

Article

Information Criteria for Signal Extraction Using Singular Spectrum Analysis: White and Red Noise

Nina Golyandina *  and Nikita Zvonarev 

Faculty of Mathematics and Mechanics, St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg 199034, Russia; nikitazvonarev@gmail.com

* Correspondence: n.golyandina@spbu.ru

Abstract: In singular spectrum analysis, which is applied to signal extraction, it is of critical importance to select the number of components correctly in order to accurately estimate the signal. In the case of a low-rank signal, there is a challenge in estimating the signal rank, which is equivalent to selecting the model order. Information criteria are commonly employed to address these issues. However, singular spectrum analysis is not aimed at the exact low-rank approximation of the signal. This makes it an adaptive, fast, and flexible approach. Conventional information criteria are not directly applicable in this context. The paper examines both subspace-based and information criteria, proposing modifications suited to the Hankel structure of trajectory matrices employed in singular spectrum analysis. These modifications are initially developed for white noise, and a version for red noise is also proposed. In the numerical comparisons, a number of scenarios are considered, including the case of signals that are approximated by low-rank signals. This is the most similar to the case of real-world time series. The criteria are compared with each other and with the optimal rank choice that minimizes the signal estimation error. The results of numerical experiments demonstrate that for low-rank signals and noise levels within a region of stable rank detection, the proposed modifications yield accurate estimates of the optimal rank for both white and red noise cases. The method that considers the Hankel structure of the trajectory matrices appears to be a superior approach in many instances. Reasonable model orders are obtained for real-world time series. It is recommended that a transformation be applied to stabilize the variance before estimating the rank.

Keywords: time series; singular spectrum analysis; model order selection; low rank approximation; information criterion



Citation: Golyandina, N.; Zvonarev, N. Information Criteria for Signal Extraction Using Singular Spectrum Analysis: White and Red Noise. *Algorithms* **2024**, *17*, 395. <https://doi.org/10.3390/a17090395>

Academic Editor: Vittorio Maniezzo

Received: 12 August 2024

Revised: 26 August 2024

Accepted: 2 September 2024

Published: 5 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Singular spectrum analysis (SSA), which is closely related to signal-subspace methods (cf. [1–5] and reviews [6,7]), has been increasingly used in recent decades for practical tasks, including preprocessing and feature extraction as part of hybrid machine learning methods [8–12]. The attractive feature of the SSA method is that it does not require specifying a time series model.

SSA is capable of addressing a wide range of problems in time series analysis, including the application of low-frequency filters for smoothing, the extraction of signals, the estimation of frequencies, the filling of gaps, and the forecasting; all but the former are based on signal subspace estimation (Golyandina, 2020 [7]). The signal is understood to be a non-random component of the time series, which may include a trend and oscillations. The SSA algorithm consists of embedding the time series into a sequence of vectors of size L , collecting them into a matrix, decomposing this matrix into elementary matrices, grouping these matrices in a sophisticated way, and then obtaining an interpreted decomposition of the original time series into a sum of interpreted components. In order to act as a low-frequency filter, the number of components is determined based on the frequency characteristics of the components in question. In the case of the majority of other problems,

signal estimation is required. Signal estimation is performed by grouping the leading r components. As a result, it is of great importance to know r , which is referred to as the signal rank or signal model order. The following section will describe the SSA algorithm for solving a particular signal estimation problem.

SSA

Let us briefly describe the SSA algorithm for signal extraction from a time series $X = (x_1, \dots, x_N)$ of length N , following [7]. We assume that

$$X = S + N,$$

where S is a signal, and N is random noise with zero expectation. The algorithm has two parameters, the window length L , $1 < L < N$, and the number of components r , $r < \min(L, K)$, where $K = N - L + 1$. First, the time series is transformed into a trajectory matrix X of size $L \times K$ of the time series X :

$$X = \mathcal{T}_L(X) = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{pmatrix}, \tag{1}$$

where the embedding operator \mathcal{T}_L denotes the bijection between \mathbb{R}^N and \mathcal{H} , and \mathcal{H} is the set of Hankel matrices of size $L \times K$ with equal values on the anti-diagonals $i + j = \text{const}$.

The SSA estimator of the signal is defined as the composition

$$\tilde{S} = \mathcal{T}_L^{-1} \circ \Pi_{\mathcal{H}} \circ \Pi_{\mathcal{M}_r} \circ \mathcal{T}_L(X), \tag{2}$$

where $\Pi_{\mathcal{H}}$ is the orthogonal projector on the set of Hankel matrices \mathcal{H} , and $\Pi_{\mathcal{M}_r}$ is the projector on the set \mathcal{M}_r of $L \times K$ matrices of rank at most r . In both cases, we consider the projector by the Frobenius norm. The projection by $\Pi_{\mathcal{H}}$ is constructed by averaging the values along the anti-diagonals [Section 6.2] in [3], and the result of $\Pi_{\mathcal{M}_r}$ can be obtained via singular value decomposition as the sum of the leading r summands (Eckart–Young theorem [13,14]).

From the description of the algorithm, it follows that to adequately estimate a signal, its signal trajectory matrix must be of rank r or well approximated by a matrix of rank r . This raises the following notion.

Signal Model

Say that a signal S is a series of rank r if its L -trajectory matrix $S = \mathcal{T}(S)$ is rank-deficient and has rank r for any $r < L < N - r + 1$. It is known that the definition of the series of rank r is equivalent to the equality to the rank r of a trajectory matrix with $L = r + 1$ if $N > 2r$ [Corollary 5.1] in [15]. If a signal has rank r , we call it a low-rank series.

For an infinite time series S of rank r , there exists a governing linear recurrence relation (LRR) of order r : $s_n = \sum_{i=1}^r a_i s_{n-i}$, $n = r + 1, \dots, N$, $a_r \neq 0$ [Chapter XVI, Section 10, Theorem 7] in [16]. A well-known result that specifies the explicit form of LRR governed series in parametric form is: $s_n = \sum_i P_i(n) \exp(\alpha_i n) \cos(2\pi\omega_i n + \psi_i)$, where $P_i(n)$ are polynomials in n [Theorem 3.1.1] in (cf. [17] and [Theorem 5.3] in [3]).

The rank of a signal is referred to as the model order, since in the complex-valued case, a sum of r complex exponentials has rank r . Consequently, the rank estimation problem is known as model order selection.

Low-Rank Approximation and SSA

If the signal is a series of known rank r , there are numerous methods for extracting it, including low-rank approximation [18–21]. In particular, the paper [21] proposes an efficient MGN (Modified Gauss-Newton) method for finding the best low-rank approximation using the least-squares method. It is significant to note that in the case of Gaussian white

noise, the least-squares approximation coincides with the maximum likelihood estimate. Alternative approaches include the Cadzow method, which consists of alternating projections to $\Pi_{\mathcal{H}}$ and $\Pi_{\mathcal{M}_r}$ at each iteration. This method has been discussed in the literature, including in the following sources: [22–24].

In all versions, low-rank approximation methods are iterative, which makes even efficient methods quite time-consuming and not guaranteed to find the global minimum. Furthermore, it is always necessary to know the rank of the signal in advance.

Let us describe the cases when low-rank approximation methods, with the choice of r equal to the signal rank, will prove ineffective. The first situation is when the noise level is too high. In this case, the approximation by a signal-rank series includes a significant portion of the noise in the result. In order to extract the signal more accurately, it is necessary to take r less than the signal rank. The second situation is when the signal is not exactly a low-rank series. This is usually the case for real-world time series. In this case, the low-rank approximation can perform poorly.

The version of SSA for signal extraction is a single iteration of the Cadzow method and has a very efficient implementation [25]. Given that the method employs a single iteration, it is not constrained to the extraction of low-rank signals. Instead, it can be used to identify trends and periodic components in real-world time series, which can then be subjected to further analysis and forecasting. By employing the intermediate SSA result in the form of the singular value decomposition of the trajectory matrix, one can perform visual identification of the signal-related components of the decomposition. It is evident that this approach is not feasible when dealing with a vast amount of data. Consequently, techniques for automated component identification within SSA have been developed. For example, in [26], a method for automated trend identification has been proposed, wherein the number of components to be identified must be specified. Consequently, it is also necessary to set r .

The Problem Statement and Contribution

In this study, we propose a novel approach to the problem of estimating the signal rank. Rather than determining the signal rank itself, our objective is to identify the optimal parameter r in the SSA algorithm that minimizes the mean square error (MSE) of the signal reconstruction. In the case of a low-rank signal and low noise, this approach will yield the same result as the conventional method of finding the signal rank. However, in situations where the signal is not exactly of low rank or the noise level is high, the signal rank may not provide the optimal r .

This study will be based on the methods of signal rank detection (model order selection). These methods can be divided into two types: those based on information criteria and those based on the properties of the SSA method (properties of the signal subspace). The information criteria that are currently available were not designed for use in the context of the SSA case; therefore, we propose modifications to them. Given that the original version of the information criteria was developed for the case of Gaussian white noise, we suggest an approach that would allow us to extend them to the case of red noise.

Structure of the Paper

Let us describe the structure of the paper. Section 2 describes known methods for estimating the model order r . In Section 3, we propose an approach that is based on signal estimation by SSA. Section 3.1 considers the case of white noise, and Section 3.2 proposes a way to transfer the methods to the case of red noise. Section 4 includes numerical studies and comparisons on artificial examples. Section 5 verifies the performance of the methods for real-world time series. Section 6 presents a summary and discussion; conclusions complete the paper.

2. Model Order Selection

This section presents an overview of the existing methods for model order selection, including both general information criteria and specific criteria that employ signal subspace models.

2.1. Information Criteria

Let us start with information criteria that are common to model order selection.

2.1.1. General Approach

Let the observed data be in the form $X = P_k(\theta_k) + \mathcal{E} \in \mathbb{R}^N$, where $P_k(\theta_k)$ is a model with the number of parameters k , \mathcal{E} is a random noise with zero expectation.

We can write out the general form of the information criterion (abbreviated as IC in formulas and algorithms) as follows [27]:

$$IC(k) = 2 \ln \mathcal{L}_k - \phi(k), \tag{3}$$

here the maximum value of the likelihood function within the model P_k is denoted by \mathcal{L}_k ; k is the number of parameters in the model; and the penalty for the number of parameters is represented by the function $\phi(k)$. In this form of criteria, the optimal model is indicated by the maximum value.

If the elements $\varepsilon_i, i = 1, \dots, N$, of the noise \mathcal{E} are independent and $\varepsilon_i \sim N(0, \sigma^2)$, then instead of searching for the maximum of (3), it is possible to search for the maximum in k of

$$IC(k) = -N \log(\hat{\sigma}^2) - \phi(k), \tag{4}$$

where

$$\hat{\sigma}^2 = \hat{\sigma}_k^2 = \min_{\theta} \|X - P_k(\theta)\|^2.$$

Different information criteria are distinguished by the penalties that are associated with them. For Akaike information criterion (AIC), $\phi(k) = 2k$ [28] and for Bayesian information criterion (BIC) [29], $\phi(k) = k \log N$.

Another criterion is the Maximum Description Length criterion (MDL [30]), where the penalty additionally depends on the particular model. In the case of independent noise, the MDL criterion is asymptotically equivalent to the BIC criterion as $N \rightarrow \infty$. However, the penalty for the number of parameters differs for finite N .

The following section will present the application of this approach to the model order selection problem for time series, specifically in the context of a low-rank signal.

2.1.2. Information Criteria for Time Series

Consider the problem of estimating the rank r of the signal $S = (s_1, \dots, s_N)^T$ by the observed noisy time series $X = S + N$.

The parametric model of signals of rank r can be considered as a nonlinear model with $k = 2r$ parameters [21]. In the case of a noisy signal with Gaussian white noise, the maximum likelihood estimate \hat{S} of the signal S obtained from the time series X can be obtained by means of the nonlinear least-squares method. Numerically, the least-squares estimate can be obtained using the VPGN method [19] or the MGN method [21]. In both cases, an initial approximation is required, which can be obtained using the Cadzow iterations [22,24].

This approach to information criteria is not without its limitations. First, the nonlinear least-squares method is a computationally expensive procedure due to its iterative nature. Additionally, the range over which the rank r is searched slows down the computation. Second, as demonstrated by numerical experiments (see Section 4), such methods may converge to local minima of the target function, which may result in an overestimation of the rank due to an overestimated error.

2.1.3. Information Criteria for a Sequence of Vectors

In [31], a model for a sequence of vectors is stated: $X_i = \mathbf{A}C_i + \mathcal{E}_i \in \mathbb{R}^L, i = 1, \dots, K$; the rank of the matrix $\mathbf{A} \in \mathbb{R}^{L \times r}$ is equal to an unknown r , and the noises are independent. Here $C_i \in \mathbb{R}^r$ consists of r stationary processes, and L noisy linear combinations of these processes are observed. In matrix form, the model is $\mathbf{X} = \mathbf{A}\mathbf{C} + \mathcal{E}$, where \mathcal{E} is a matrix composed of independent noises.

Thus, in such a model, X_i is a stationary process, allowing us to discuss the model of its auto-covariance matrix. It has a form $\mathbf{R} = \mathbf{A}\mathbf{S}\mathbf{A}^T + \sigma^2\mathbf{I}$, where \mathbf{S} represents the auto-covariance matrix of the signals.

The same model can be valid in a more general formulation of the problem; for instance, the matrix \mathbf{C} is not necessarily random and thus \mathbf{R} is not necessarily a covariance matrix. Additionally, the stationarity condition is not a prerequisite.

Let $\{\mu_i\}$ be the singular values of the matrix \mathbf{R} . Then, as shown in [31], the logarithm of the maximum value of the likelihood function has the form

$$\ln \mathcal{L} = -K \sum_{i=r+1}^L \ln \mu_i + (L - r) \ln(\hat{\sigma}^2), \tag{5}$$

where

$$\hat{\sigma}^2 = \sum_{i=r+1}^L \mu_i / (L - r).$$

The number of parameters is equal to $r(2L - r) + 1$ [31]. After adjusting for the penalty, we obtain the following:

$$\text{MDL}(r) = -2 \ln \mathcal{L} - r(2L - r) \ln K. \tag{6}$$

In this case, we consider the variant of MDL that is the same as BIC. However, to distinguish it from BIC in time series form, we will keep the name MDL.

Remark 1. *Let us show the similarities and differences with the problem given in the time series form. The transfer from time series to trajectory matrices yields the following: \mathbf{X} is the trajectory data matrix, $\mathbf{A}\mathbf{C}$ is a possible form for the presentation of a trajectory signal matrix of rank r , and \mathcal{E} is the trajectory noise matrix. Then, \mathbf{R} , which is formally not an auto-covariance matrix here, can be estimated as $\mathbf{X}\mathbf{X}^T / K$. The distinction lies in the fact that, in the case of time series, the elements of the noise matrix are not independent. This is due to the specific structure of the trajectory noise matrix, which is Hankel, meaning that the noise elements on each anti-diagonal are identical.*

2.2. Subspace-Based Criteria

In this section, we consider two methods, namely ESTER [32] and SAMOS [33], which are not related to information criteria. These criteria are based on signal subspace properties and, in the specific case where the signal subspace has rank r and is extracted exactly, they yield the value $+\infty$ for rank r . These methods are based on the concept of separability in SSA [Sections 1.5 and 6.1] in [3] and aim to identify separability points. The criteria are applicable when the signal is approximately separable from the noise. However, within the signal, its components may also be separable, which makes the methods appear unstable despite the strong advantage that they are directly based on SSA properties. Another disadvantage of the criteria is that they cannot specify $r = 0$, i.e., they cannot determine the case of no signal.

The criterion ESTER is based on the principles of the ESPRIT method [34], which is a subspace-based method used for estimating signal parameters, particularly, frequencies. We define $\mathbf{E}(r) = \overline{\mathbf{U}}_r - \mathbf{U}_r\mathbf{A}$, where $\mathbf{A} = \mathbf{U}_r^{-1}\overline{\mathbf{U}}_r$, \mathbf{U}_r consists of the r leading left singular vectors of the trajectory matrix of the series \mathbf{X} . In the case of an exactly separable signal \mathbf{S} of

rank r , the matrix $\mathbf{E}(r)$ is the null matrix. Therefore, the criterion proposes to consider the maximum of

$$\text{ESTER}(r) = \|\mathbf{E}(r)\|_2^{-2};$$

here the norm $\|\cdot\|_2$ equals the maximum eigenvalue.

Subsequently, in [33], the method SAMOS was proposed, which was also based on the matrix \mathbf{U}_r . Let us denote τ_i the singular values of the matrix $[\overline{\mathbf{U}}_r : \underline{\mathbf{U}}_r]$. Then, we seek the maximum of

$$\text{SAMOS}(r) = \frac{1}{\sum_{i=r+1}^{2r} \tau_i / r}.$$

3. Modifications of Information Criteria for the Case of Hankel Noise

It is known that the set \mathcal{M}_r of matrices of rank at most r in the neighborhood of a matrix of rank r is a smooth manifold of order $r(L + K) - r^2$ (ref. [21], [Ex.13, p.27] in [35]). This allows one to consider the linearization of the projector $\Pi_{\mathcal{M}_r}$, and to approximate the projection on the set \mathcal{M}_r as a linear projection onto a tangent subspace at the desired point.

Our approach is to estimate the variance σ^2 used in information criteria (4) for a given rank r by using the matrix

$$\mathbf{D}_r = \mathcal{T}_L \mathbf{X} - \Pi_{\mathcal{M}_r} \circ \mathcal{T}_L(\mathbf{X}), \tag{7}$$

rather than by constructing the maximum likelihood estimate of the signal. Since the computation of $\Pi_{\mathcal{M}_r}$ is reduced to the summation of the first r summands of the singular value decomposition of the matrix $\mathcal{T}_L(\mathbf{X})$, this allows for fast recalculation for different r and thus provides a fast method for estimating the rank of a signal.

3.1. White Noise

Let \mathbf{N} be a Gaussian white noise with zero expectation and variance σ^2 . The noise series \mathbf{N} can be estimated as $\mathcal{T}^{-1} \circ \Pi_{\mathcal{H}}(\mathbf{D}_r)$. In the proposed approach, the estimation of σ^2 will be conducted without proceeding from the matrix \mathbf{D}_r to a time series.

Let $\mathbf{D}_r = (d_{i,j}^{(r)}) = (d_{i,j})$. Then the estimate of σ^2 can be given as follows (let us call this version ‘SVD’):

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^L \sum_{j=1}^K d_{i,j}^2}{LK}. \tag{8}$$

By employing the singular value decomposition, the same noise variance estimate can be obtained as $\tilde{\sigma}^2 = \sum_{i=r+1}^L \lambda_i / (LK)$, where λ_i are the squares of the singular values of the trajectory matrix \mathbf{X} .

It can be seen from (1) that the operator \mathcal{T}_L repeats each time series element in the trajectory matrix as many times as there are elements on the corresponding anti-diagonal. Consequently, given the Hankel structure of the input matrices, we put forward a more accurate weighted version of the $\tilde{\sigma}^2$ estimator (which we shall henceforth refer to as ‘TRMAT’):

$$\tilde{\sigma}^2 = \frac{1}{N} \sum_{i=1}^L \sum_{j=1}^K \frac{1}{w_{i+j-1}} d_{i,j}^2, \tag{9}$$

where $w_i, i = 1, \dots, N$, are the numbers of elements on the i -th anti-diagonals of an $L \times K$ matrix. The division by N is a consequence of the fact that the number of such diagonals is N . Note, that if the window length L is a relatively small value compared to N , there is a negligible difference with the SVD criterion, since the weights $w_L, w_{L+1}, \dots, w_{N-L}$ are the same for both methods.

In both cases, the alternative estimate of k (instead of $k = 2r$) to substitute into the Formula (4) is

$$k = \frac{N(r(K + L) - r^2)}{LK}, \tag{10}$$

that is, it is the dimension $r(K + L) - r^2$ of the smooth manifold \mathcal{M}_r , reweighted taking into account the “replacement” of the dimension LK of the matrix space by the dimension N of the time series space. Let us explain this normalization. One can look at the contribution of number k of parameters to the penalty in the AIC and BIC values in the form (4) by dividing the expression by N . This shows that the parameter penalty depends on k/N .

We have numerically checked that the non-normalized number of parameters $k = 2r$ that corresponds to the non-Hankel case leads to a severe underestimation of the penalty term $\phi(k)$ in (4) when (8) or (9) is taken as an estimate of σ^2 ; therefore, we will not consider this case further.

We will consider (8) (SVD) and (9) (TRMAT) with the normalized number of parameters given by (10) using the information criteria (4). Recall that the best rank corresponds to the maximum value of a criterion. Preliminary experiments have shown that only the BIC-penalty criteria turned out to be working, and we will consider it further. A graph of $AIC(k)$ flattens out after growth to the correct rank, so the maximum point is determined unstably and the rank is usually overestimated.

3.2. Red Noise

Let the noise be stationary and Gaussian. The procedure that makes the noise white is called whitening and consists of multiplying the time series by the square root of the inverse of the noise autocovariance matrix. The whitening operation affects both the signal and the noise. Since the matrix-form model is stable concerning a linear transformation (multiplication by a full-rank matrix), we can apply the methods of the previous section to the result of the whitening. To apply the criterion, it is sufficient to know what the variance of the white noise after whitening will be; more precisely, it is sufficient to know how the variance of the noise after whitening is expressed through the variance of the original noise. To conduct this, we need to know the covariance matrix of the noise.

Let us denote the variance estimate using the Formula (8) or (9) as $\hat{\sigma}_0^2$. Then, we need to substitute the variance $\hat{\sigma}^2$ of the whitened noise into the Formula (4). For example, in the case of an AR(1) model with coefficient ϕ ,

$$\hat{\sigma}^2 = \hat{\sigma}_0^2(1 - \phi^2). \tag{11}$$

Recall that red noise is an AR(1) process with a positive coefficient.

To implement this approach, it is sufficient to estimate the coefficient ϕ , which is equal to the correlation coefficient between successive observations. As before, $\mathbf{D} = \mathbf{D}_r$ is the residual matrix defined in (7). It is not exactly Hankel. Since in the case of wrong rank, its structure is far from Hankel and diagonal averaging should distort it considerably, let us estimate the correlation using the matrix before averaging by shifting the rows of the residual matrix.

Let

$$c = \sum_{i=1}^L \sum_{j=1}^{K-1} d_{i,j}d_{i,j+1}, \quad v_1 = \sum_{i=1}^L \sum_{j=1}^{K-1} d_{i,j}^2, \quad v_2 = \sum_{i=1}^L \sum_{j=1}^{K-1} d_{i,j+1}^2.$$

Then, as an estimate of ϕ we take

$$c / \sqrt{v_1 v_2}. \tag{12}$$

Remark 2. The idea behind TRMAT can also be applied to the evaluation of ϕ by considering the same weights $\frac{1}{w_{i+j-1}}$ in each sum as in (9). We will apply such an estimate with weights to the algorithm TRMAT.

An alternative is to estimate ϕ as the correlation between successive observations in the series $\mathbf{D}_r = \mathcal{T}^{-1} \circ \Pi_{\mathcal{H}}(\mathbf{D}_r)$, but we will not consider it since it did not lead to an improvement in preliminary numerical experiments.

Unfortunately, especially if the signal is not stationary and even more so if it is not a low-rank series, the calculated estimate of ϕ based on nonstationary residuals in the case of

wrong rank can be accidental and lead to an incorrect maximum of an information criterion. Therefore, we consider the following variant in which the information criterion is assigned $-\infty$ when the residual is clearly nonstationary (recall that the best model corresponds to the maximum of criteria):

1. Check the series D_r for nonstationarity (this is optional, not necessary), e.g., using the KPSS test [36]. If the hypothesis is rejected (e.g., p -value < 0.05), then $\phi = 1$ and STOP.
2. Find the better model for D_r from white and red noise models. If it is detected as white noise, then $\phi = 0$, and if it is closer to red noise, then estimate parameter ϕ in the red-noise model, for example, by the MLE method without requiring the model to be stationary.
3. If $\phi \geq 1$, then the criterion returns $-\infty$; otherwise, the value is calculated using the formula of the corresponding information criterion.

The option of criteria with the adjustment (11) according to the estimated ϕ will be referred to by adding `_AR`.

3.3. Case of Zero Signal

Information criteria allow one to consider the absence of a signal as one of the models. In this case, in the form of the criterion (4), the mean square of the values of the initial time series serves as the estimate of σ^2 .

Accordingly, the signal values are 0. Recall that the ESTER and SAMOS criteria do not consider the case $r = 0$. Formally, criteria ESTER and SAMOS return $-\infty$ in the case $r = 0$, so the value $r = 0$ will never provide the maximum for these criteria.

3.4. Algorithm

Algorithm 1 describes how to compute BIC versions of the proposed criteria for white and red noise cases.

Algorithm 1 Calculation of TRMAT and SVD

Input: Time series X , window length L , rank r , type of IC (TRMAT or SVD), indicator CHECKSTAT if the stationarity check is needed, significance level α for checking stationarity, indicator NOISETYPE.

Result: Value of IC.

1. Apply SSA with L and r to X of length N .
 2. Calculate the residuals in the matrix (D_r) and time-series (D_r) forms.
 3. Calculate $\hat{\sigma}^2$ by (9) if IC = TRMAT or by (8) if IC = SVD.
 4. If NOISETYPE = RED:
 - Estimate ϕ by (12) using matrix residuals D_r (with weights if IC = TRMAT; Remark 2).
 - If CHECKSTAT = TRUE and the stationarity of D_r is rejected with significance level α or if $\phi > 1$, set $\phi = 1$ and return IC = $-\infty$.
 - Recalculate $\hat{\sigma}^2 \leftarrow \hat{\sigma}^2(1 - \phi^2)$;
 5. Calculate the number of parameters k by (10).
 6. Return IC = $-N \log \hat{\sigma}^2 - k \log N$ (4).
-

For MDL and white noise, the values of the criterion are calculated by (6). If the noise is red, then $\hat{\sigma}^2$ in (6) is recalculated with the substitution of $\hat{\sigma}^2(1 - \phi^2)$ instead of $\hat{\sigma}^2$.

4. Numerical Experiments

Let us numerically compare the considered methods and study their accuracy as a function of the noise level.

4.1. Approach to Comparison

One of the criterion quality characteristics used in practice for model order estimation is the proportion of correct order (rank) estimates or the bias of the average rank estimate. However, overestimation or underestimation of the rank can have different effects on the result, because the decomposition components are arranged by decreasing their contribution, and therefore, overestimation of the rank leads to a smaller increase in the signal estimation error compared to underestimation of the rank. Therefore, we will consider the RMSE of the signal estimation as the main characteristic. Note, also that the trend identification methods [26,37] are robust to rank overestimation when the number of identified components is chosen according to the estimated rank.

In the problem statement considered in this paper, the correct order of models is generally not defined. Therefore, we will compare the estimated ranks (model orders) with the optimal rank r , which gives the minimum error of the signal estimation obtained by (2). Accordingly, we will compare the RMSE of the signal estimation at the estimated rank with the minimum error at the optimal rank. Since the best approximation depends on the noise level, we will consider the quality of the criteria depending on the noise level.

Thus, in most cases, we will compare the signal estimation error with the average minimum error and the average rank with the average optimal rank. In addition, we will consider the proportion of rank estimate matches with individual optimal ranks that yield the minimum errors for each series separately.

4.2. White Noise

In this section, we consider the case of a noisy signal, where the noise is Gaussian white with zero mean and variance σ^2 .

4.2.1. Sum of Two Sinusoids

Let us start with a simple example, a signal in the form of the sum of two sinusoids:

$$s_n = 10 \sin(2\pi n/20) + \sin(2\pi n/10). \quad (13)$$

The rank of the signal (13) is 4 [Example 5.2] in [3]. Since deterministic signals are asymptotically separated from noise [Section 6.1.3] in [3], the optimal value of r will be 4. However, as the noise level increases, the second sinusoid will start to mix with the noise and after some period of uncertainty, the optimal rank will become equal to 2, the rank of one sinusoid. It is clear that for any signal, as the noise increases, at some noise level the optimal rank becomes zero, i.e., the best estimate of the signal is the zero series.

We will consider 20 values of σ from 0.01 to 100 with equal logarithmic steps. Separately, we will focus on four noise levels, $\sigma_1 = 0.1128838$ (optimal rank 4), $\sigma_2 = 1.274275$ (transition period from 4 to 2), $\sigma_3 = 5.45559478$ (optimal rank 2) and $\sigma_4 = 61.58482111$ (optimal rank 0).

We begin by examining the proportion of matches between the estimated ranks and the individual optimal ranks (which yield the smallest RMSE of the signal estimates for the given series).

In comparison to the methods SVD, TRMAT, and MDL, the methods ESTER and SAMOS demonstrate the most favorable outcome, with a proportion of matches reaching 0.998 (out of 1000 trials) at a low noise level ($\sigma = 0.01$) and window length $L = 50$. This outcome aligns with the findings presented in the studies [32,33]. However, as the noise level increases, the methods lose efficacy, yielding near-zero matches. Figure 1 illustrates the dependence of standardized criterion values on rank. For comparability, the criteria were standardized (i.e., the mean was subtracted and the values were divided by the standard deviation depicted in the graph) as the criterion scales may be incomparable.

To more effectively illustrate the impact, we have selected the window length $L = 40$ as a multiple of both sinusoids' periods and the standard deviation $\sigma = \sigma_1$. It is readily apparent that the ESTER and SAMOS criteria lead to an erroneous determination of the rank in this instance, particularly the ESTER method, with a maximum at the separability point $r = 2$. In contrast, the considered information criteria, such as TRMAT, effectively resolve the rank to be four.

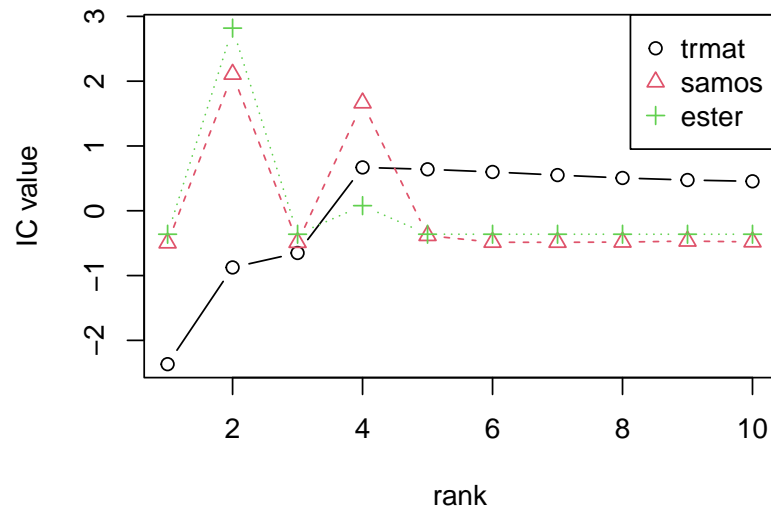


Figure 1. Example 'sine': The dependence of criteria $IC(k)$ on rank k for one realization of noise.

Thereby, in the following sections, we will not consider the ESTER and SAMOS criteria, particularly given that they are not applicable in the absence of a signal. Thus, in what follows, we will examine the criteria SVD, TRMAT, and MDL in greater detail.

As illustrated in Table 1, for noise levels corresponding to the stable rank detection (i.e., all except the level σ_2), the TRMAT, SVD, and MDL methods consistently yield optimal rank estimates. For each of the aforementioned methods, the proportion of matches with optimal ranks is nearly one. However, at the noise level σ_2 , the TRMAT and SVD methods were unsuccessful, with only a small proportion of matches being identified. In comparison, the MDL method demonstrated better performance, with a success rate of 0.48, as opposed to 0.235 and 0.277 for the TRMAT and SVD methods, respectively.

Table 1. Example 'sine': proportion of ranks coinciding with the individual optimal ranks.

σ	TRMAT	SVD	MDL
σ_1	0.983	0.990	0.969
σ_2	0.235	0.277	0.480
σ_3	0.984	0.988	0.966
σ_4	0.992	0.994	0.974

Figure 2 illustrates the distinction in the behavior of the criteria for the two noise levels, namely, σ_1 and σ_3 . As previously, the criterion values have also been standardized.

Figure 3 depicts the mean estimated ranks and the optimal ranks for the smallest root of average MSEs over 1000 realizations for the TRMAT and MDL methods (SVD has been omitted due to its similar behavior to TRMAT). It can be observed that the optimal rank graph as a function of noise levels contains plateaus with identical rank values and transition periods.

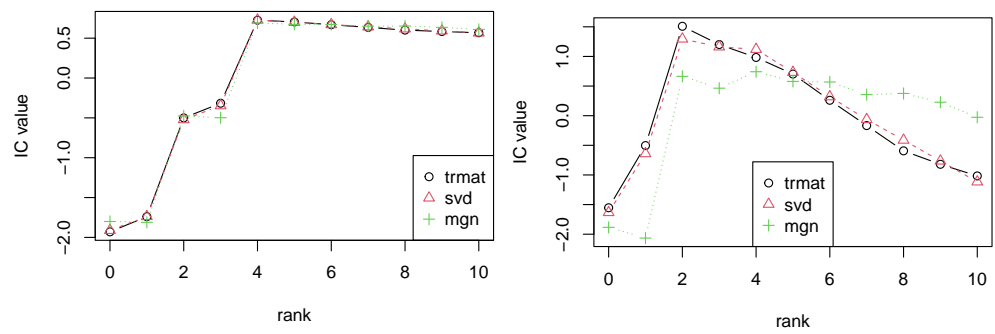


Figure 2. Example ‘sine’: The dependence of criteria $IC(k)$ on rank k for one realization of noise (noise level σ_1 on the left, σ_3 on the right).

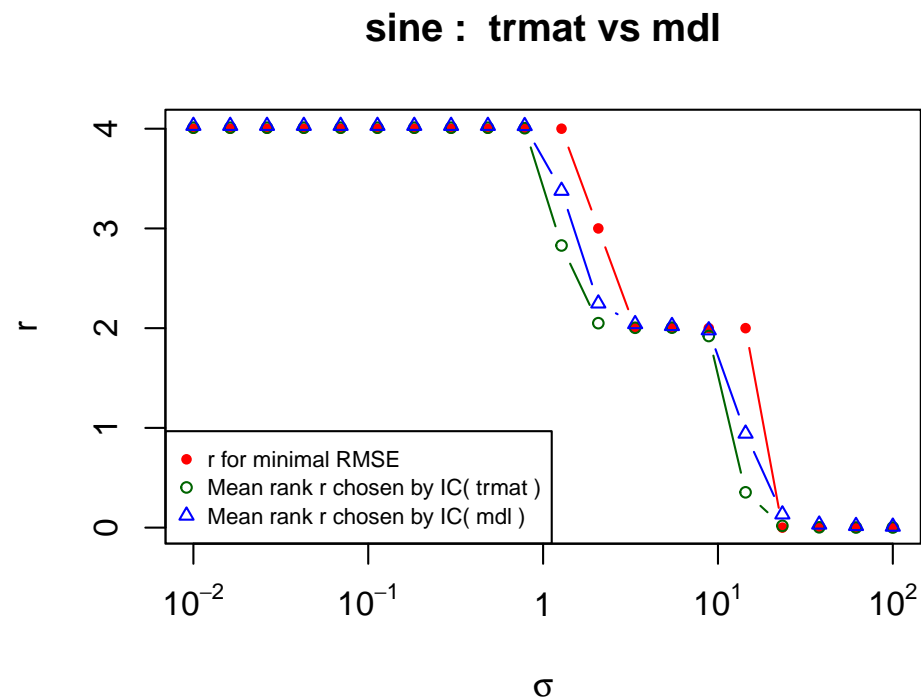


Figure 3. Example ‘sine’: Dependence of rank estimation on white noise level, TRMAT and MDL.

At plateaus, both methods yield accurate estimates of the rank. During transition periods, both methods underestimate the rank, but the MDL method does so to a lesser extent.

The graphs of RMSE versus noise level are presented in Figure 4. We will now explain the specifics of the graph depicted in this figure. First, as a baseline, we consider the RMSE at the maximum possible rank, that is, when the signal estimate is the entire original series and the RMSE is equal to the root of the mean squared time series values. All errors depicted in the plot are presented on a relative scale, wherein they are divided by the baseline RMSE. Accordingly, the value of 1 corresponds to a signal estimate that is equal to the original series. In a sense, this represents the most unfavorable scenario. Figure 4 also presents the optimal case, which corresponds to the rank associated with the lowest average error. The lines exhibit minimal divergence at the plateau, indicating stable rank detection, while diverging at the transition periods. It is evident that the error of both criteria exceeds the minimum error at the transitions. However, the error of the MDL criterion is slightly smaller than that of TRMAT, which is consistent with the results presented in Figure 3.

Let us include MGN in the consideration. When the criterion MGN is applied, we obtain an MGN signal estimate and use this estimate for calculating the MSE of the signal estimate; in particular, for finding the optimal ranks.

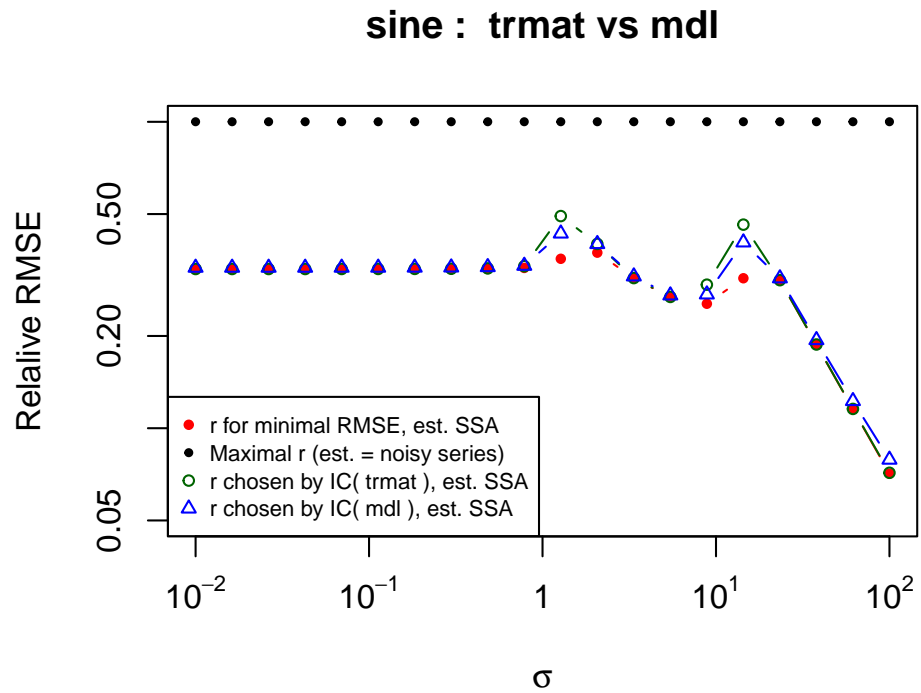


Figure 4. Example ‘sine’: Dependence of relative RMSE on white noise level, TRMAT and MDL.

Figure 5, which depicts the mean of the rank estimates as a function of noise level, shows that MGN overestimates the rank at plateaus and thus is much closer to the optimal rank at transitions than the other criteria, which estimate the rank accurately at plateaus and underestimate at transitions. It should be noted that the optimal ranks depicted in Figure 5 may differ from those presented in Figure 3. This discrepancy arises from the fact that, in the MGN criterion, the estimate of the signal is obtained through the MGN method, rather than the SSA. A comparison of the two figures reveals that the optimal MGN ranks do not exhibit a transition range of noise levels, whereas the optimal SSA ranks display noise levels that correspond to an intermediate rank of 3. Consequently, it is possible to calculate both minimal mean squared errors (MSE) using the MGN estimates and minimal MSEs using the SSA estimates. The former are, naturally, smaller, given that the MGN estimates are obtained by the least-squares method.

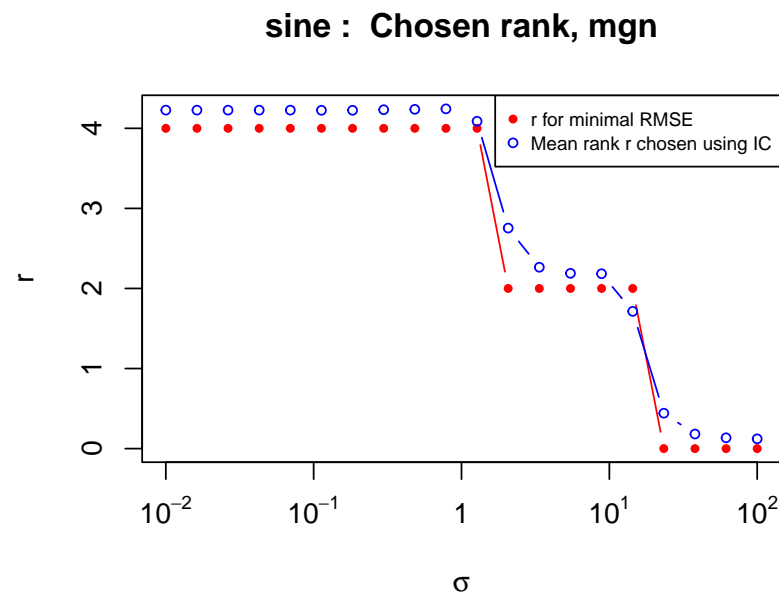


Figure 5. Example ‘sine’: Dependence of rank estimation on white noise level, MGN.

As a consequence of rank overestimation, the resulting RMSE for the MGN-estimated rank is observed to be larger at plateaus than that of the minimal MGN error. Therefore, at plateaus, the MGN criterion provides the same level of accuracy (Figure 6) as TRMAT, with smaller errors occurring at transitions. Figure 6 also presents a hybrid scenario in which the criterion is TRMAT and the signal estimation is conducted using MGN. This combination results in the smallest errors, as TRMAT determines rank with greater precision and MGN generates a more precise estimate of the signal.

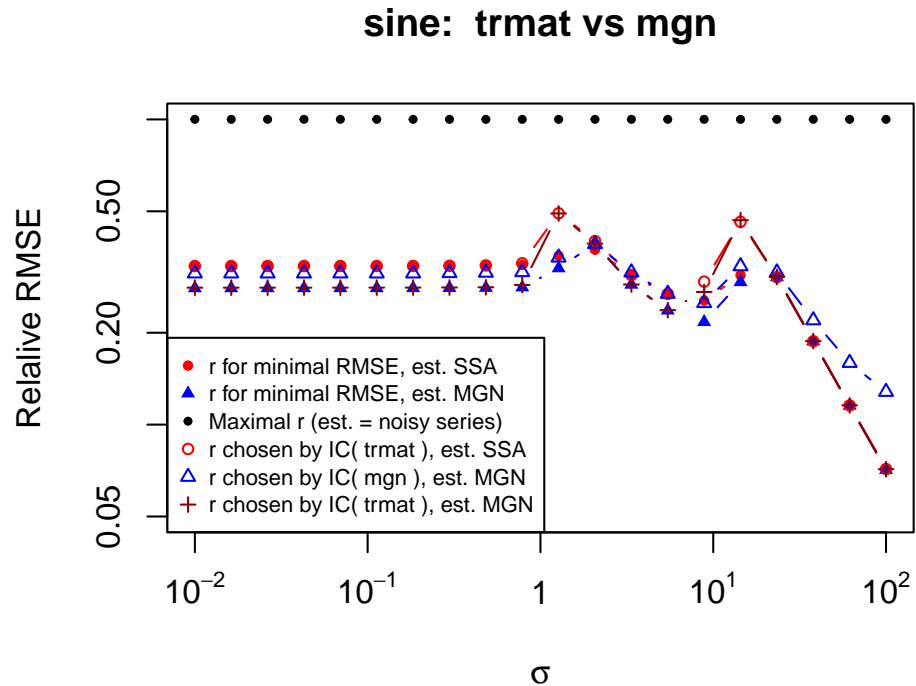


Figure 6. Example ‘sine’: Dependence of relative RMSE on white noise level, TRMAT and MGN.

4.2.2. Logarithmic Signal

As an example of a signal that is not low-rank, consider the signal in the form of a logarithmic series:

$$s_n = \log n. \tag{14}$$

As the signal is not low-rank, it is not possible to determine a proper rank; however, the selection of an appropriate model order can be discussed. Figure 7 depicts the mean estimated ranks and the optimal ranks (model orders) for the smallest root of the average mean squared errors (MSEs) over 1000 realizations. As the noise level increases, the optimal rank values decrease from four to zero. The plateaus are relatively short, and there are significantly more transition regions than in the previous example involving a finite-rank signal. Figure 8 illustrates the dependence of the root mean square error (RMSE) on the noise level.

In this example, we will consider the same noise levels as previously described in Section 4.2.1. In contrast with the previous example, the noise level σ_1 falls on the transition period between ranks 3 and 2, σ_2 is almost on the plateau (rank 1), σ_3 lies exactly on the plateau (rank 1), and for σ_4 , the signal is not detected.

Table 2 correlates with this description of noise levels. In the first row, the accuracy of the criteria is generally less precise; in the second row, it is more precise and in the third and fourth rows, it is highly precise. It can be seen that the TRMAT method produces the best result, providing a good match of ranks on transition periods as well.

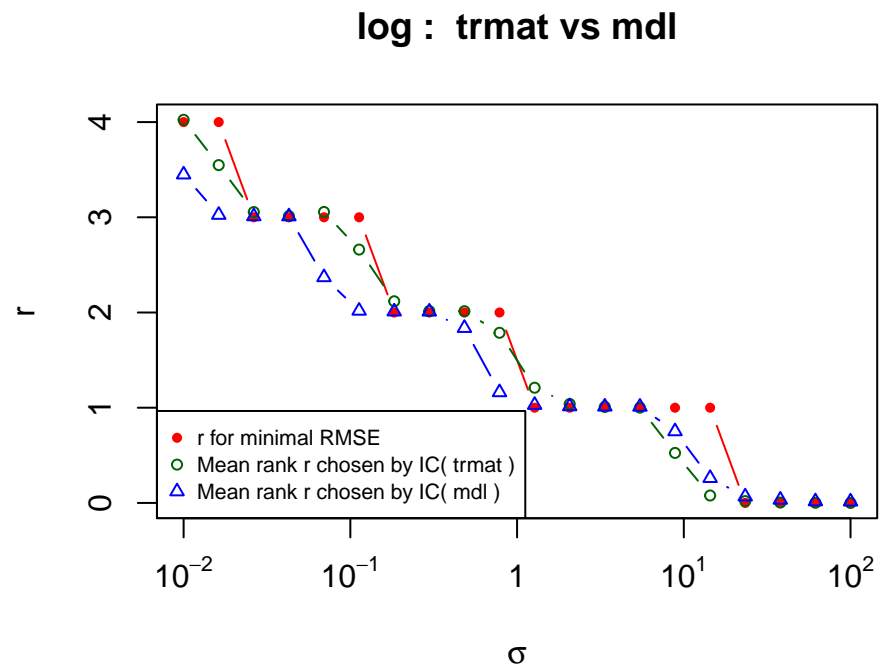


Figure 7. Example ‘log’: Dependence of rank estimation on white noise level, TRMAT and MDL.

Table 2. Example ‘log’: proportion of ranks coinciding with the individual optimal ranks.

σ	TRMAT	SVD	MDL
σ_1	0.744	0.214	0.206
σ_2	0.751	0.603	0.589
σ_3	0.983	0.983	0.976
σ_4	0.997	0.998	0.978

In the preceding example, the TRMAT and MDL criteria exhibited comparable accuracy, with MDL demonstrating a slight advantage at transition sections in the noise levels. In the case of the considered signal that is not low-rank, the MDL criterion exhibited instability, while the TRMAT criterion demonstrated a notable advantage. These observations are illustrated in Table 2 and Figure 8.

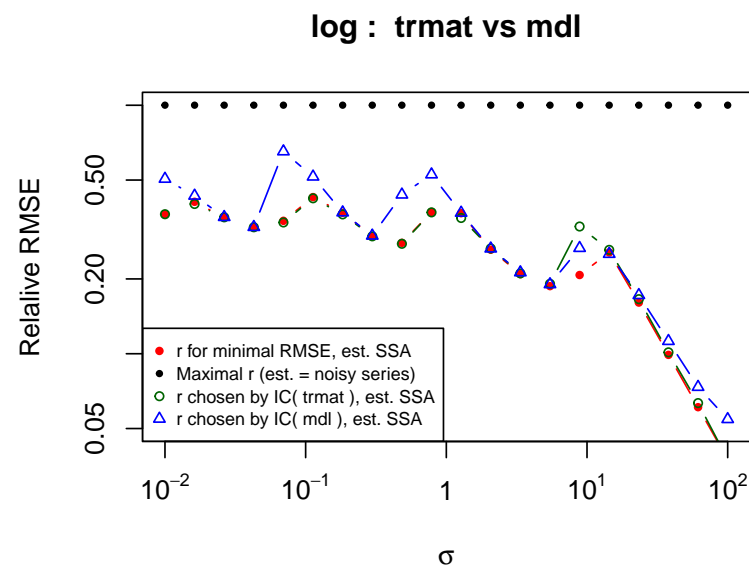


Figure 8. Example ‘log’: Dependence of relative RMSE on white noise level, TRMAT and MDL.

The MGN criterion is now to be incorporated into the comparison (see Figures 9 and 10). One can see that the error lines for TRMAT and MGN are intertwined, indicating that there is no clear advantage of one criterion over the other. In this instance, the combination of the TRMAT criterion for rank detection and the MGN signal estimation with the obtained rank does not result in any improvement and thus is not depicted in the graph.

log : Chosen rank, mgn

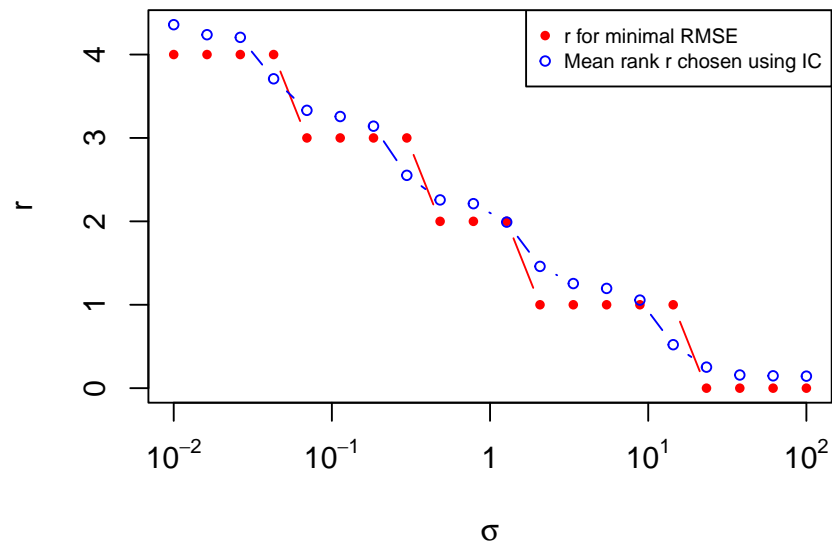


Figure 9. Example ‘log’: Dependence of rank estimation on white noise level, MGN.

log: trmat vs mgn

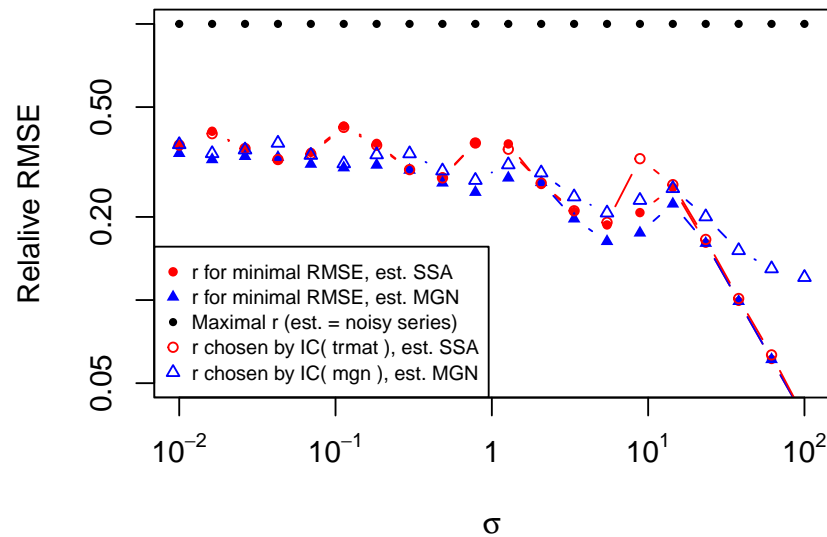


Figure 10. Example ‘log’: Dependence of relative RMSE on white noise level, TRMAT and MGN.

4.3. Red Noise

In this section, we examine a more complicated case of red noise. Since there is no effective implementation of the MGN method for red noise with unknown autoregression parameters, we will not consider it in this section.

In this study, we consider the red noise process in the form $\zeta_{n+1} = \phi\zeta_n + \sigma\epsilon_n$, where $\epsilon_n \sim N(0,1)$ are independent between themselves and with $\zeta_n, 0 \leq \phi < 1$;

$\xi_n \sim N(0, \sigma^2 / (1 - \phi^2))$. Accordingly, when calculating the baseline RMSE, it is necessary to normalize it by dividing by $\sqrt{1 - \phi^2}$ in order to obtain 1 for the maximum relative RMSE.

4.3.1. Sum of Two Sinusoids

In this section, we will examine an example with the signal defined in (13). We recall that the rank of the signal in this case is 4. As with white noise, an increase in the variance results in the second sinusoid becoming mixed with noise. As the variance continues to increase, after a period of uncertainty, the optimal rank becomes equal to 2 and then to 0.

As in the preceding analysis, 20 values of σ will be considered, ranging from 0.01 to 100 in equal logarithmic steps.

Figure 11 depicts the mean estimated ranks and the optimal ranks for the smallest root of average MSEs over 1000 realizations. As demonstrated in Figure 12, the methods exhibit a high degree of similarity in their performance. At plateaus, the criteria accurately determine the rank of the signal, thereby confirming their capacity to estimate the rank correctly. However, at the transition from model order 2 to 0, both methods demonstrate a notable decline in performance, with a pronounced tendency to underestimate the rank.

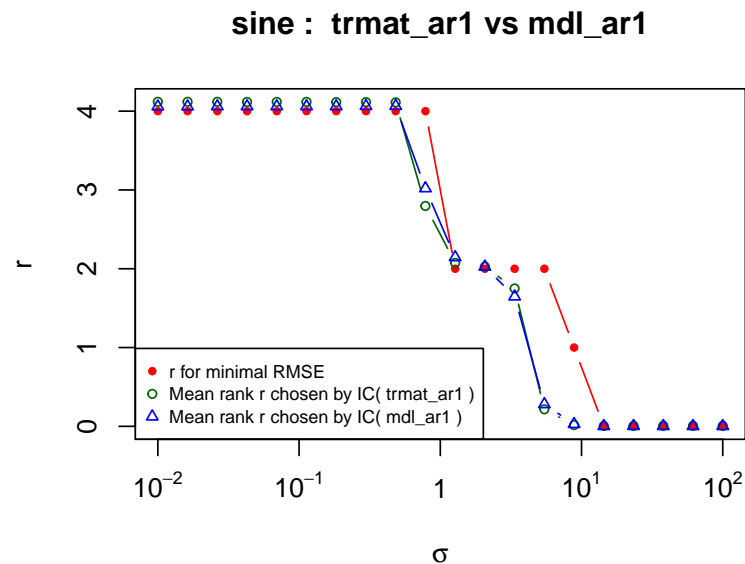


Figure 11. Example ‘sine’: Dependence of rank estimation on red noise level, TRMAT and MDL.

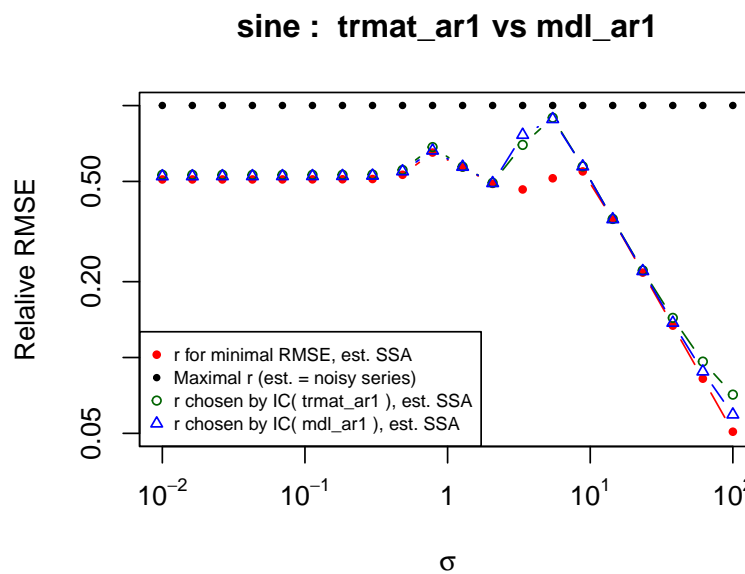


Figure 12. Example ‘sine’: Dependence of relative RMSE on red noise level, TRMAT and MDL.

4.3.2. Logarithmic Signal

As an example of a signal that is not low-rank, consider a series with the signal (14). In this case, we use the version of the criteria with the stationarity check (Algorithm 1, CHECKSTAT = TRUE).

Figure 13 depicts the mean estimated ranks and the optimal ranks for the smallest root of average MSEs over 1000 realizations. In this case, the superiority of TRMAT is clearly evident at relatively low noise levels, as illustrated in Figure 14. At high noise levels, however, both methods perform poorly, largely due to a significant underestimation of ranks.

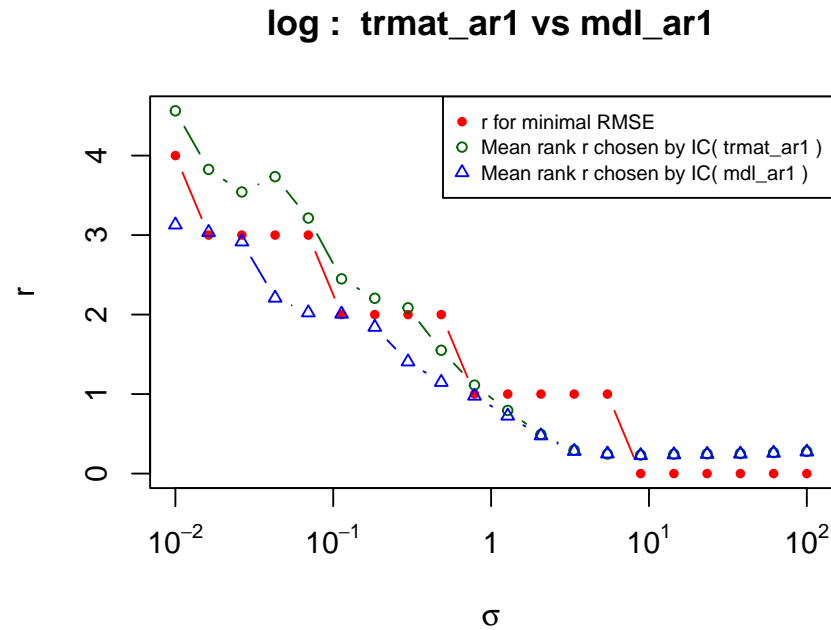


Figure 13. Example ‘log’: Dependence of rank estimation on red noise level, TRMAT and MDL.

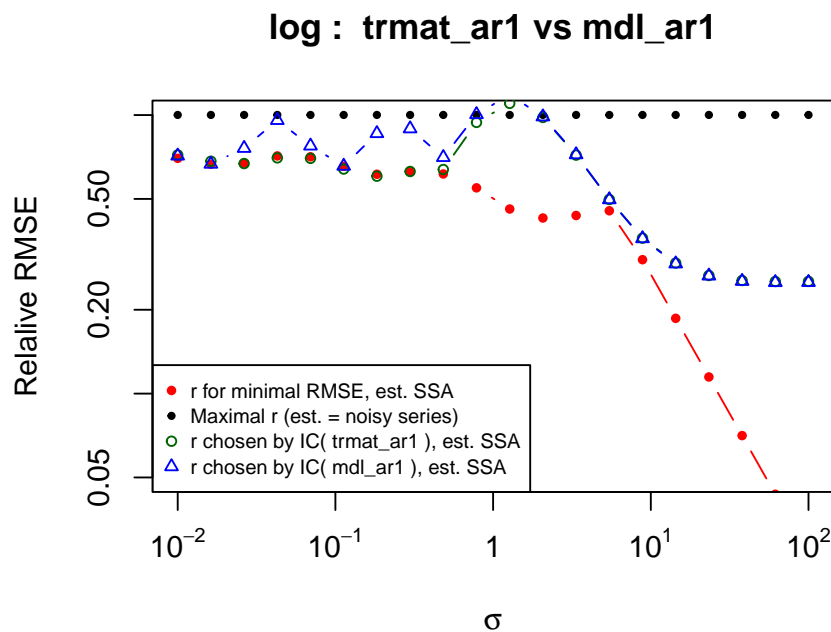


Figure 14. Example ‘log’: Dependence of relative RMSE on red noise level, TRMAT and MDL.

4.4. Zero Signal

In the case of white noise, all three methods (TRMAT, SVD, and MDL) explicitly indicate that the rank is equal to zero. The noise level plays no role in this particular scenario.

In the case of red noise, it is recommended that a criterion without an additional check for stationarity of the residual be used for the stationary signal. In this case, for $\phi = 0.7$, the three methods, TRMAT_AR, SVD_AR, and MDL_AR, almost always yield a rank estimate equal to 0.

5. Real-Life Time Series

Consider ten real-world time series, they are described at <https://ssa-with-r-book.github.io/datasets.html> (accessed on 10 August 2024):

- wines (from 1 to 6), Australian wine sales (Total, Drywhite, Fortified, Red, Rose, Sparkling), monthly, length 174;
- co2, Atmospheric concentrations of CO₂, monthly, length 468;
- dwarf, Time variation of the intensity of the variable white dwarf star, every 10 s, length 618;
- oilprod, Crude oil and natural gas plant liquids production, monthly, length 300;
- treerings, Tree ring indices, annual, length 669.

We take the window length for each series to be half the length of the series.

Table 3 presents the model order estimates by the specified criteria in three versions: for the white noise case and with the estimation of the parameter ϕ of the autoregressive process AR(1), without checking for stationarity (labeled “no”) and with checking for stationarity. The last column contains the ranks that were estimated visually. It is not always possible to determine the number of components associated with a given signal. In such cases, a number of values are provided. In general, the estimates of the model orders produced by the various criteria are consistent with the visual estimates. As can be observed, the MGN criterion produces an overestimate of the rank in the majority of cases, while the TRMAT criterion produces an underestimate in several instances. The stationarity check has a negligible impact on the outcome. The results are, for the most part, satisfactory. In particular, the signal rank for the series “tree rings” is indeed equal to 1, as the series is similar to a random process with a non-zero constant mean.

Table 3. Original series: Ranks estimated by the criteria and visual rank estimation.

ts	MGN	TRMAT	MDL	TRMAT ar1 (no)	MDL AR (no)	TRMAT AR	MDL AR	Visual Est.
wine1	14	13	13	14	13	14	13	13
wine2	12	11	11	11	11	11	11	11–14
wine3	15	11	11	7	11	7	11	11
wine4	13	5	10	6	10	6	10	14
wine5	22	2	11	1	12	1	13	12
wine6	12	12	12	12	12	12	12	12
co2	11	12	21	6	6	6	6	6, 15, 25
oilprod	22	21	21	16	16	16	16	16+
dwarf	10	15	15	11	14	11	14	12–14
treerings	5	1	3	1	1	1	1	1

Since the criteria were developed for the case of stationary noise, we will assess whether the performance of the criteria improves after the variance is stabilized. Table 4 illustrates the outcomes of the criteria applied to the series after the Box-Cox transformation [38] with the automatically estimated parameter λ , the value of which is provided in the last column. It can be observed that the results have been equalized, and the effects of over-ranking by the MGN criterion and under-ranking by the TRMAT criterion have been mitigated. The

effect of the stationarity test is entirely absent. Therefore, although the SSA method does not require stationarity, it is recommended to perform a variance stabilizing transformation for more accurate and stable rank estimation.

Table 4. Series after the Cox-Box transformation: Rank values estimated by the criteria and estimates of the transformation parameters.

ts	MGN	TRMAT	MDL	TRMAT AR (no)	MDL AR (no)	TRMAT AR	MDL AR	λ
wine1	14	13	13	14	13	14	13	0.35
wine2	12	11	11	11	11	11	11	0.38
wine3	15	11	11	7	11	7	11	0.19
wine4	14	10	12	6	12	6	12	0.51
wine5	12	12	12	12	12	12	12	−0.24
wine6	12	9	9	9	9	9	9	−0.27
co2	21	12	21	6	6	6	6	−0.03
oilprod	18	21	21	16	16	16	16	−0.33
dwarf	-	-	-	-	-	-	-	-
treerings	-	-	-	-	-	-	-	-

6. Summary and Discussion

In this paper, we considered a variety of criteria for determining the model order (signal rank) in SSA, including non-conventional cases, namely, signals that are not low-rank or mixed with noise. The criteria considered were ESTER, SAMOS, MGN, MDL, SVD, and TRMAT (the latter two were proposed by us).

The ESTER and SAMOS criteria appeared to be unsuitable for determining model order in the majority of the considered cases. While the information criteria can be employed, none of them have a comprehensive theoretical justification for application to the considered problem statement. It is also important to note that this justification is unlikely to be obtained due to the overly general formulation of the problem, which allows for the possibility of signals that are not low-rank and have a high noise level.

An exception is the MGN criterion, which is theoretically justified in the case of a low-rank unmixed signal in the presence of white noise. This is due to the fact that it numerically searches for the maximum likelihood estimate (MLE) of the signal. The MGN method, which is used in the MGN criterion for signal estimation, is computationally expensive, while it is implemented as computationally fast as possible. The used optimization method is local, as are many others, and can converge to a local extremum, leading to an excessive estimation of the standard error and thus an overestimation of the rank by the MGN criterion. In general, an overestimation may not be a significant issue, as the higher the number of decomposition components, the smaller their contribution. However, the high computational cost represents a significant obstacle to the application of the MGN criterion, even in the case of a low-rank signal.

In the case of SSA, when the singular value decomposition is performed once and the signal estimation is generally not a low-rank series, three variants, MDL, SVD, and TRMAT, were considered in the BIC version, since the penalty from the AIC criterion leads to a significant overestimation of the signal rank. In this case, the MSE signal estimation error (an estimate of σ^2) used in the information criteria is not based on the time series decomposition after diagonal averaging, but on the estimated noise matrix before hankelization. In our proposed version of TRMAT, the Hankel structure of the trajectory matrix is accounted for using weights.

Numerical studies have demonstrated that for relatively simple cases, such as a noisy sum of sinusoids with a low noise level, the methods yield approximately the same results. A slight advantage of the MDL method can be observed in transition regions where the optimal order of the model changes. In the case of a signal that is not low-rank, such as a

logarithmic signal, the proposed TRMAT criterion is preferred. In both cases, the MGN criterion is comparable to TRMAT.

Let us turn our attention to the issue of computational cost. In the case of white noise, the SVD and MDL methods have the same cost as SSA itself, which is $O(N \log N)$. The TRMAT method requires an additional computation that can be implemented at the same asymptotic cost as $N \rightarrow \infty$. Consequently, TRMAT is slightly more expensive. The MGN method is significantly more expensive, as demonstrated in [21], with a computational cost of one iteration being $O(N \log N)$. However, the number of iterations required for convergence can be considerable.

In order to apply the given criteria to the case of red noise, the well-known technique of noise whitening was employed. Due to the linearity of the model, the approach can be reduced to multiplying the estimated variance by $1 - \phi^2$, where ϕ is the AR(1) coefficient.

In the case of an incorrect model, ϕ can be estimated by an irrelevant value; this may lead to an incorrect estimation of the signal rank. The proposed approach is as follows. If the series has a trend, one can first test the residual of the extracted signal of rank r for stationarity and if the hypothesis is rejected, not consider this value of rank as a candidate for the rank estimation. Numerical experiments have shown that the stationary check improved the accuracy in the considered examples. However, in the absence of a trend, such a check leads to a worsening of the rank estimation. For the TRMAT method, the estimate of ϕ can be improved using the same weights caused by the Hankel structure of trajectory matrices as it is in the TRMAT method itself.

We do not discuss the costs of the methods for the red noise case because estimating ϕ and checking the stationarity of the residuals can be conducted in different ways and it is currently difficult to choose the best one.

A general recommendation on the choice of a criterion that is based on numerical experiments is as follows: If one does not take into account the considerable time consumption associated with MGN, TRMAT can be recommended for both white and red noise cases.

In addition, the paper presented an approach to comparing methods based on the relative MSE as a function of the noise variance. The range of noise levels is divided into plateaus and transitions, providing a more structured framework for interpreting the comparison results. Without this approach, the comparison results were less organized and difficult to interpret.

The application of the methods to real-world data sets demonstrated satisfactory outcomes, with the suggestion that a variance-stabilizing transformation should be performed.

7. Conclusions

A number of information criteria have been proposed and investigated for estimating the amount of signal-related SSA components in the noisy signal model, including applications to real-world time series. The approach that considers the matrix residuals and the weights resulting from the Hankel structure of the trajectory matrices enabled the construction of appropriate criteria, particularly in the case of white noise. The whitening technique was employed in the processing of the red noise case. As might be expected, the red noise case is more complex and offers less accurate results. Consequently, it is necessary to implement additional verification procedures within the information criteria for signals exhibiting trends. We have studied the possibility of testing the stationarity of the residuals. In general, the proposed approach can be extended to the case of AR(p) noise.

Since the model order is typically used for a specific purpose, it is necessary to further investigate the methods discussed in this paper from the perspective of addressing a particular problem, such as trend extraction, parameter estimation, and forecasting.

Author Contributions: Conceptualization, N.G. and N.Z.; methodology, N.G. and N.Z.; software, N.G. and N.Z.; validation, N.G. and N.Z.; formal analysis, N.G. and N.Z.; investigation, N.G. and N.Z.; resources, N.G.; writing—original draft preparation, N.G. and N.Z.; writing—review and editing,

N.G. and N.Z.; visualization, N.G. and N.Z.; supervision, N.G.; project administration, N.G.; funding acquisition, N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the RSF, project number 23-21-00222.

Data Availability Statement: The scripts in R can be found in <https://doi.org/10.5281/zenodo.13298984> (accessed on 1 September 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SSA	Singular spectrum analysis
LRR	Linear recurrent relation
IC	Information criterion
AIC	Akaike information criterion
BIC	Bayesian information criterion
MDL	Minimum description length
AR(p)	Autoregressive process of order p

References

- Broomhead, D.; King, G. Extracting qualitative dynamics from experimental data. *Phys. D* **1986**, *20*, 217–236. [CrossRef]
- Elsner, J.B.; Tsonis, A.A. *Singular Spectrum Analysis: A New Tool in Time Series Analysis*; Plenum: Ljubljana, Slovenia, 1996.
- Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A. *Analysis of Time Series Structure: SSA and Related Techniques*; Chapman&Hall/CRC: Boca Raton, FL, USA, 2001.
- Tufts, D.; Kumaresan, R.; Kirsteins, I. Data adaptive signal estimation by singular value decomposition of a data matrix. *Proc. IEEE* **1982**, *70*, 684–685. [CrossRef]
- De Moor, B. The singular value decomposition and long and short spaces of noisy matrices. *IEEE Trans. Signal Process.* **1993**, *41*, 2826–2838. [CrossRef]
- Hermus, K.; Wambacq, P.; Van hamme, H. A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition. *EURASIP J. Adv. Signal Process.* **2006**, *2007*, 045821. [CrossRef]
- Golyandina, N. Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing. *WIREs Comput. Stat.* **2020**, *12*, e1487. [CrossRef]
- Gao, Y.; Qu, C.; Zhang, K. A Hybrid Method Based on Singular Spectrum Analysis, Firefly Algorithm, and BP Neural Network for Short-Term Wind Speed Forecasting. *Energies* **2016**, *9*, 757. [CrossRef]
- Shang, Q.; Lin, C.; Yang, Z.; Bing, Q.; Zhou, X. A Hybrid Short-Term Traffic Flow Prediction Model Based on Singular Spectrum Analysis and Kernel Extreme Learning Machine. *PLoS ONE* **2016**, *11*, e0161259.
- Liu, H.; Mi, X.; Li, Y.; Duan, Z.; Xu, Y. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression. *Renew. Energy* **2019**, *143*, 842–854. [CrossRef]
- Bojang, P.O.; Yang, T.C.; Pham, Q.B.; Yu, P.S. Linking Singular Spectrum Analysis and Machine Learning for Monthly Rainfall Forecasting. *Appl. Sci.* **2020**, *10*, 3224. [CrossRef]
- Wang, J.; Peng, X.; Wu, J.; Ding, Y.; Ali, B.; Luo, Y.; Hu, Y.; Zhang, K. Singular spectrum analysis (SSA) based hybrid models for emergency ambulance demand (EAD) time series forecasting. *IMA J. Manag. Math.* **2023**, *35*, 45–64. [CrossRef]
- Schmidt, E. Zur Theorie der linearen und nichtlinearen Integralgleichungen. *Math. Ann.* **1907**, *63*, 433–476. [CrossRef]
- Eckart, C.; Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1936**, *1*, 211–218. [CrossRef]
- Heinig, G.; Rost, K. *Algebraic Methods for Toeplitz-like Matrices and Operators (Operator Theory: Advances and Applications)*; Birkhäuser: Basel, Switzerland, 1985.
- Gantmacher, F.R. *The Theory of Matrices*; Chelsea Publishing Company: New York, NY, USA, 1959; Volume 2, p. 68.
- Hall, M. *Combinatorial Theory*; Wiley-Interscience: New York, NY, USA, 1998.
- Markovsky, I. *Low Rank Approximation: Algorithms, Implementation, Applications*; Springer: Cham, Switzerland, 2019.
- Usevich, K.; Markovsky, I. Variable projection for affinely structured low-rank approximation in weighted 2-norms. *J. Comput. Appl. Math.* **2014**, *272*, 430–448. [CrossRef]
- Gillard, J.W.; Usevich, K. Hankel low-rank approximation and completion in time series analysis and forecasting: A brief review. *Stat. Interface* **2023**, *16*, 287–303. [CrossRef]
- Zvonarev, N.; Golyandina, N. Fast and stable modification of the Gauss–Newton method for low-rank signal estimation. *Numer. Linear Algebra Appl.* **2021**, *29*, 2428. [CrossRef]
- Cadzow, J.A. Signal enhancement: A composite property mapping algorithm. *IEEE Trans. Acoust.* **1988**, *36*, 49–62. [CrossRef]

23. Gillard, J. Cadzow's basic algorithm, alternating projections and singular spectrum analysis. *Stat. Interface* **2010**, *3*, 335–343. [[CrossRef](#)]
24. Zvonarev, N.; Golyandina, N. Mixed Alternating Projections with Application to Hankel Low-Rank Approximation. *Algorithms* **2022**, *15*, 460. [[CrossRef](#)]
25. Korobeynikov, A. Computation- and space-efficient implementation of SSA. *Stat. Interface* **2010**, *3*, 357–368. [[CrossRef](#)]
26. Golyandina, N.; Dudnik, P.; Shlemov, A. Intelligent Identification of Trend Components in Singular Spectrum Analysis. *Algorithms* **2023**, *16*, 353. [[CrossRef](#)]
27. Chakrabarti, A.; Ghosh, J.K. AIC, BIC and Recent Advances in Model Selection. In *Philosophy of Statistics*; Bandyopadhyay, P.S., Forster, M.R., Eds.; Handbook of the Philosophy of Science; North-Holland: Amsterdam, The Netherlands, 2011; Volume 7, pp. 583–605. [[CrossRef](#)]
28. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
29. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
30. Rissanen, J. Modeling by shortest data description. *Automatica* **1978**, *14*, 465–471. [[CrossRef](#)]
31. Wax, M.; Kailath, T. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech Signal Process.* **1985**, *33*, 387–392. [[CrossRef](#)]
32. Badeau, R.; David, B.; Richard, G. A new perturbation analysis for signal enumeration in rotational invariance techniques. *IEEE Trans. Signal Process.* **2006**, *54*, 450–458. [[CrossRef](#)]
33. Papy, J.M.; De Lathauwer, L.; Van Huffel, S. A Shift Invariance-Based Order-Selection Technique for Exponential Data Modelling. *IEEE Signal Process. Lett.* **2007**, *14*, 473–476. [[CrossRef](#)]
34. Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [[CrossRef](#)]
35. Guillemin, V.; Pollack, A. *Differential Topology*; AMS Chelsea Publishing: New York, NY, USA, 1974.
36. Kwiatkowski, D.; Phillips, P.C.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econom.* **1992**, *54*, 159–178. [[CrossRef](#)]
37. Golyandina, N.; Dudnik, P. Extraction and Forecasting of Trends in Cases of Signal Rank Overestimation. *Eng. Proc.* **2024**, *68*, 20. [[CrossRef](#)]
38. Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B (Methodol.)* **1964**, *26*, 211–252. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.