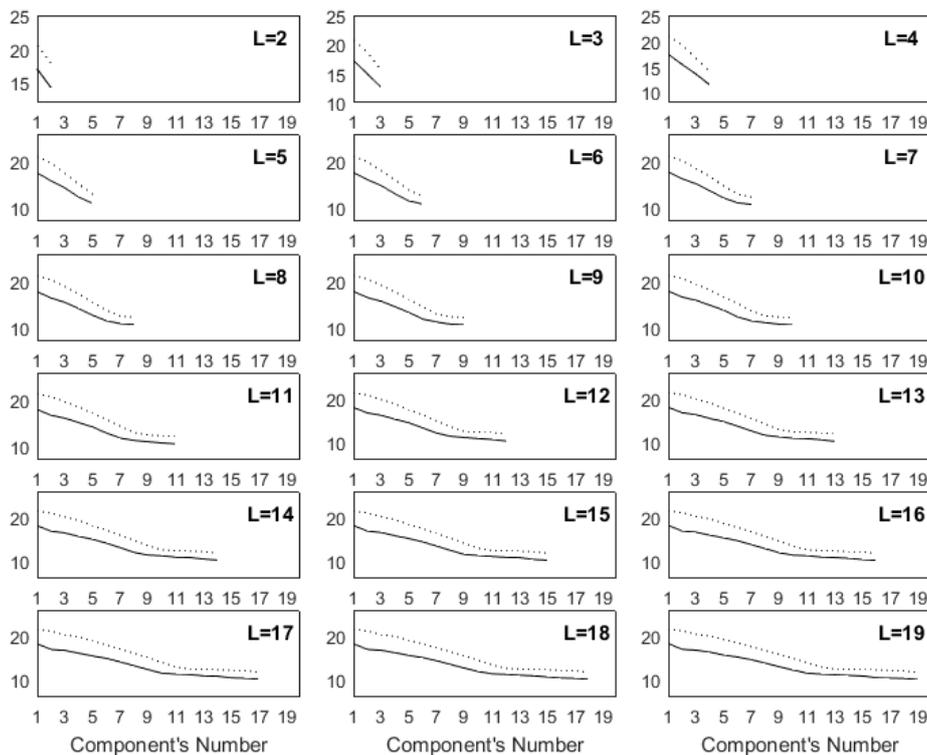
Figure 6. Logarithm of average eigenvalues for EEG patterns A (solid line) and E (dotted line) with different window lengths, *L*: The horizontal axis shows the component's number and the vertical axis shows the logarithm of the average eigenvalues.

According to Figure 6, for all window lengths, the first two components show most differences between the two patterns. Furthermore, as seen here, a larger window length does not lead to a major difference between the first two components. As such, when *L* = 3, SSA removes the noise from two first components and reveals differences between categories.

Given the results discussed above, the following features are then extracted based on SSA with *L* = 3, to discriminate between five categories, i.e., the logarithm of the first two eigenvalues, the logarithm of the first two variances, the logarithm of the maximum value of the first two components and the logarithm of the absolute minimum value of the first two components.

It is noteworthy that all programming and computations were completed in R package [48]. The average time for each EEG feature extraction was 0.035 s, whilst the average time for training the kernel classifier to estimate the category of one EEG recording was 21.74 s.

4. Discrimination between EEG Recordings Based on SSA

The EEG recording discrimination follows these steps:

(1) Use the SSA with window length *L* = 3 and *r* = 2 (without integration) to decompose each EEG recording to two components.

(2) Extract the six features as follows:

Logarithm of the variance of each two components, logarithm of the eigenvalues corresponding to two components, logarithm of maximum value of two components, logarithm of minimum value of two components (The SSA analysis and original feature extraction can be implemented using package "Rssa" in R [48]). The features extracted from five categories of EEG recording are shown in Figure 7.

(3) Use PCA and calculate the six principal components of the features extracted in step 2 (These are the features we use for discrimination).

(4) Estimate the six variate probability density functions for each category (say $f_1(x), \ldots, f_5(x)$, where $x = [x_1, \ldots, x_6]'$) using the kernel method [49] (Since some of the features are not well separated, a nonparametric (kernel) classifier is used to classify the features):

$$\hat{f}_i(x) = \frac{1}{n \det(H)} \sum_{j=1}^{n_i} c\left[1 - u'u\right] I_{u'u<1}$$

where H is bandwidth matrix, c is the some normalizing constant, $u_j^i = H^{-1}\left(x - x_j^i\right)$ and x_j^i are the jth features' vector (corresponding to the jth observation) in the ith category.

In this research the estimation is based on Epanechnikov kernel function, since it gave the better results in this dataset, and the bandwidth is selected using the plug-in bandwidth selection method [50] (The "ks" package in R is used for bandwidth selection and kernel density estimation).

(5) As the discrimination rule, a new set of features (say $x_1, \ldots, x_6$) is allocated to the ith category if $\arg \max_{i=1, \ldots, 5} \left\{\hat{f}_i(x)\right\}$.
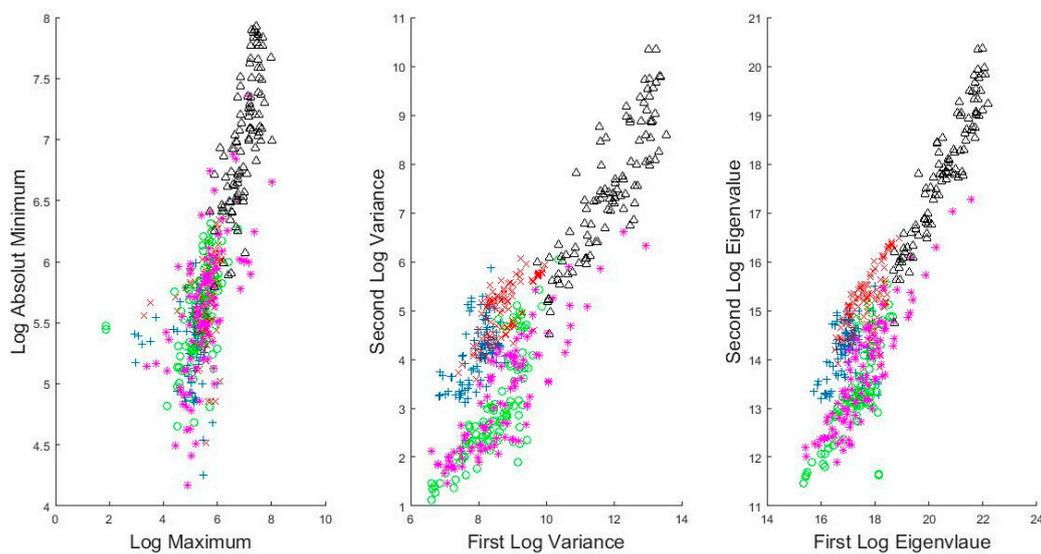

Figure 7. Features extracted from categories A (+), B ($\times$), C (o), D (*) and E (Δ).
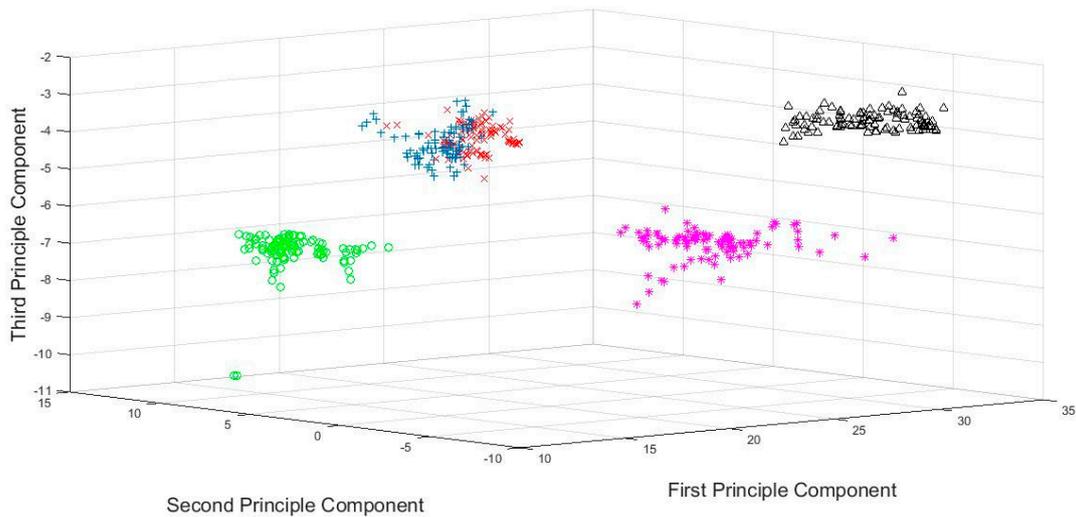
5. Results

The results of discrimination (number of misclassifications) between patterns A, B, C, D and E (before applying PCA) are given in Table 2. According to these results, using features extracted from SSA and the kernel classifier, there are only 3 misclassifications, which makes the total error rate 0. 6% (99.4% accuracy).

Table 2. Classification results for five patterns based on SSA, before applying principal components analysis (PCA). The diagonal numbers show the number of correct classifications in each category.

Predicted Category→ Real Category↓	A	B	C	D	E
A	100	0	0	0	0
B	0	100	0	0	0
C	0	0	100	0	0
D	2	0	1	97	0
E	0	0	0	0	100

To further improve the performance of the proposed method, one may use principal components analysis (PCA) to reduce the dependence between features. Figures 8 and 9 show the first and last three principal components of features. It is evident that most categories are well separated. Using all six components of PCA gives 100% accuracy. Applying PCA will increase the computational cost of recognition by $O(n^2)$ (Applying PCA, in estimating the category of the given observation, the original vector of features will be multiplied with an n by n coefficient matrix) where *n* is the number of features). Since the proposed procedure involves only six features, the additional computational cost of recognition regarding applying PCA is relatively low. On the other hand, before applying PCA, the procedure classifies some of the recordings from epileptogenic area of epileptic subject as healthy EEG recordings.

The accuracy of the presented procedure in discriminating observations out of the training dataset is investigated using the leave-one-out method. Table 3 shows the leave-one-out accuracy and misclassification results for discriminating between five categories using six components (from PCA) and the kernel classifier. According to the results, the leave-one-out misclassification rate after applying PCA is 2.8%, which gives a total accuracy of 97.2%.

Figure 8. First three principal components of features extracted from categories A (+), B (×), C (o), D (*) and E (Δ).
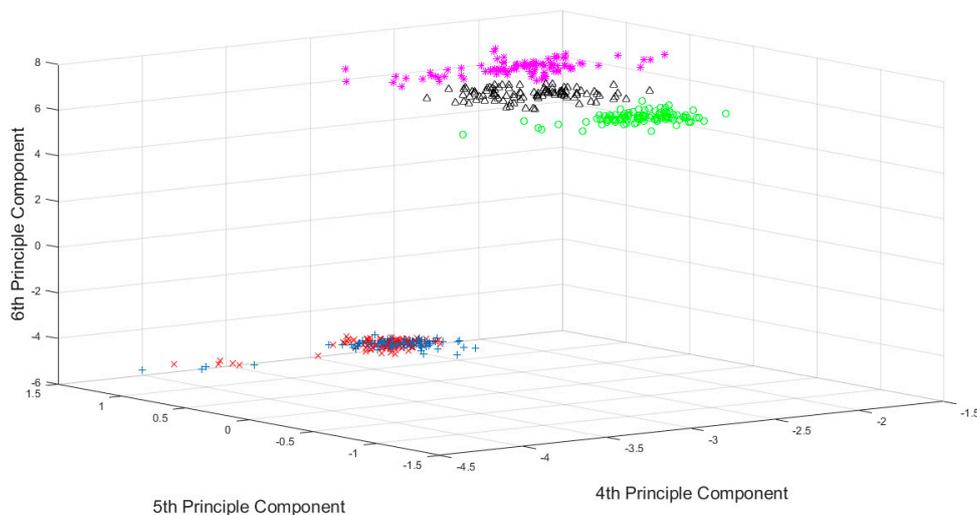
Figure 9. First three principal components of features extracted from categories A (+), B (×), C (o), D (*) and E (Δ).

Table 3. Leave-one-out classification results for five patterns based on SSA and PCA. The diagonal numbers show the number of correct classifications in each category.

Predicted Category → Real Category ↓	A	B	C	D	E
A	91	9	0	0	0
B	5	95	0	0	0
C	0	0	100	0	0
D	0	0	0	100	0
E	0	0	0	0	100

Maharaj and Alonso [15] used discrete wavelet transform(DWT) to discriminate between these patterns. The best accuracy achieved using DWT is reported in Table 1. According to Table 1, discrimination using DWT and a kernel classifier has a 90.2% total accuracy. The comparison of SSA and DWT to extract EEG features clearly indicates the superiority of the SSA-based extractions, which are far more accurate than DWT-based extractions.

6. Conclusions

This paper begins with the aim of introducing a new and more effective approach for discriminating EEG signals. Accordingly, we propose the use of the nonparametric singular spectrum analysis technique as a viable tool and evaluate its performance at discriminating EEG signals using the [1] benchmark data set for epileptic study. We find promising results which show that highly accurate feature extractions are possible via SSA and a kernel classifier. Motivated by this, we innovate further by incorporating PCA within the proposed classification model to make additional gains in classification accuracy.

As a result, we have succeeded in presenting several contributions. Firstly, previous researchers working on this same data set (see Introduction) have relied upon the wavelet transform and the Fourier transform for feature extraction. In contrast, we have succeeded in proposing and introducing a successful SSA and kernel-based approach for discriminating between EEG signals.

Secondly, it is important to note that the primary challenge in this dataset is our attempt to classify all five categories. For instance, categories A and B are very similar, as both are EEG recordings of healthy subjects, with their eyes are open and closed, respectively. Moreover, categories C and D are also similar, since both are EEG recordings from epileptic patients, taken from the epileptic parts of the brain and the other from the normal parts of the brain, when they are not under attack. In contrast, the last category is completely different as it is recorded during an epileptic attack. Interestingly, most of the previous research focused on discrimination between categories A and E, or C and E, or D and E, or A and D, or A and C categories in pairs. Whilst [15] considered discriminating all categories at the same time, their findings were not satisfactory and thus forced them to switch to categories A and E, which represents the simplest case for discriminating between EEG recordings. The fact that our findings, using the spectral method of SSA, produced betters results by discriminating between all five categories at the same time is extremely important and useful in practice.

Thirdly, the initial results were extremely positive, as the SSA combined with a kernel classifier reported a 99.4% accuracy rate. We also show that incorporating PCA within the proposed methodology can further improve the classification accuracy, up to 100% (97.2% for out-of-sample classification). These results are far better than the accuracy rate reported in [15], where DWT was used to discriminate between EEG signals. Whilst historically the DWT technique was popular for discriminating between EEG signals, the findings here not only question its applicability in future, but also clearly demonstrate that the SSA and kernel-based approach for classification has comparatively better features and accuracy rates.

Finally, this paper also opens several research avenues related to EEG discrimination and (DWT) provides researchers and practitioners with a new tool for consideration as a solution to other classification problems within the medical industry. For example, since our algorithm selects SSA choices (window length and number of eigenvalues) to gain the maximum difference between categories, it has the potential to be exploited for automatic discrimination between EEG recordings.

References

1. Andrzejak, R.G.; Lehnertz, K.; Rieke, C.; Mormann, F.; David, P.; Elger, C.E. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **2001**, *64*, 061907. [CrossRef] [PubMed]
2. McBride, J.; Zhao, X.; Munro, N.; Smith, C.; Jicha, G.; Jiang, Y. Resting EEG Discrimination of Early Stage Alzheimer's Disease from Normal Aging Using Inter-Channel Coherence Network Graphs. *Ann. Biomed. Eng.* **2013**, *41*, 1233–1242. [CrossRef] [PubMed]
3. Lana, E.P.; Adorno, B.V.; Tierra-Criollo, C.J. Detection of movement intention using EEG in a human-robot interaction environment. *Res. Biomed. Eng.* **2015**, *31*. [CrossRef]
4. Panagiotis, C.; Hadjileontiadis, L.J. Emotion Recognition from EEG Using Higher Order Crossing. *IEEE Trans. Inform. Technol. Biomed.* **2010**, *14*, 186–197. [CrossRef]
5. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications, with R Examples*; Springer: New York, NY, USA, 2011.
6. Dargahi-Noubary, G.R. Discrimination between Gaussian time series based on their spectral differences. *Commun. Stat. Theory Methods* **1992**, *21*, 2439–2458. [CrossRef]
7. Shumway, R.H.; Unger, A.N. Linear discriminant function for stationary time series. *J. Am. Stat. Assoc.* **1974**, *69*, 948–956. [CrossRef]
8. Kakizawa, Y.; Shumway, R.; Taniguchi, M. Discrimination and clustering for multivariate time series. *J. Am. Stat. Assoc.* **1998**, *93*, 328–340. [CrossRef]
9. Shumway, R.H. Time-frequency clustering and discriminant analysis. *Stat. Probab. Lett.* **2003**, *63*, 307–314. [CrossRef]
10. Sakiyama, K.; Taniguchi, M. Discriminant Analysis for locally stationary processes. *J. Multivar. Anal.* **2004**, *90*, 282–300. [CrossRef]
11. Dahlhaus, R. Fitting time series models to non-stationary processes. *Ann. Stat.* **1997**, *25*, 1–37. [CrossRef]
12. Nason, G.P.; von Sachs, R.; Kroisandt, G. Wavelet processes and adaptive estimation of evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B* **2000**, *62*, 271–292. [CrossRef]
13. Huang, H.S.; Ombao, H.; Stoffer, D. Discrimination and classification of non-stationary time series using the SLEX model. *J. Am. Stat. Assoc.* **2004**, *99*, 763–774. [CrossRef]
14. Yegangi, M.R.; Mansouri, B.; Chinipardaz, R. Discrimination and clustering of earthquake and explosion based on NDWT. *Commun. Stat. Case Stud. Data Anal. Appl.* **2015**, *1*, 1–8. [CrossRef]
15. Maharaj, E.A.; Alonso, A.M. Discrimination of locally stationary time series using wavelets. *Comput. Stat. Data Anal.* **2007**, *52*, 879–895. [CrossRef]
16. Fryzlewicz, P.; Ombao, H. Consistent classification of nonstationary time series using stochastic wavelet representations. *J. Am. Stat. Assoc.* **2009**, *104*, 299–312. [CrossRef]
17. Maharaj, E.A.; Alonso, A.M. Discrimination analysis of multivariate time series: Application to diagnosis based on ECG signals. *Comput. Stat. Data Anal.* **2014**, *70*, 67–87. [CrossRef]
18. Nigam, V.P.; Graupe, D. A neural-network-based detection of epilepsy. *Neurol. Res.* **2004**, *26*, 55–60. [CrossRef] [PubMed]
19. Kannathal, N.; Choo, M.L.; Acharya, U.R.; Sadasivan, P.K. Entropies in the detection of epilepsy in EEG. *Comput. Methods Programs Biomed.* **2005**, *80*, 187–194. [CrossRef] [PubMed]
20. Guler, N.F.; Ubeyli, E.D.; Guler, I. Recurrent neural networks employing Lyapunov exponents in EEG recordings. *Expert Syst. Appl.* **2005**, *29*, 506–514. [CrossRef]

21. Ubeyli, E.D. Wavelet/mixture of experts network structure for EEG classification. *Expert Syst. Appl.* **2008**, *37*, 1954. [CrossRef]

22. Ghosh-Dastidar, S.; Adeli, H.; Dadmehr, N. Principal component analysis enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 512–518. [CrossRef] [PubMed]

23. Orhan, U.; Hekim, M.; Ozer, M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst. Appl.* **2011**, *38*, 13475. [CrossRef]

24. Wang, D.; Miao, D.; Xie, C. Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Syst. Appl.* **2011**, *38*, 14314–14320. [CrossRef]

25. Guo, L.; Rivero, D.; Dorado, J.; Munteanu, C.R.; Pazos, A. Automatic feature extraction using genetic programming: An application to epileptic EEG classification. *Expert Syst. Appl.* **2011**, *38*, 1042. [CrossRef]

26. Bajaj, V.; Rai, K.; Kumar, A.; Sharma, D. Time-frequency image-based features for classification of epileptic seizures from EEG signals. *Biomed. Phys. Eng. Express* **2017**, *3*, 015012. [CrossRef]

27. Broomhead, D.S.; King, G.P. Extracting Qualitative Dynamics from Experimental Data. *Phys. D Nonlinear Phenom.* **1986**, *20*, 217–236. [CrossRef]

28. Broomhead, D.S.; King, G.P. On the Qualitative Analysis of Experimental Dynamical Systems. In *Nonlinear Phenomena and Chaos*; Sarkar, S., Ed.; Adam Hilger: Bristol, UK, 1986.

29. Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A. *Analysis of Time Series Structure: SSA and Related Techniques*; Chapman and Hall/CRC: New York, NY, USA; London, UK, 2001.

30. Silva, E.S.; Hassani, H.; Heravi, S. Modeling European industrial production with multivariate singular spectrum analysis: A cross-industry analysis. *J. Forecast.* **2018**, *37*, 371–384. [CrossRef]

31. Hassani, H.; Silva, E.S.; Antonakakis, N.; Filis, G.; Gupta, R. Forecasting accuracy evaluation of tourist arrivals. *Ann. Tour. Res.* **2017**, *63*, 112–127. [CrossRef]

32. Hassani, H.; Silva, E.S.; Ghodsi, Z. Optimizing bicoid signal extraction. *Math. Biosci.* **2017**, *294*, 46–56. [CrossRef] [PubMed]

33. Ghodsi, Z.; Silva, E.S.; Hassani, H. Bicoid signal extraction with a selection of parametric and nonparametric signal processing techniques. *Genom. Proteom. Bioinform.* **2015**, *13*, 183–191. [CrossRef] [PubMed]

34. Hassani, H.; Mahmoudvand, R.; Zokaei, M.; Ghodsi, M. On the separability between signal and noise in singular spectrum analysis. *Fluct. Noise Lett.* **2012**, *11*, 1250014. [CrossRef]

35. Hassani, H.; Dionisio, A.; Ghodsi, M. The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. *Nonlinear Anal. Real World Appl.* **2010**, *11*, 492–502. [CrossRef]

36. Sanei, S.; Hassani, H. *Singular Spectrum Analysis of Biomedical Signals*; CRC Press: London, UK, 2015.

37. Hassani, H.; Thomakos, D. A review on singular spectrum analysis for economic and financial time series. *Stat. Its Interface* **2010**, *3*, 377–397. [CrossRef]

38. Silva, E.S.; Hassani, H. On the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recession. *Int. Econ.* **2015**, *141*, 34–49. [CrossRef]

39. Hassani, H.; Silva, E.S. Forecasting with Big Data: A Review. *Ann. Data Sci.* **2015**, *2*, 5–19. [CrossRef]

40. Hassani, H.; Huang, X.; Silva, E.S. Banking with blockchain-ed big data. *J. Manag. Anal.* **2018**. [CrossRef]

41. De Oliviera, M.A.; Filho, J.V.; Lopes, V.; Inman, D.J. A new approach for structural damage detection exploring the singular spectrum analysis. *J. Intell. Mater. Syst. Struct.* **2016**, *28*, 1160–1174. [CrossRef]

42. Loh, C.H.; Chen, C.H.; Hsu, T.Y. Application of advanced statistical methods for extracting long-term trends in static monitoring data from an arch dam. *Struct. Health Monit.* **2011**, *10*, 587–601. [CrossRef]

43. Khan, M.A.R.; Poskitt, D.S. A note on window length selection in Singular Spectrum Analysis. *Aust. N. Z. J. Stat.* **2013**, *55*, 87–108. [CrossRef]

44. Hassani, H.; Webster, A.; Silva, E.S.; Heravi, S. Forecasting U.S. tourist arrivals using optimal singular spectrum analysis. *Tour. Manag.* **2015**, *46*, 322–335. [CrossRef]

45. Khan, M.A.R.; Poskitt, D.S. Signal Identification in Singular Spectrum Analysis. *Aust. N. Z. J. Stat.* **2016**, *58*, 71–98. [CrossRef]

46. Hassani, H.; Ghodsi, Z.; Silva, E.S.; Heravi, S. From nature to maths: Improving forecasting performance in subspace-based methods using genetics Colonial Theory. *Digit. Signal Process.* **2016**, *51*, 101–109. [CrossRef]

47. Ghodsi, M.; Hassani, H.; Rahmani, D.; Silva, E.S. Vector and recurrent singular spectrum analysis: Which is better at forecasting? *J. Appl. Stat.* **2018**, *45*, 1872–1899. [CrossRef]

48. Hassani, H.; Mahmoudvand, R. *Singular Spectrum Analysis, Using R*; Palgrave Macmillan: London, UK, 2018.

49. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986.
50. Chacon, J.E.; Duong, T. Multivariate plug-in bandwidth selection with unconstrained pilot matrices. *Test* **2010**, *19*, 375–398. [CrossRef]