

Original Paper

Automatic Near Real-Time Outlier Detection and Correction in Cardiac Interbeat Interval Series for Heart Rate Variability Analysis: Singular Spectrum Analysis-Based Approach

Michael Lang, MSc

Graduate School of Excellence Computational Engineering, Technische Universität Darmstadt, Darmstadt, Germany

Corresponding Author:

Michael Lang, MSc

Graduate School of Excellence Computational Engineering

Technische Universität Darmstadt

Dolivostraße 15

Darmstadt, 64293

Germany

Phone: 49 61511624401

Fax: 49 61511624404

Email: michael.lang@ieee.org

Abstract

Background: Heart rate variability (HRV) is derived from the series of R-R intervals extracted from an electrocardiographic (ECG) measurement. Ideally all components of the R-R series are the result of sinoatrial node depolarization. However, the actual R-R series are contaminated by outliers due to heart rhythm disturbances such as ectopic beats, which ought to be detected and corrected appropriately before HRV analysis.

Objective: We have introduced a novel, lightweight, and near real-time method to detect and correct anomalies in the R-R series based on the singular spectrum analysis (SSA). This study aimed to assess the performance of the proposed method in terms of (1) detection performance (sensitivity, specificity, and accuracy); (2) root mean square error (RMSE) between the actual N-N series and the approximated outlier-cleaned R-R series; and (3) how it benchmarks against a competitor in terms of the relative RMSE.

Methods: A lightweight SSA-based change-point detection procedure, improved through the use of a cumulative sum control chart with adaptive thresholds to reduce detection delays, monitored the series of R-R intervals in real time. Upon detection of an anomaly, the corrupted segment was substituted with the respective outlier-cleaned approximation obtained using recurrent SSA forecasting. Next, N-N intervals from a 5-minute ECG segment were extracted from each of the 18 records in the MIT-BIH Normal Sinus Rhythm Database. Then, for each such series, a number (randomly drawn integer between 1 and 6) of simulated ectopic beats were inserted at random positions within the series and results were averaged over 1000 Monte Carlo runs. Accordingly, 18,000 R-R records corresponding to 5-minute ECG segments were used to assess the detection performance whereas another 180,000 (10,000 for each record) were used to assess the error introduced in the correction step. Overall 198,000 R-R series were used in this study.

Results: The proposed SSA-based algorithm reliably detected outliers in the R-R series and achieved an overall sensitivity of 96.6%, specificity of 98.4% and accuracy of 98.4%. Furthermore, it compared favorably in terms of discrepancies of the cleaned R-R series compared with the actual N-N series, outperforming an established correction method on average by almost 30%.

Conclusions: The proposed algorithm, which leverages the power and versatility of the SSA to both automatically detect and correct artifacts in the R-R series, provides an effective and efficient complementary method and a potential alternative to the current manual-editing gold standard. Other important characteristics of the proposed method include the ability to operate in near real-time, the almost entirely model-free nature of the framework which does not require historical training data, and its overall low computational complexity.

(*JMIR Biomed Eng* 2019;4(1):e10740) doi: [10.2196/10740](https://doi.org/10.2196/10740)

KEYWORDS

change-point detection; cumulative sum; forecasting; heart rate variability; R-R series; singular spectrum analysis; ventricular premature complexes

Introduction

Background on Heart Rate Variability

Oscillations in the time interval between successive heart beats, referred to as heart rate variability (HRV), have long been known to allow for insight into the intricate control mechanisms of the autonomic nervous system (ANS) [1-4]. Accordingly, research into HRV has attracted considerable attention over the past 4 decades, as evidenced by an exponential growth of published work [5,6].

In a nutshell, heart rhythm and rate can be said to be governed by the 2 competing divisions of the ANS, that is, the sympathetic nervous system and the parasympathetic nervous system. An increased sympathetic input to the sinoatrial (SA) node yields an increase in heart rate (HR) mediated by the release of adrenaline and noradrenaline and the subsequent activation of β -1-adrenoceptors, resulting in an accelerated diastolic depolarization. On the other hand, an increase in the parasympathetic outflow decreases the HR through the release of acetylcholine by the vagus nerve, which binds to and activates M2 muscarinic acetylcholine receptors in the SA node and eventually slows down the diastolic depolarization rate [4,7-9].

Various HRV measures have been established and are usually categorized as either time domain, frequency domain, or nonlinear analysis techniques [2,4]. To a large extent, the popularity of HRV is because of its easy acquisition and seemingly straightforward interpretation.

The Necessity of Preprocessing the R-R Series

It is crucial to realize that by virtue of its very definition, the HRV encompasses only oscillatory phenomena between heart beats resulting from the SA node depolarization [4,10]. While, ideally, all components of an R-R series originate from the SA node, the actual R-R series in both healthy and diseased subjects are contaminated by outliers due to artifacts and heart rhythm disturbances such as ectopic beats (ie, heart beats whose origin is outside of the SA node). Thus arises the necessity to ensure that the HRV analysis is performed on a series representing only the actual normal sinus rhythm (NSR) interbeat intervals, commonly referred to as N-N series (as in normal-to-normal).

The detrimental impact of ectopics on HRV measures is pronounced and well-documented [6,11-17]. In a recent study Stapelberg et al [18] examined the sensitivity of 38 time domain, frequency domain and nonlinear HRV measures to the addition of artifacts in real and artificial 24-hour recordings. In accordance with previous findings, they concluded short-term time domain HRV measures to be more sensitive to the presence of artifacts than their long-term counterparts. Furthermore, frequency domain measures were found to generally be more sensitive than time domain measures, whereas the less commonly used nonlinear measures were shown to exhibit some inherent robustness properties.

Ectopic heart beats are ubiquitous phenomena and not limited to patients with cardiac disorders and diseases. Hingorani et al [19] retrospectively examined the prevalence of cardiac arrhythmias by scrutinizing 24-hour Holter recordings of 1273 healthy volunteers from 22 phase I clinical trials; note that this specific population is not a representative sample of the overall population. The sample was heavily biased in that, consistent with the requirements of phase I clinical trials, subjects were extensively screened by history, physical examination, and laboratory tests and were therefore significantly healthier than the overall population. Crucially, all subjects underwent 12-lead pretrial electrocardiography (ECG) and were excluded if they had any cardiac conduction disorder or disease, including a personal or familiar history thereof. Those exhibiting ≥ 2 consecutive ectopics as well as bigeminy, trigeminy and quadrigeminy in their 12-lead ECG examination were also excluded. Despite these rigorous exclusion criteria, Hingorani et al [19], among other findings, found premature atrial complexes in 60.8% of the examined Holter records, followed by premature ventricular complexes (PVCs), which were observed in 43.4% of healthy volunteers. While multifocal PVCs occurred in 5.3%, 3.3% exhibited >200 PVCs per 24-hour ECG recordings. A relatively high prevalence of occasional ectopic beats has been reported by numerous other authors as well both in healthy and diseased subjects [20-23].

In summary, the detection and correction of ectopic beats in tachograms are not merely imperative for formal consistency with the HRV definition but arises from the impact of ectopics on HRV measures and the general prevalence of ectopics.

Prior Work and State-of-the-Art

Compared with the overall research interest directed toward HRV applications, the crucial issue of the discrepancy between most real-world R-R series and their respective N-N series has arguably not received as much attention. The actual gold standard of manual R-R series editing, advocated for in the field's relevant guideline by the European Society of Cardiology and the North American Society of Pacing and Electrophysiology Task Force [4], remains unaltered to this day [11,24].

It should be noted that, aiding and abetting the aforementioned exponential growth of the HRV-related (applied) research, a number of well-designed and widely used software packages for the HRV analysis have been developed and made available to the general public; these include, but are not limited to, Kubios [25,26], Nevrokard's aHRV [27], and others [28-30]. More often than not, these commercial software packages tend to be closed source and therefore provide only a limited benefit to algorithm developers. On the other hand, it ought to be acknowledged that they are well suited for the needs of practitioners, as they combine advanced preprocessing and analysis features, intuitive graphical interfaces, and support for various file formats.

In automated approaches the task of the detection and correction of outliers is usually tackled as a two-step process.

First, outliers in the R-R tachogram are detected, usually by straightforward thresholding based on a fixed percentage difference to previous intervals or their SD [31], although some more sophisticated methods have also been proposed [12].

Second, the detected (regions of) artifacts from step 1 are either discarded (thereby reducing the effective length of the tachogram), substituted by preceding uncorrupted intervals or by intervals obtained through interpolation, with linear and cubic spline interpolation being the most popular techniques [11].

Objective of This Work

The objective of this study was to introduce a general framework of low computational complexity that allows for an automatic near real-time detection and correction of outliers in the R-R series based on the singular spectrum analysis (SSA). The main novel contributions of this work include (1) the application of a recently proposed model-free lightweight SSA change-point detection (l-SSA-CPD) algorithm [32]; (2) a modification of l-SSA-CPD through the use of an adaptive control limit sequential ranks (AC-SRC) control chart [33] to drastically reduce detection delays; and (3) upon detection of an anomaly, the substitution of the corrupted tachogram segment by an approximation obtained through recurrent SSA forecasting based on a small outlier-free tachogram segment.

An extensive simulation study comprising 198,000 5-minute long R-R series, obtained by randomly inserting varying amounts of simulated ectopic beats in records taken from the MIT-BIH Normal Sinus Rhythm Database (NSRDB), is conducted to assess the artifact detection and correction performance of the proposed algorithm.

Fundamentals of Basic Univariate Singular Spectrum Analysis and Singular Spectrum Analysis Based Change-Point Detection

Recently, this author proposed a low-complexity model-free approach based on SSA and nonparametric cumulative sum (CUSUM) control charts for real-time cardiac anomaly detection, referred to as l-SSA-CPD [32]. It was shown that l-SSA-CPD reliably detects anomalies even when directly applied to unprocessed (ie, no preprocessing was performed) raw ECG and photoplethysmographic records from common databases publicly available through Physiobank [34].

In this study, modifications to the original l-SSA-CPD algorithm will be introduced, and its capabilities will be expanded by adding a recurrent forecasting feature to generate appropriate substitutes of the corrupted tachogram entries. First, however, a brief restatement of SSA and l-SSA-CPD [32] is required.

Basic Singular Spectrum Analysis Algorithm

SSA is a powerful technique of time series analysis owing much of its appeal to an inherent model-free concept combining elements of conventional time series analysis, multivariate geometry and statistics, and signal processing [35]. The univariate basic SSA can be considered as the application of

the principal component analysis to the Hankel matrix (obtained through an embedding of the original univariate time series) with the subsequent attempt to reconstruct the original series.

Consider N observations $\mathbb{X}_N = (x_1, \dots, x_N)$ of a univariate time series and a lag-integer or embedding dimension M ($1 < M < N$) also commonly referred to as the window length. The basic SSA algorithm then encompasses the following 4 steps:

1. Embedding $\mathbb{X}_N = (x_1, \dots, x_N) \rightarrow \mathbf{X} \in \mathbb{R}^{M \times K}$

A trajectory matrix \mathbf{X} is constructed by mapping \mathbb{X}_N into a sequence of $K = N - M + 1$ lagged column vectors $\mathbf{X}_j = (x_j, \dots, x_{j+M-1})^T, j = 1, \dots, K$ of size M , yielding

$$\mathbf{X} = \begin{bmatrix} x_{n+1} & x_{n+2} & \cdots & x_{n+K} \\ x_{n+2} & x_{n+3} & \cdots & x_{n+K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+M} & x_{n+M+1} & \cdots & x_{n+N} \end{bmatrix}. \quad (1)$$

\mathbf{X} is called a Hankel matrix because of its characteristic of having equal elements on the antidiagonals. One can think of \mathbf{X} as multivariate data with M characteristics and K observations and accordingly \mathbf{X}_j of \mathbf{X} as column vectors in the M -dimensional space \mathbb{R}^M .

2. Singular Value Decomposition of \mathbf{X}

Taking the singular value decomposition (SVD) of \mathbf{X} decomposes the trajectory matrix into its orthogonal bases and yields a set of M eigenvalues and eigenvectors. Let the eigenvalues of $\mathbf{X}\mathbf{X}^T$ be $\lambda_1 \geq \dots \geq \lambda_M \geq 0$ and U_1, \dots, U_M be the respective eigenvectors. Then, with $d = \max(i, \text{such that } \lambda_i > 0)$ denoting the rank of \mathbf{X} , the SVD of \mathbf{X} can be expressed as the sum of d elementary matrices

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d, \quad (2)$$

with rank 1 matrices $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ and $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ being the eigenvectors of $\mathbf{X}^T \mathbf{X}$. Accordingly, the decomposition in Equation (2) is completely characterized by the set of so-called ‘‘eigen triples’’ $(\sqrt{\lambda_i}, U_i, V_i)$.

Note that owing to the symmetry of left and right singular vectors, the SVDs of \mathbf{X} with lag M and $K = N - M + 1$ are equivalent, entailing the lack of any additional benefit from the use of larger embedding dimensions $M > N / 2$ [32,35-39].

3. Eigen triple Grouping

Consider the task of extracting a particular component or signal of interest from observed data contaminated by various artifacts such as noise. To do so, during the third stage of the basic SSA, disjoint subsets of indices $\{1, \dots, d\}$ are determined such that the respective systems of eigenvectors span the subspaces associated with those signal components.

In the example of aiming at the separation of a signal from unwanted noise and other disturbances, this entails determining

an appropriate subset of indices $I = \{i_1, \dots, i_l\}$, $l < d \leq M$ that span an l -dimensional subspace in \mathbb{R}^M , denoted as $\mathbf{X}_I \subset \mathbb{R}^M = \text{span}\{U_I\} = \text{span}\{U_{i_1}, \dots, U_{i_l}\}$, representing the signal whereas the remaining eigentriples with $\bar{I} = \{i_1, \dots, i_d\} / I$ are said to span the noise subspace $\mathbf{X}_{\bar{I}} \subset \mathbb{R}^M = \text{span}\{U_{\bar{I}}\}$.

The trajectory matrix component \mathbf{X}_I corresponding to the subset I of eigentriples associated with the signal of interest is then

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_l}, \quad (3)$$

and the component $\mathbf{X}_{\bar{I}}$ corresponding to the subset $\bar{I} = \{i_1, \dots, i_d\} / I$ associated with the remainder of the observed signal is

$$\mathbf{X}_{\bar{I}} = \sum_{i \in \bar{I}} \mathbf{X}_i, \quad (4)$$

such that the overall SVD of \mathbf{X} can be rewritten as

$$\mathbf{X} = \mathbf{X}_I + \mathbf{X}_{\bar{I}} = \sum_{i \in I} \mathbf{X}_i + \sum_{i \in \bar{I}} \mathbf{X}_i. \quad (5)$$

In the case of separability (see, eg, [36] at 17), the contribution of \mathbf{X}_I to the entire observed signal \mathbf{X} is represented by the respective share of eigenvalues $\frac{\sum_{i \in I} \lambda_i}{\sum_{i=1}^d \lambda_i}$.

4. Diagonal Averaging

For perfectly separable signal components, all matrices in the expansion of Equation (5) are Hankel matrices, which require that they all have equal entries on their antidiagonals. In real-world problems, however, perfect separability is rarely achievable. While an approximate separability usually suffices, the final fourth step in the SSA algorithm is required, as the matrices \mathbf{X}_i from step 3 would have unequal entries on their antidiagonals. Thus, the Hankelization of all \mathbf{X}_i is performed, that is, the compliance with the Hankel structure is enforced by taking the average of each antidiagonal and then replacing each element of that antidiagonal with the determined average. This yields

$$\tilde{\mathbf{X}} = \tilde{\mathbf{X}}_I + \tilde{\mathbf{X}}_{\bar{I}} = \sum_{i \in I} \tilde{\mathbf{X}}_i + \sum_{i \in \bar{I}} \tilde{\mathbf{X}}_i, \quad (6)$$

with $\tilde{\mathbf{X}}_i$ being the diagonally averaged matrices.

One can then, for example, easily reconstruct the approximation of the signal of interest through the eigentriples with indices I through the one-to-one correspondence between $\tilde{\mathbf{X}}_I$ and the respective time series $\tilde{\mathbf{X}}_N = (\tilde{x}_1, \dots, \tilde{x}_N)$ which provides an approximation of either the entire time series \mathbf{X}_N or some components of it, depending on the particular choice of indices I .

Singular Spectrum Analysis Based Change-Point Detection Rationale

The sequential application of SSA for change detection is based on Moskvina and Zhigljavsky [40] (see [39] for an exhaustive discussion; see also [35] for an earlier account of the basic idea).

The rationale is to slide 2 (possibly intersecting) moving windows over the time series and to apply the first 3 stages of the basic SSA each time. Think of the first window as adaptively generating a low-dimensional base subspace representation (capturing the main structure in the series) and the second window, which contains at least M new data samples, as generating a test subspace. As both windows are slid over the series, some sort of distance between the base and the test subspace is monitored. If no change-point occurs, the distance remains small, whereas it spikes in the presence of a sudden significant change in the main structure of the series.

The algorithm by Moskvina and Zhigljavsky has been successfully applied to various change detection problems and some improved variations of it have been presented as well (see [32] for a detailed background). Computational complexity, however, ought to be acknowledged as a potentially significant limitation of the said algorithm. Recently, 1-SSA-CPD [32] has been proposed as a lightweight alternative. Contrary to the algorithm by Moskvina and Zhigljavsky, which recalculates the SVD of 2 sliding windows each time new observations become available, 1-SSA-CPD relies on a nominal low-dimensional signal subspace representation computed only once at startup. As new observations become available, 1-SSA-CPD simply allocates them into a vector and calculates a test statistic based on the squared Euclidean distance and the angle between this new data vector and the previously determined nominal reference subspace.

Singular Spectrum Analysis Parameter Selection

The behavior of SSA is largely determined by 2 tuning parameters, namely the window length M and the number l of eigenvalues to be used in the embedding and grouping steps, respectively. While it is undisputed that the successful application of SSA hinges on an appropriate choice of M and l , practitioners are faced with the nuisance that no clear-cut rules as to how they ought to be determined exist.

For a series of length N and window length (or embedding dimension) $1 < M < N$, as previously discussed, there is no additional benefit from using $M > N/2$ (see, eg, [35] at 69 and [36] at 47). It is widely acknowledged that as the appropriate window length M depends on both the underlying data and the particular application, no universal rule for determining it exists [35-39]. General recommendations vary between $M \approx N/4$ [37] and $M \approx N/2$ [35,36]. In the case of a periodic signal with period T , it is important for M to be, at least, equal to T for the SSA to capture the main structure of the series. Furthermore, M should be proportional to T [35,36,38]. Of note, small windows act like a smoothing filter of width $2M - 1$ [36].

Similarly, it is well known that selecting a proper l is also crucial, for issues are known to arise when eigentriple grouping comprises either an insufficient or an excessive number of eigenvalues. More specifically, if l is too small, the SSA is unable to capture the entire signal whereas if it is too large noise components are unwittingly captured. Since the contribution of a particular series component (associated with the corresponding eigentriple) to the entire series is proportional to the respective eigenvalue, it is common practice to select l by identifying the leading eigenvalues through visual inspection of the eigenvalue

spectra or by setting a predefined eigenvalue share (see, eg, [35,36,38] for a more exhaustive discussion).

In addition to the above-referred monographs, interested readers are also referred to some fairly recent contributions by Hassani et al [41] and Khan and Poskitt [42,43], which approach the issue through the concept of separability and information theory, respectively.

In the light of the pragmatic suboptimal approach pursued here, a detailed discussion of the SSA parameter selection is, however, beyond the scope of this study and therefore omitted.

Methods

The Proposed Outlier Detection and Correction Framework

The proposed framework consisted of 2 separate stages, namely the detection and the correction of outlier corrupted segments of the tachogram.

Outlier Detection Using Lightweight Singular Spectrum Analysis Change-Point Detection and Adaptive Control Limit Sequential Ranks Control Charts

Let M, N, l, p, q be fixed integers such that $l < M < N/2$ and $0 \leq p < q$. Then 1-SSA-CPD proceeds as follows:

1. Initialization at $n = 0$

Akin to MZ, the first 3 steps of the basic SSA are applied on the interval $[n + 1, n + N]$ to obtain a low-dimensional nominal subspace $I = I^{(n=0)}$ which captures the main structure of the series. This involves embedding as in Equation (1) to obtain a trajectory matrix, which, hereafter, we refer to as the base matrix $X_B = X_B^{(0)} = X_B^{(n=0)}$, taking the SVD of X_B and obtaining I through an appropriate choice of $I = \{i_1, \dots, i_l\}$, $l < d \leq M$ with $d = \max(i, \text{such that } \lambda_i > 0)$.

2. Then, for each $n = 0, 1, \dots$ 1-SSA-CPD proceeds as follows:

- Construction of a test matrix $X_T^{(n)}$ on the interval $[n + p + 1, n + q + M - 1]$ as

$$X_T^{(n)} = \begin{bmatrix} x_{n+p+1} & x_{n+p+2} & \dots & x_{n+q} \\ x_{n+p+2} & x_{n+p+3} & \dots & x_{n+q+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+p+M} & x_{n+p+M+1} & \dots & x_{n+q+M-1} \end{bmatrix} \quad (7)$$

- Computation of the detection statistic $D_{n,I,p,q}$ as

$$D_{n,I,p,q} = D_{n,I,p,q}^{\dagger 1} \circ D_{n,I,p,q}^{\dagger 2}, \quad (8)$$

with

$$D_{n,I,p,q}^{\dagger 1} = \frac{1}{Q} \sum_{j=p+1}^q \left[\left(X_j^{(n)} \right)^T X_j^{(n)} - \left(X_j^{(n)} \right)^T U_I \left(U_I \right)^T X_j^{(n)} \right]$$

$$D_{n,I,p,q}^{\dagger 2} = 1 - \cos \left[\angle \left(X_T^{(n)}, \mathcal{L}_I \right) \right],$$

and with the angle $\angle \left(X_T^{(n)}, I \right)$ taking values in $[0, \pi / 2]$ and accordingly $D_{n,I,p,q}^{\dagger 1} \in [0, 1]$. $X_j^{(n)} = [x_{n+j}, \dots, x_{n+j+M-1}]^T$ are the column vectors of $X_T^{(n)}$ whereas $U_I = [U_{i_1}, \dots, U_{i_l}]$ denotes the $M \times l$ matrix of eigenvectors spanning I and \circ denotes the Hadamard (or element-wise) product. Note that throughout this study, $p = N, q = N + 1$ is used, that is, the width $Q = q \times p$ of the test matrices $X_T^{(n)}$ used always equals 1. In other words, each test matrix is actually a single column vector containing M new observations. $Q = 1$ is primarily chosen to allow for an agile response time and to minimize the computational burden [32].

- Monitoring of $D_{n,I,p,q}$ using the AC-SRC Control Chart

To reduce detection delays, instead of the sequential ranks CUSUM (SRC) by McDonald [44], as used in the conventional 1-SSA-CPD [32], a sequential ranks-based CUSUM with adaptive control limits referred to as AC-SRC [33] is used. AC-SRC is inspired by a distribution-free CUSUM proposed by Chatterjee and Qiu [45] where instead of a fixed threshold a sequence of control limits, determined by the bootstrap estimate of the conditional distribution of the CUSUM test statistic given the last time it was zero, is used. AC-SRC carries the approach of Chatterjee and Qiu over to McDonald's SRC; the main advantage of the SRC (which is that it does not require training data and control limits can thus easily be determined in advance, eg, through Monte Carlo simulations) is retained while detection delays are significantly reduced because of the use of an adaptive sequence of control limits. Furthermore, owing to a trade-off favoring the lower computational burden and simplicity over the optimal performance, AC-SRC lacks some of the refined fine-tuning routines used in the Chatterjee and Qiu method. The author refrains from a detailed discussion of AC-SRC and refers the interested reader to previous studies [33,45].

Let the sequential rank of $D_{n,I,p,q}$ be denoted as

$$R_n = 1 + \sum_{r=1}^{n-1} \max(0, D_{n,I,p,q} - D_{r,I,p,q}) \quad (9)$$

The AC-SRC then starts with the same recursive equation as the SRC, ie,

$$C_n = \max \left(0, C_{n-1} + \frac{R_n}{n+1} - k \right), \quad n \geq 1 \quad (10)$$

with $C_0 = 0$ and k a reference constant. Let T_n denote the so-called "sprint length" expressing the time elapsed since C_n was last 0, ie,

$$T_n = \begin{cases} 0 & \text{if } C_n = 0 \\ j & \text{if } C_n \neq 0, \dots, C_{n-j+1} \neq 0, \\ & C_{n-j} = 0; \quad j = 1, \dots, n \end{cases}$$

Furthermore, let Y_j be a random variable with distribution

$$Y_j \sim C_n | T_n = j \quad (11)$$

Chatterjee and Qiu showed that the conditional distributions of $C_n | T_n$ in Equation (11) are easier to handle than the unconditional distribution of C_n and under some regularity conditions (see [45] and references therein for details), depend only on j and the underlying process generating distribution, but not on n . Then, for any positive integer $j_{\max} \leq n$, the distribution of C_n can be approximated by means of the conditional distributions in Equation (11) as

$$C_n \sim \sum_{j=1}^{j_{\max}} Y_j I_{T_n=j} + Y^* I_{T_n > j_{\max}}, \quad (12)$$

with $Y^* \sim C_n | T_n > j_{\max}$ and I being the indicator function.

It can be shown that [44], given the observed process is in-control, the quantities $R_n / (n + 1)$ in Equation (10) are independent and discrete uniform on $\{\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}\}$. Herein lies the key advantage of AC-SRC, in that for some fixed j_{\max} and k one can determine a sequence of control limits $\{h_{AC-SRC_j}\}_{j=1}^{j_{\max}}$ from

$$Y_j \sim [\tilde{C}_n | T_n = j], \quad (13)$$

with \tilde{C}_n representing a simulation of Equation (10) for up to some N_{AC-SRC} wherein the quantities $R_n / (n + 1)$ are substituted by the set of random variables $\{Z_n\}_{n=1}^{N_{AC-SRC}}$ as discrete uniform on $\{\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}\}$. The sequence of control limits $\{h_{AC-SRC_j}\}_{j=1}^{j_{\max}}$ can then easily and without the need for training data samples be determined by means of Monte Carlo simulations as outlined elsewhere [33].

The AC-SRC signals a change to have occurred if

$$T_n = j \quad \text{and} \quad C_n > h_{AC-SRC_j}, \quad \text{for } 1 \leq j \leq j_{\max},$$

or if

$$T_n > j_{\max} \quad \text{and} \quad C_n > h_{AC-SRC_{j_{\max}}}.$$

To limit the computational burden, it is recommended to calculate control limits only up to a reasonably small j_{\max} after which the control limit is kept fixed at $h_{AC-SRC_{j_{\max}}}$. In addition, following recommendations by Chatterjee and Qiu, the reference constant k is set proportionate to the expected average sprint length $E\{T_n\}$. Throughout this work, k was empirically calibrated such that $k \approx \lfloor \frac{3j_{\max}}{4} \rfloor$, whereas the average run length

(ARL) was set to a nominal $ARL_0=500$. A detailed discussion as to the empirical calibration of control charts would, however, be beyond the scope of this work [33]. Note that as k increases, accordingly j_{\max} decreases $E\{T_n\}$, that is, the likelihood of C_n bouncing back to 0 increases [33,45]. For the particular use case examined here, the use of a small j_{\max} to allow for agile change detection is recommended (see Tables 1-4 and the respective discussion in the next section).

Finally, as with the SRC in the conventional 1-SSA-CPD algorithm, the control chart is reinitialized after a change has been detected (by setting $C_n=0$ and $T_n=0$) to allow for the detection of multiple and potentially nearby change-points [32].

Monitoring of the 1-SSA-CPD test statistic $D_{n,l,p,q}$ with $N=20$, $M=10$ by means of AC-SRC with $j_{\max}=8$, $E\{T_n\}=6$ is showcased on a 5-minute tachogram excerpt of record 16,273 MIT-BIH NSRDB artificially contaminated with 2 and 8 ectopics in Figures 1 and 2, respectively. Note the remarkably low detection delay due to the use of AC-SRC.

Outlier Correction Using Recurrent Singular Spectrum Analysis Forecasting

Once an outlier has been detected at say $n=\tau$, accounting for the inherent detection delay and accordingly an uncertainty as to the exact outlier location, the interval $[\tau - \Delta_1, \tau]$, is designated as corrupted tachogram segment. Note that Δ_1 depends on the (average) detection delay and should be chosen carefully. However, it need not be exact or otherwise optimal. Then, a second interval of length Δ_2 immediately preceding the corrupted one is used to obtain an anomaly-free forecast to substitute the corrupted segment with.

The rationale of the proposed framework is depicted in Figure 3, with green and yellow rectangles representing the designated uncorrupted and corrupted segments and the detected change highlighted in red.

As the proposed method operates sequentially as new observations arrive, an appropriate uncontaminated data segment (either uncorrupted to begin with or previously cleaned) to serve as a basis for forecasting is always available. While it may be beneficial to use large values for Δ_2 (assuming sufficient past observations were available), keeping Δ_1, Δ_2 as small as possible is recommended; as doing so allows for near real-time operation and significantly reduces the computational burden. For the purpose of this study, simplicity was prioritized over (more often than not unnecessary) optimality.

Table 1. Detection performance of adaptive control limit sequential ranks with $j_{max}=6$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

Record	Sensitivity	Specificity	Accuracy
16,265	0.9712	0.9827	0.9825
16,272	0.9415	0.9802	0.9794
16,273	0.9905	0.9910	0.9909
16,420	0.9771	0.9778	0.9777
16,483	0.9953	0.9846	0.9846
16,539	0.9326	0.9878	0.9870
16,773	0.9692	0.9832	0.9829
16,786	0.9722	0.9758	0.9755
16,795	0.8800	0.9922	0.9905
17,052	0.9521	0.9830	0.9825
17,453	0.9768	0.9904	0.9902
18,177	0.9549	0.9757	0.9754
18,184	0.9719	0.9805	0.9803
19,088	0.9711	0.9917	0.9914
19,090	0.9900	0.9745	0.9745
19,093	0.9744	0.9837	0.9834
19,140	0.9689	0.9920	0.9917
19,830	0.9948	0.9909	0.9909

Table 2. Detection performance of adaptive control limit sequential ranks with $j_{max}=8$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

Record	Sensitivity	Specificity	Accuracy
16,265	0.9335	0.9860	0.9854
16,272	0.8595	0.9850	0.9828
16,273	0.9199	0.9922	0.9913
16,420	0.9141	0.9810	0.9803
16,483	0.9569	0.9893	0.9888
16,539	0.7494	0.9889	0.9863
16,773	0.8781	0.9858	0.9844
16,786	0.9038	0.9779	0.9767
16,795	0.7210	0.9936	0.9897
17,052	0.8259	0.9851	0.9829
17,453	0.9013	0.9933	0.9921
18,177	0.8101	0.9785	0.9771
18,184	0.8508	0.9833	0.9819
19,088	0.8667	0.9924	0.9910
19,090	0.9403	0.9793	0.9789
19,093	0.8920	0.9869	0.9854
19,140	0.8796	0.9946	0.9933
19,830	0.9510	0.9945	0.9941

Table 3. Detection performance of adaptive control limit sequential ranks with $j_{max}=12$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

Record	Sensitivity	Specificity	Accuracy
16,265	0.8816	0.9875	0.9864
16,272	0.8240	0.9871	0.9844
16,273	0.8710	0.9940	0.9925
16,420	0.8741	0.9850	0.9839
16,483	0.8990	0.9900	0.9889
16,539	0.7218	0.9908	0.9879
16,773	0.8202	0.9871	0.9850
16,786	0.8421	0.9814	0.9794
16,795	0.6619	0.9943	0.9898
17,052	0.7753	0.9893	0.9865
17,453	0.8450	0.9939	0.9921
18,177	0.7539	0.9824	0.9805
18,184	0.7952	0.9853	0.9832
19,088	0.8030	0.9930	0.9909
19,090	0.8869	0.9824	0.9814
19,093	0.8379	0.9892	0.9869
19,140	0.8004	0.9951	0.9931
19,830	0.9063	0.9948	0.9939

Table 4. Detection performance of adaptive control limit sequential ranks with $j_{max}=16$ and lightweight singular spectrum analysis change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples.

Record	Sensitivity	Specificity	Accuracy
16,265	0.8623	0.9878	0.9866
16,272	0.7947	0.9885	0.9854
16,273	0.8774	0.9942	0.9929
16,420	0.8585	0.9851	0.9838
16,483	0.8950	0.9904	0.9893
16,539	0.7260	0.9920	0.9890
16,773	0.7951	0.9880	0.9857
16,786	0.8022	0.9814	0.9790
16,795	0.6913	0.9941	0.9900
17,052	0.7866	0.9905	0.9881
17,453	0.8176	0.9951	0.9932
18,177	0.7644	0.9841	0.9823
18,184	0.7846	0.9866	0.9847
19,088	0.8119	0.9936	0.9917
19,090	0.8733	0.9837	0.9827
19,093	0.8299	0.9898	0.9875
19,140	0.7949	0.9953	0.9934
19,830	0.8979	0.9952	0.9944

Figure 1. Adaptive control limits sequential ranks cumulative sum (AC-SRC) monitoring with $j_{max}=8$, $E\{T_n\}=6$, $ARL_0=500$ of $D_{n,l,p,q}$ with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples (top) for a 5-minute tachogram excerpt of record 16273 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 2 artificially inserted ectopics (bottom).

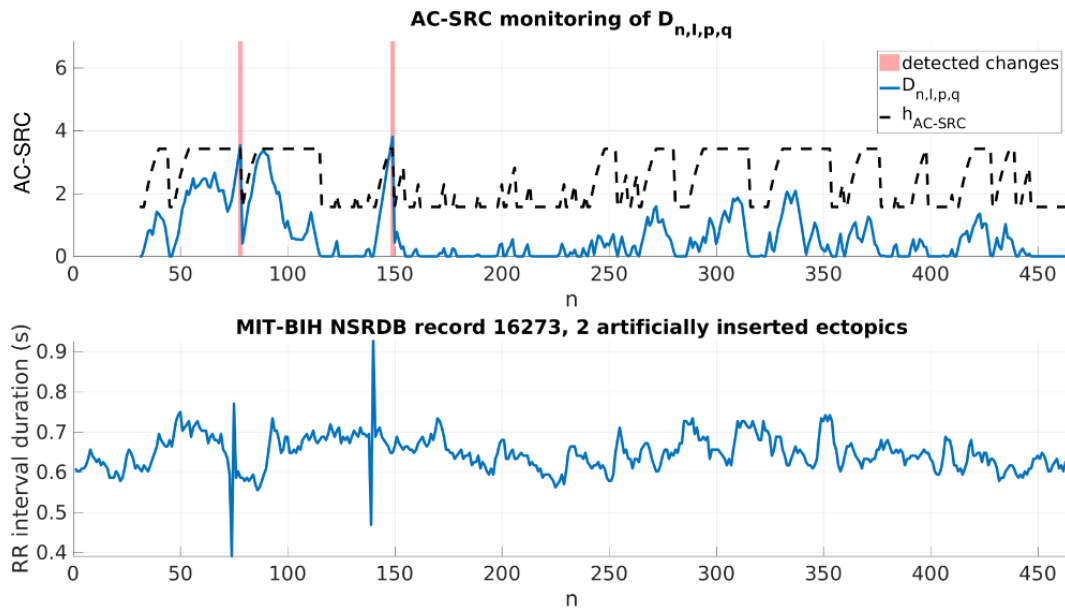


Figure 2. Adaptive control limits sequential ranks cumulative sum (AC-SRC) monitoring with $j_{max}=8$, $E\{T_n\}=6$, $ARL_0=500$ of $D_{n,l,p,q}$ with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples (top) for a 5-minute tachogram excerpt of record 16273 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 8 artificially inserted ectopics (bottom).

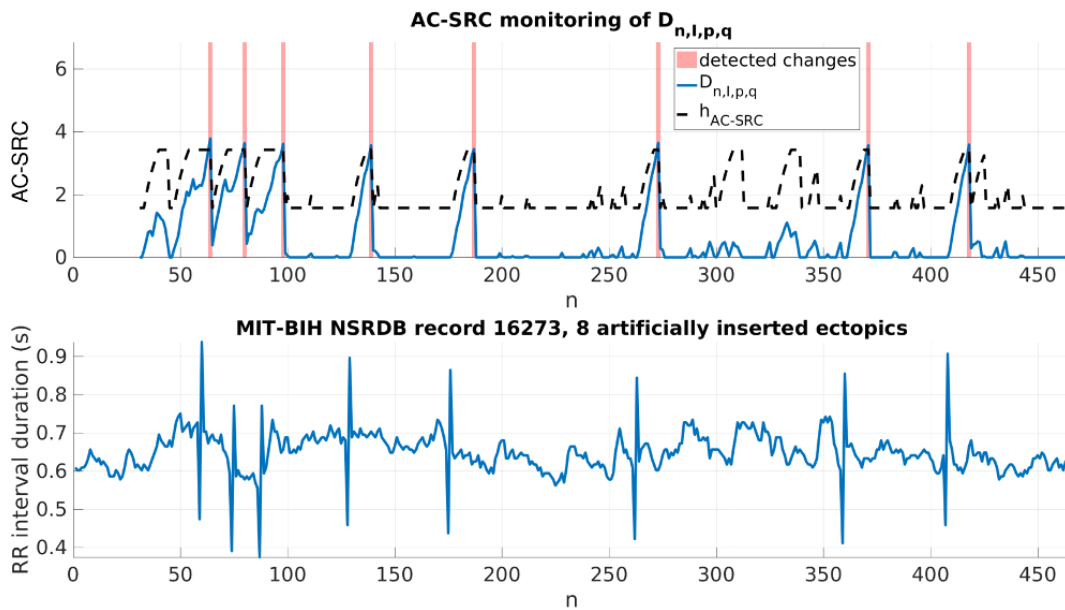
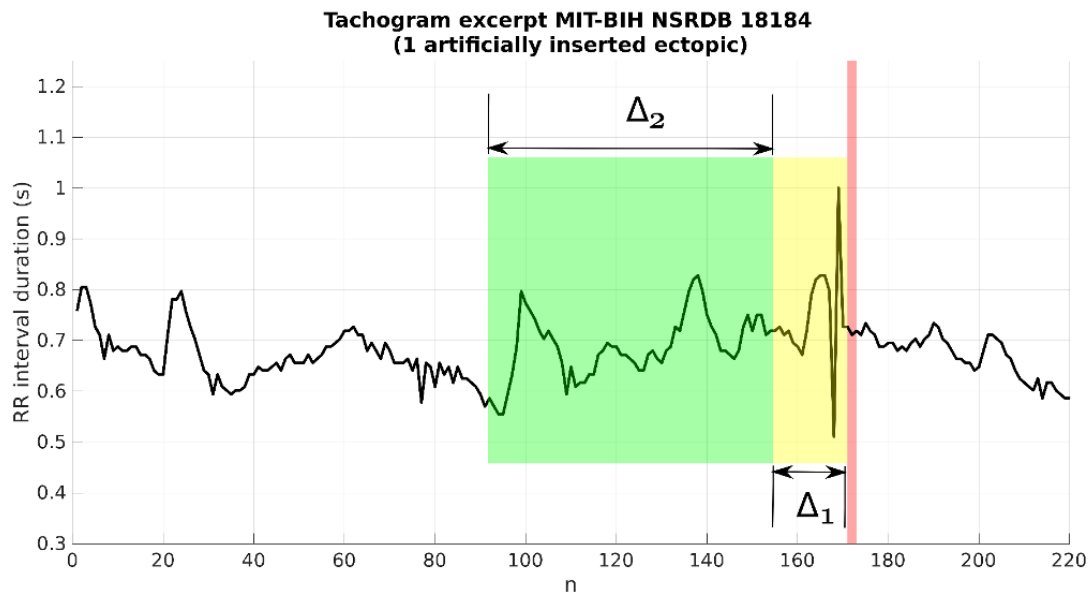


Figure 3. Schematic depiction of the proposed R-R series outlier detection and correction framework. Adaptive control limits sequential ranks cumulative sum with low detection delay is used to monitor lightweight singular spectrum analysis (SSA) change-point detection’s test statistic; when it signals (red vertical bar), a segment of length Δ_1 immediately preceding the detection is labeled as being corrupted (yellow rectangle) and a larger, yet still reasonably small segment of length Δ_2 (green rectangle) is used to construct a trajectory matrix for the recurrent SSA forecasting which eventually yields a short-term SSA forecast of length Δ_1 which the corrupted segment is then substituted with.



Time series forecasting plays a crucial role in many scientific fields and arguably also represents one of the most popular (real-world) application areas of the SSA. Note that there exist different approaches for SSA forecasting [46], most notably recurrent and vector forecasting. Refraining from any detailed discussion (see [35-38] for excellent monographs on the subject matter), the recurrent SSA forecasting approach used in this study will briefly be outlined. As with other more advanced capabilities that go beyond the realm of basic SSA, forecasting and missing value imputation (note that the 2 are closely related with the latter being the more general problem incorporating forecasting as a special case) in a way sacrifice some of the model-free beauty and appeal of SSA by invoking certain model assumptions. On the other hand, this appears to be outweighed by the successful application of SSA forecasting in numerous areas [47-52]. Furthermore, Golyandina and Zhigljavsky [36] point out that for short-term forecasts, there is actually only very little use of the imposed model, that is, the series approximately satisfy some linear recurrent formulas.

Consider again a univariate time series $\mathbb{X}_N = (x_1, \dots, x_N)$ of length N , a fixed embedding dimension M ($1 < M < N$) and the respective trajectory matrix $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_K]$ with $K = N - M + 1$ as in Equation (1). Let $\mathcal{r} \subset \mathbb{R}^M$ be a linear space of dimension $r < M$ and U_1, \dots, U_r (ie, the eigenvectors obtained through the SVD of \mathbf{X}) be an orthonormal basis in \mathcal{r} . Furthermore, let $\hat{x} = [\hat{x}_1, \dots, \hat{x}_K] = \sum_{i=1}^r U_i U_i^T \mathbf{x}$, that is, \hat{X}_i is the orthogonal projection of the column vector \mathbf{X}_i of \mathbf{X} onto \mathcal{r} , and its Hankelized counterpart $\tilde{\mathbf{x}} = \mathcal{H} \hat{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_K]$ be the trajectory matrix corresponding to the time series $\tilde{\mathbb{X}}_N = (\tilde{x}_1, \dots, \tilde{x}_N)$. In addition, given a vector $Y = (y_1, \dots, y_M)^T \in \mathbb{R}^M$, let $Y^\nabla \in \mathbb{R}^{M-1}$ denote the vector consisting of the first $M - 1$ components of

Y and similarly $Y \Delta \in \mathbb{R}^{M-1}$ the vector consisting of the last $M - 1$ components of Y . Set $v^2 = \pi_1^2 + \dots + \pi_r^2$ where $\pi_i = |U_i^T Y|$, $i = 1, \dots, r$ denotes the last component of the eigenvector U_i . Assuming $M = (0, 0, \dots, 0, 1)^T \notin \mathcal{r}$ (ie, \mathcal{r} not to be a vertical space), v^2 represents the squared cosine of the angle between M and \mathcal{r} , and $v^2 < 1$ holds. The last component y_M of any vector $Y = (y_1, \dots, y_M)^T \in \mathbb{R}^M$ can then be shown to be a linear combination of its first $M - 1$ components y_1, \dots, y_{M-1} (ie, of Y^∇)

$$y_M = \alpha_1 y_{M-1} + \alpha_2 y_{M-2} + \dots + \alpha_{M-1} y_1, \quad (14)$$

with vector $A = (a_1, \dots, a_{M-1})^T$ being

$$A = \frac{1}{1 - v^2} \sum_{i=1}^r \pi_i U_i^\nabla. \quad (15)$$

Note that the representation in Equation (15) does not depend on the choice of the basis U_1, \dots, U_r in \mathcal{r} .

With the required notation now having been introduced, the time series $\mathbb{Y}_{N+N_F} = (y_1, \dots, y_N, y_{N+1}, \dots, y_{N+N_F})$ can be defined as

$$y_i = \begin{cases} \tilde{y}_i & | i=1, \dots, N \\ \sum_{j=1}^{M-1} \alpha_j y_{i-j} & | i=N+1, \dots, N+N_F \end{cases}, \quad (16)$$

with $\tilde{y}_i |_{i=1, \dots, N}$ being the reconstructed series as in basic SSA and $y_{N+1}, \dots, y_{N+N_F}$ the recurrent SSA forecast of length N_F .

Two examples of the automatic detection and correction of outliers using the method proposed in this study are shown in Figures 4 and 5 and compared with the benchmark approach of replacing the corrupted segment with an immediately preceding

block of R-R data, referred to as block replacement [28]. Note how in both cases, while all outliers have been detected and corrected, there are ectopic-free segments of tachogram which have also been corrected because of false alarms.

Figure 4. Ectopic beats detected on a 5-minute tachogram excerpt of record 19093 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 6 artificially inserted ectopics using lightweight SSA change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples and adaptive control limits sequential ranks cumulative sum with $j_{\max}=6$, $E\{T_n\}=4$, $ARL_0=500$ are corrected by means of the proposed method with $\Delta_1=M$, $\Delta_2=N$ (top) or by replacing the corrupted segment of length Δ_1 with the immediately preceding block of equal length (bottom).

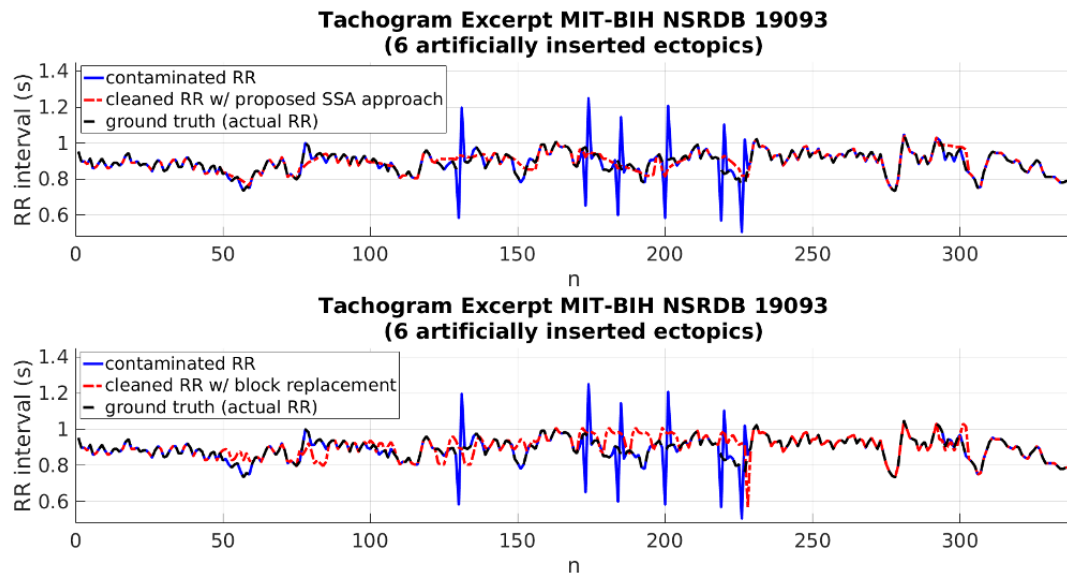
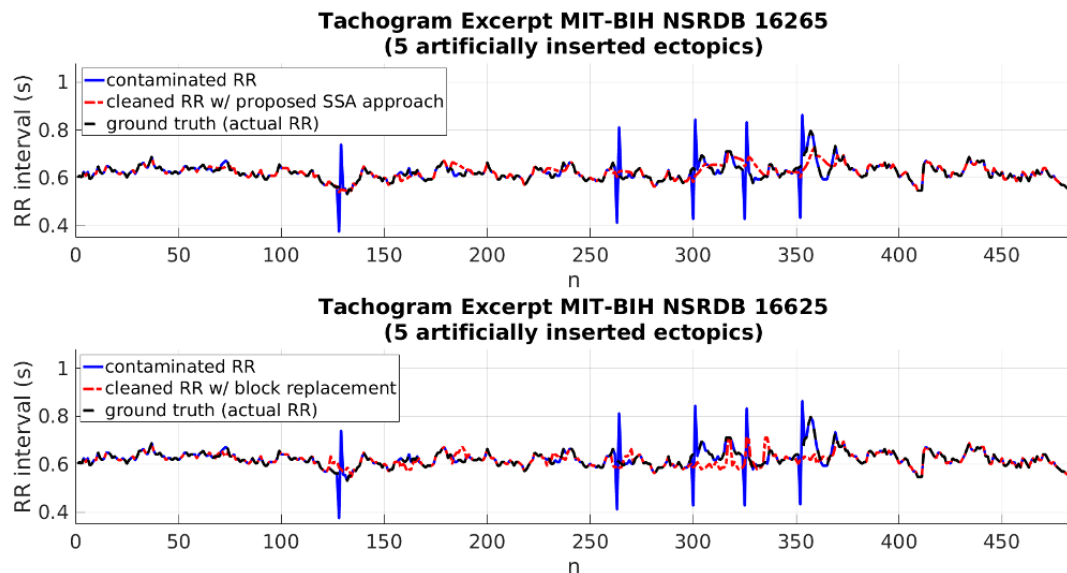


Figure 5. Ectopic beats detected on a 5-minute tachogram excerpt of record 16265 MIT-BIH Normal Sinus Rhythm Database (NSRDB) with 5 artificially inserted ectopics using lightweight SSA change-point detection with $N=20$, $M=10$, $Q=1$, 75% of leading eigentriples and adaptive control limits sequential ranks cumulative sum with $j_{\max}=6$, $E\{T_n\}=4$, $ARL_0=500$ are corrected by means of the proposed method with $\Delta_1=M$, $\Delta_2=N$ (top) or by replacing the corrupted segment of length Δ_1 with the immediately preceding block of equal length (bottom).



Performance Evaluation

The performance and utility of the proposed method are evaluated using records that are publicly available through Physiobank [34]. More specifically, excerpts of all 18 records in the MIT-BIH NSRDB are used. The European Society of Cardiology/North American Society of Pacing and Electrophysiology Task Force guideline [4] recommends to perform the HRV analysis on ECG recordings of at least 5

minutes, and the said length can be considered as the accepted standard for the short-term HRV analysis. Accordingly, for each of the 18 records in MIT-BIH NSRDB excerpts corresponding to the first 5-minute period of NSR are collected. Note that the records in MIT-BIH NSRDB are cardiologist-annotated, that is, they have manually been checked for errors in beat detection and labeling. Thus, one can circumvent the step of QRS complex-detection and use the R-R intervals as provided (and validated by experts) by Physiobank directly. For all 18 records

the first 5 minutes were free of ectopic beats and therefore selected. However, it will be pointed out that 4 records contained

a couple of extreme outliers (most likely measurement errors), which were removed manually.

Textbox 1. Artificial premature ventricular complex obtained by modifying two consecutive tachogram values.

1. $RR(\tau) = RR(\tau) - RR(\tau)/3$;
2. $RR(\tau+1) = RR(\tau+1) + RR(\tau+1)/3$;

Specifically, the 4 affected records were records number 16,272, 18,177, 19,088, and 19,140, for which 2, 1, 2, and 1 data points were removed, respectively. Besides being readily discernible because they deviate by some order of magnitude from surrounding interbeat intervals, these outliers are also marked as artifacts in the respective annotation files.

Next, the effect of a PVC on the R-R interval was simulated by decreasing the respective entry to 2/3 of its actual value and by increasing the following entry to 4/3 of its value (Textbox 1). That is, for a PVC at say $n = \tau$, one sets

which is consistent with the actual signature of a single PVC (see, eg, [53]) in an R-R series characterized by a premature beat (and accordingly a significantly smaller R-R interval) followed by a compensatory pause (longer R-R interval) followed by a return to baseline (NSR).

Extensive Monte Carlo simulations were then used to determine the performance characteristics of the proposed method, wherein a large number of artificial R-R series based on the 18 5-minute MIT-BIH NSRDB records were created by first randomly drawing an integer between 1 and 6 and then randomly selecting the position(s) within the series where the respective PVCs would be placed (with the constraint that PVCs be at least 5 samples apart).

Results

With the proposed method consisting of 2 stages (detection and correction), a separate analysis of each stage appears appropriate.

Detection Step

In reporting results pertaining to the detection performance of the first stage, established metrics commonly used in the literature are relied upon the sensitivity (Se), specificity (Sp), and accuracy (Acc), which are defined as

$$Se = \frac{TP}{(TP + FN)}$$

$$Sp = \frac{TN}{(TN + FP)}$$

and

$$Acc = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

with TP, FP, TN, and FN representing the number of true positives, false positives, true negatives, and false negatives, respectively. Intuitively, sensitivity quantifies the ability to correctly detect (actual) anomalies, specificity quantifies the

proportion of nonabnormal segments that are correctly identified as such, and accuracy combines both of the aforementioned aspects. More formally, Se expresses the empirical statistical power, $1 - Sp$ the type I (false positive), and $1 - Se$ the type II (missed detection/false negative) error rate.

As reported previously [32], for an event occurring at a time instance $n = \tau$, one allows for a certain detection delay Δ_d and considers a signal from the AC-SRC control chart as true positive if it falls in the interval $[\tau, \tau + \Delta_d]$. All results presented in this study were obtained using $\Delta_d = \Delta_1 = M$.

Note that the design and calibration of a detection algorithm involves an inherent trade-off between false alarms and missed detections and that furthermore reducing the false alarm rate usually entails an increase in the detection delay. For this specific use case, high sensitivity and short detection delays are arguably of the highest priority; for one wants a high likelihood of detecting all actual outliers (to reliably “clean” the R-R series, thereby transforming it into the respective N-N series) with a manageable (ie, small) detection delay to be able to narrow down the corrupted segments in the series as much as possible. The latter is particularly relevant because the proposed method substitutes the (allegedly) corrupted segment by an approximation of the outlier-free segment obtained using recurrent SSA forecasting; one aims to keep the forecasting horizon small to obtain acceptable results without having to tweak the SSA parameters extensively. This comes at the expense of higher false alarm rates, as is evidenced by the results presented in Tables 1-4.

As in the considered design, j_{\max} is linked to the average sprint length $E\{T_n\}$, better agility is generally achieved with rather small j_{\max} in a use case, as the one considered here, were AC-SRC monitors for large deviations [33,45].

While AC-SRC achieved satisfactory performance over a wide range of values j_{\max} (results not reported here), upon examination of the results reported in Tables 1-4, which are based on 18,000 R-R series, $j_{\max}=6$ was chosen to proceed to the second stage of the proposed method (ie, the correction of corrupted segments of the R-R series).

Correction Step

As mentioned earlier this design choice entails higher false alarm rates, as discernible in the 2 examples shown in Figures 4 and 5, which both exhibit actually uncorrupted segments that have been “corrected” by the proposed algorithm because of false alarms, thereby introducing some distortions. More importantly though, both examples also clearly show that all artificially inserted outliers because of PVCs were detected and appropriately corrected.

Table 5. Average root mean square errors of artificially contaminated R-R series.

Record	RMSE _{SSA} ^{a,b}	RMSE _{block}	RRMSE ^c	Percent increase or decrease
16,265	0.0225	0.0251	0.8964	10.36
16,272	0.0239	0.0354	0.6751	32.49
16,273	0.0136	0.0221	0.6154	38.46
16,420	0.0246	0.0326	0.7546	24.54
16,483	0.0162	0.0196	0.8265	17.35
16,539	0.0282	0.0342	0.8246	17.54
16,773	0.0299	0.0509	0.5874	41.26
16,786	0.0249	0.0366	0.6803	31.97
16,795	0.0323	0.0450	0.7178	28.22
17,052	0.0213	0.0414	0.5145	48.55
17,453	0.0256	0.0344	0.7442	25.58
18,177	0.0296	0.0357	0.8291	17.09
18,184	0.0298	0.0450	0.6622	33.78
19,088	0.0151	0.0198	0.7626	23.74
19,090	0.0187	0.0245	0.7633	23.67
19,093	0.0350	0.0439	0.7973	20.27
19,140	0.0123	0.0156	0.7885	21.15
19,830	0.0110	0.0125	0.8800	12.00
Average	0.0230	0.0319	0.7210	27.9

^aRMSE: root mean square error.

^bSSA: singular spectrum analysis.

^cRRMSE: relative root mean square error.

Results pertaining to the second stage of the proposed method are reported in Table 5, wherein the root mean square error (RMSE) is used to quantify the discrepancy between the outlier-cleaned series and the actual (uncontaminated) series.

Let X_{RR} be the uncontaminated N-N series (ie, the ground truth) and $X_{\hat{R}\hat{R}_N} = (\hat{x}_1, \dots, \hat{x}_N)$ denote our allegedly cleaned R-R series.

Calculation of the RMSE is straightforward,

$$\text{RMSE} = \sqrt{\sum_{i=1}^N (\hat{x}_i - x_i)^2}. \quad (17)$$

Furthermore, to better compare the RMSE of the proposed method to its competitor (the commonly used block replacement method is used here for benchmarking purposes), consider also the relative RMSE (RRMSE) given by

$$\text{RRMSE} = \frac{\text{RMSE}_{\text{proposed}}}{\text{RMSE}_{\text{benchmark}}}. \quad (18)$$

Note that if the RRMSE, as given in Equation (18) is <1 , it implies that the novel approach presented in this paper outperforms the competing benchmark approach by $100 \cdot (1 - \text{RRMSE})$. It shall be emphasized, however, that the block

replacement method as it is applied here (to correct for outliers in the R-R series) uses the first stage of the proposed method as well, its performance should thus be seen in the context of being applied to pretty tight segments of the series that ought to be substituted, which is all because of the AC-SRC I-SSA-CPD algorithm of the first stage of the framework presented in this study.

That being said, the results reported in Table 5, which are based on 180,000 R-R series, show that the proposed method, which substitutes corrupted tachogram segments by an approximation of the respective outlier-free segment obtained by

means of recurrent SSA forecasting, clearly outperforms the competing block replacement method. In fact, it outperforms its competitor for all 18 records (on which the extensive simulations built upon), on average by almost 30%.

Discussion

Background and Relevance

The HRV has long been known to allow for valuable insights into the intricate control mechanisms of the ANS and has a long track record of exponentially growing research interest and output. A crucial prerequisite for any HRV analysis is the exclusion of all artifacts and abnormalities, that is, a clean N-N series is required. Such a clean N-N series is virtually never

available, mostly because of various interferences, beats not detected by the QRS complex-detector, and various highly prevalent heart rhythm disturbances.

Compared with the overall research effort directed toward the HRV, the issue of automatic R-R series (pre)processing has received rather little attention. In fact, manual editing still represents the gold standard. It should be noted that every automatic approach comes with risks and benefits; the former lie in the inherently nonzero type I and type II error rates, whereas the latter include, but are not limited to, considerable savings in both cost and time. Considering that error rates for manual approaches are nonzero as well and that they are most likely an increasing function of time (whereas in automatic approaches they are usually not), the choice of which one or which particular combination to use should always be made on an individual basis, considering all requirements and circumstances of the particular case. Accordingly, while the proposed method could be construed as a potential alternative to manual editing, in the light of the above considerations, this author would rather recommend it as a complementary method and or a preprocessing tool.

The main contribution of this work is the introduction of a general framework of low computational complexity that allows for an automatic near real-time detection and correction of outliers in R-R series based on the SSA. While related work pertaining to the use of the SSA in ECG and R-R data processing exists [54], to the best of the author's knowledge, this study is the first to propose a general SSA-based framework that handles both detection and correction of (unwanted) anomalies in the R-R series.

Principal Findings

An extensive simulation study comprising 198,000 5-minute R-R series, obtained by randomly inserting varying amounts of simulated ectopic beats in records taken from the MIT-BIH NSRDB was conducted to assess the artifact detection and correction performance of the proposed algorithm.

It was shown that the proposed algorithm reliably detects outliers in the R-R series and achieved an overall sensitivity of 96.6%, specificity of 98.4%, and accuracy of 98.4%. Furthermore, it compares favorably in terms of discrepancies of the cleaned R-R series compared with the actual N-N series, outperforming a block replacement approach on average by almost 30%. It

should also be emphasized that a suboptimal pragmatic approach was pursued, deliberately refraining from an optimization of the various (SSA and AC-SRC) tuning parameters, which would most likely have resulted in further performance improvements at the expense of simplicity and computational complexity.

Other important characteristics of the proposed method include the ability to operate in near real-time, the almost entirely model-free nature of the framework, and the low computational complexity. Moreover, it should be pointed out that the proposed method is not limited to the R-R series but can be applied broadly, as all of its components are deliberately kept as general as possible. This entails the additional benefit of not being limited to the removal of PVCs (although only this has extensively been investigated and tested so far) but rather being able to deal with all kinds of anomalies that may contaminate the observed series of interest.

Limitations and Future Research

This work has several limitations, some of which open (new) avenues for future research.

A major limitation is because of the lack of established and generally recognized and validated procedures other than manual editing, which transforms the performance comparison among competing methods in a nontrivial task. Second, to increase the readability and decrease the length, the paper fails to provide general recommendations and guidelines as to how various tuning parameters should be chosen. Finally, the SSA represents a very active research field, as in the last couple of years, and several promising advances and new developments have been reported, but thus far have not been considered in the research reported in this work. This, however, does not infringe on the relevance of this work, as the objective was to introduce a simple, general framework of low computational complexity and to investigate the use of SSA in both the detection and correction step.

Several of the abovementioned recent developments, however, appear to open promising avenues for future research. In particular, the author intends to expand and improve the work presented here by focusing on the issue of the SSA parameter selection and by considering latest developments, especially pertaining to SSA forecasting. For the latter, the interested reader is referred to recent publications by Hassani and Kalantari [55-58].

Acknowledgments

ML was supported by the "Excellence Initiative" of the German Federal and State Governments and the Graduate School of Excellence Computational Engineering at Technische Universität Darmstadt. The views expressed in this article are solely those of the author in his private capacity and do not necessarily reflect the views of Technische Universität Darmstadt or any other organization.

Conflicts of Interest

None declared.

References

1. Billman GE. Heart rate variability - a historical perspective. *Front Physiol* 2011;2:86 [FREE Full text] [doi: [10.3389/fphys.2011.00086](https://doi.org/10.3389/fphys.2011.00086)] [Medline: [22144961](https://pubmed.ncbi.nlm.nih.gov/22144961/)]

2. Billman GE, Huikuri HV, Sacha J, Trimmel K. An introduction to heart rate variability: methodological considerations and clinical applications. *Front Physiol* 2015;6:55 [FREE Full text] [doi: [10.3389/fphys.2015.00055](https://doi.org/10.3389/fphys.2015.00055)] [Medline: [25762937](https://pubmed.ncbi.nlm.nih.gov/25762937/)]
3. Rajendra AU, Paul JK, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. *Med Biol Eng Comput* 2006 Dec;44(12):1031-1051. [doi: [10.1007/s11517-006-0119-0](https://doi.org/10.1007/s11517-006-0119-0)] [Medline: [17111118](https://pubmed.ncbi.nlm.nih.gov/17111118/)]
4. ESC/NASPE. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation* 1996 Mar 01;93(5):1043-1065 [FREE Full text] [Medline: [8598068](https://pubmed.ncbi.nlm.nih.gov/8598068/)]
5. Huikuri HV, Mäkikallio T, Airaksinen KE, Mitrani R, Castellanos A, Myerburg RJ. Measurement of heart rate variability: a clinical tool or a research toy? *J Am Coll Cardiol* 1999 Dec;34(7):1878-1883 [FREE Full text] [Medline: [10588197](https://pubmed.ncbi.nlm.nih.gov/10588197/)]
6. Sassi R, Cerutti S, Lombardi F, Malik M, Huikuri HV, Peng C, et al. Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society. *Europace* 2015 Sep;17(9):1341-1353. [doi: [10.1093/europace/euv015](https://doi.org/10.1093/europace/euv015)] [Medline: [26177817](https://pubmed.ncbi.nlm.nih.gov/26177817/)]
7. Laske TG, Shrivastav M, Iaizzo PA. The cardiac conduction system. In: Iaizzo PA, editor. *Handbook of cardiac anatomy, physiology, and devices*, 3rd ed. Cham, Switzerland: Springer International Publishing; 2015:215-234.
8. Iaizzo PA, Fitzgerald K. Autonomic nervous system. In: Iaizzo PA, editor. *Handbook of cardiac anatomy, physiology, and devices*, 3rd ed. Cham, Switzerland: Springer International Publishing; 2015:235-250.
9. Karemaker JM. An introduction into autonomic nervous function. *Physiol Meas* 2017 May;38(5):R89-R118. [doi: [10.1088/1361-6579/aa6782](https://doi.org/10.1088/1361-6579/aa6782)] [Medline: [28304283](https://pubmed.ncbi.nlm.nih.gov/28304283/)]
10. Berntson GG, Bigger JT, Eckberg DL, Grossman P, Kaufmann PG, Malik M, et al. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 1997 Nov;34(6):623-648. [Medline: [9401419](https://pubmed.ncbi.nlm.nih.gov/9401419/)]
11. Peltola MA. Role of editing of R-R intervals in the analysis of heart rate variability. *Front Physiol* 2012;3:148 [FREE Full text] [doi: [10.3389/fphys.2012.00148](https://doi.org/10.3389/fphys.2012.00148)] [Medline: [22654764](https://pubmed.ncbi.nlm.nih.gov/22654764/)]
12. Nabil D, Bereksi Reguig F. Ectopic beats detection and correction methods: A review. *Biomedical Signal Processing and Control* 2015 Apr;18:228-244. [doi: [10.1016/j.bspc.2015.01.008](https://doi.org/10.1016/j.bspc.2015.01.008)]
13. Salo MA, Huikuri HV, Seppänen T. Ectopic beats in heart rate variability analysis: effects of editing on time and frequency domain measures. *Ann Noninvasive Electrocardiol* 2001 Jan;6(1):5-17. [Medline: [11174857](https://pubmed.ncbi.nlm.nih.gov/11174857/)]
14. Kim KK, Lim YG, Kim JS, Park KS. Effect of missing RR-interval data on heart rate variability analysis in the time domain. *Physiol Meas* 2007 Dec;28(12):1485-1494. [doi: [10.1088/0967-3334/28/12/003](https://doi.org/10.1088/0967-3334/28/12/003)] [Medline: [18057513](https://pubmed.ncbi.nlm.nih.gov/18057513/)]
15. Kim KK, Kim JS, Lim YG, Park KS. The effect of missing RR-interval data on heart rate variability analysis in the frequency domain. *Physiol Meas* 2009 Oct;30(10):1039-1050. [doi: [10.1088/0967-3334/30/10/005](https://doi.org/10.1088/0967-3334/30/10/005)] [Medline: [19713596](https://pubmed.ncbi.nlm.nih.gov/19713596/)]
16. Kim KK, Baek HJ, Lim YG, Park KS. Effect of missing RR-interval data on nonlinear heart rate variability analysis. *Comput Methods Programs Biomed* 2012 Jun;106(3):210-218. [doi: [10.1016/j.cmpb.2010.11.011](https://doi.org/10.1016/j.cmpb.2010.11.011)] [Medline: [21194782](https://pubmed.ncbi.nlm.nih.gov/21194782/)]
17. Lippman N, Stein KM, Lerman BB. Comparison of methods for removal of ectopy in measurement of heart rate variability. *Am J Physiol* 1994 Jul;267(1 Pt 2):H411-H418. [doi: [10.1152/ajpheart.1994.267.1.H411](https://doi.org/10.1152/ajpheart.1994.267.1.H411)] [Medline: [7519408](https://pubmed.ncbi.nlm.nih.gov/7519408/)]
18. Stapelberg NJC, Neumann DL, Shum DHK, McConnell H, Hamilton-Craig I. The sensitivity of 38 heart rate variability measures to the addition of artifact in human and artificial 24-hr cardiac recordings. *Ann Noninvasive Electrocardiol* 2018 Jan;23(1). [doi: [10.1111/anec.12483](https://doi.org/10.1111/anec.12483)] [Medline: [28670841](https://pubmed.ncbi.nlm.nih.gov/28670841/)]
19. Hingorani P, Karnad DR, Rohekar P, Kerkar V, Lokhandwala YY, Kothari S. Arrhythmias Seen in Baseline 24-Hour Holter ECG Recordings in Healthy Normal Volunteers During Phase 1 Clinical Trials. *J Clin Pharmacol* 2016 Dec;56(7):885-893. [doi: [10.1002/jcph.679](https://doi.org/10.1002/jcph.679)] [Medline: [26626443](https://pubmed.ncbi.nlm.nih.gov/26626443/)]
20. Manolio TA, Furberg CD, Rautaharju PM, Siscovick D, Newman AB, Borhani NO, et al. Cardiac arrhythmias on 24-h ambulatory electrocardiography in older women and men: the Cardiovascular Health Study. *J Am Coll Cardiol* 1994 Mar 15;23(4):916-925. [Medline: [8106697](https://pubmed.ncbi.nlm.nih.gov/8106697/)]
21. Simpson RJ, Cascio WE, Schreiner PJ, Crow RS, Rautaharju PM, Heiss G. Prevalence of premature ventricular contractions in a population of African American and white men and women: the Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J* 2002 Mar;143(3):535-540. [Medline: [11868062](https://pubmed.ncbi.nlm.nih.gov/11868062/)]
22. Messineo FC. Ventricular ectopic activity: prevalence and risk. *Am J Cardiol* 1989 Dec 05;64(20):53J-56J. [Medline: [2480710](https://pubmed.ncbi.nlm.nih.gov/2480710/)]
23. Kostis JB, McCrone K, Moreyra AE, Gotzoyannis S, Aglitz NM, Natarajan N, et al. Premature ventricular complexes in the absence of identifiable heart disease. *Circulation* 1981 Jun;63(6):1351-1356. [Medline: [7226480](https://pubmed.ncbi.nlm.nih.gov/7226480/)]
24. Laborde S, Mosley E, Thayer JF. Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research - Recommendations for Experiment Planning, Data Analysis, and Data Reporting. *Front Psychol* 2017;8:213 [FREE Full text] [doi: [10.3389/fpsyg.2017.00213](https://doi.org/10.3389/fpsyg.2017.00213)] [Medline: [28265249](https://pubmed.ncbi.nlm.nih.gov/28265249/)]
25. Kubios HRV. URL: <http://www.kubios.com/> [accessed 2018-05-13] [WebCite Cache ID 6zO6WmKpu]
26. Tarvainen MP, Niskanen J, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV-heart rate variability analysis software. *Comput Methods Programs Biomed* 2014;113(1):210-220. [doi: [10.1016/j.cmpb.2013.07.024](https://doi.org/10.1016/j.cmpb.2013.07.024)] [Medline: [24054542](https://pubmed.ncbi.nlm.nih.gov/24054542/)]
27. nevrokard HRV. URL: <http://www.nevrokard.eu/hrv.htm> [accessed 2018-05-13] [WebCite Cache ID 6zO68ONoy]

28. Pichot V, Roche F, Celle S, Barthélémy J, Chouchou F. HRVanalysis: A Free Software for Analyzing Cardiac Autonomic Activity. *Front Physiol* 2016 Nov;7:557 [FREE Full text] [doi: [10.3389/fphys.2016.00557](https://doi.org/10.3389/fphys.2016.00557)] [Medline: [27920726](https://pubmed.ncbi.nlm.nih.gov/27920726/)]
29. Perakakis P, Joffily M, Taylor M, Guerra P, Vila J. KARDIA: a Matlab software for the analysis of cardiac interbeat intervals. *Comput Methods Programs Biomed* 2010 Apr;98(1):83-89. [doi: [10.1016/j.cmpb.2009.10.002](https://doi.org/10.1016/j.cmpb.2009.10.002)] [Medline: [19896748](https://pubmed.ncbi.nlm.nih.gov/19896748/)]
30. Kaufmann T, Sütterlin S, Schulz SM, Vögele C. ARTiiFACT: a tool for heart rate artifact processing and heart rate variability analysis. *Behav Res Methods* 2011 Dec;43(4):1161-1170. [doi: [10.3758/s13428-011-0107-7](https://doi.org/10.3758/s13428-011-0107-7)] [Medline: [21573720](https://pubmed.ncbi.nlm.nih.gov/21573720/)]
31. Clifford G, McSharry P, Tarassenko L. Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold. *Computers in Cardiology* 2002;29:129-132. [doi: [10.1109/CIC.2002.1166724](https://doi.org/10.1109/CIC.2002.1166724)]
32. Lang M. A Low-Complexity Model-Free Approach for Real-Time Cardiac Anomaly Detection Based on Singular Spectrum Analysis and Nonparametric Control Charts. *Technologies* 2018 Feb 15;6(1):26. [doi: [10.3390/technologies6010026](https://doi.org/10.3390/technologies6010026)]
33. Lang M, Zoubir AM. A nonparametric cumulative sum scheme based on sequential ranks and adaptive control limits. In: *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO)*. 2015 Presented at: 23rd European Signal Processing Conference (EUSIPCO); 2015 Aug 31; Nice, France p. 1984-1988. [doi: [10.1109/EUSIPCO.2015.7362731](https://doi.org/10.1109/EUSIPCO.2015.7362731)]
34. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000 Jun 13;101(23):E215-E220 [FREE Full text] [Medline: [10851218](https://pubmed.ncbi.nlm.nih.gov/10851218/)]
35. Golyandina N, Nekrutkin V, Zhigljavsky A. *Analysis of Time Series Structure: SSA and Related Techniques*. Boca Raton, FL, USA: Chapman and Hall/CRC; 2001.
36. Golyandina N, Zhigljavsky A. *Singular Spectrum Analysis for Time Series (SpringerBriefs in Statistics)*. Berlin/Heidelberg, Germany: Springer; 2013.
37. Elsner JB, Tsonis AA. *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. New York, NY, USA: Springer; 1996:a.
38. Sanei S, Hassani H. *Singular Spectrum Analysis of Biomedical Signals*. Boca Raton, FL, USA: CRC Press; 2015.
39. Moskvina V. *Application of the singular spectrum analysis for change-point detection in time series*. Ph.D. Thesis, Cardiff University: Cardiff, UK 2001.
40. Moskvina V, Zhigljavsky A. An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection. *Communications in Statistics - Simulation and Computation* 2003 Jan 06;32(2):319-352. [doi: [10.1081/SAC-120017494](https://doi.org/10.1081/SAC-120017494)]
41. Hassani H, Mahmoudvand R, Zokaei M. Separability and window length in singular spectrum analysis. *Comptes Rendus Mathématique* 2011 Sep;349(17-18):987-990. [doi: [10.1016/j.crma.2011.07.012](https://doi.org/10.1016/j.crma.2011.07.012)]
42. Atikur Rahman Khan M, Poskitt DS. A Note on Window Length Selection in Singular Spectrum Analysis. *Aust. N. Z. J. Stat* 2013 Jun 19;55(2):87-108. [doi: [10.1111/anzs.12027](https://doi.org/10.1111/anzs.12027)]
43. Khan AR, Poskitt DS. Signal Identification in Singular Spectrum Analysis. *Aust. N. Z. J. Stat* 2016 Mar 23;58(1):71-98. [doi: [10.1111/anzs.12143](https://doi.org/10.1111/anzs.12143)]
44. McDonald D. A cusum procedure based on sequential ranks. *Naval Research Logistics* 1990:37-646. [doi: [10.1002/1520-6750\(199010\)37:5<627::AID-NAV3220370504>3.0.CO;2-F](https://doi.org/10.1002/1520-6750(199010)37:5<627::AID-NAV3220370504>3.0.CO;2-F)]
45. Chatterjee S, Qiu P. Distribution-free cumulative sum control charts using bootstrap-based control limits. *Ann. Appl. Stat* 2009 Mar;3(1):349-369. [doi: [10.1214/08-AOAS197](https://doi.org/10.1214/08-AOAS197)]
46. Ghodsi M, Hassani H, Rahmani D, Silva ES. Vector and recurrent singular spectrum analysis: which is better at forecasting? *Journal of Applied Statistics* 2017 Nov 15;45(10):1872-1899. [doi: [10.1080/02664763.2017.1401050](https://doi.org/10.1080/02664763.2017.1401050)]
47. Hassani H, Ghodsi Z, Gupta R, Segnon M. Forecasting Home Sales in the Four Census Regions and the Aggregate US Economy Using Singular Spectrum Analysis. *Comput Econ* 2015 Dec 17;49(1):83-97. [doi: [10.1007/s10614-015-9548-x](https://doi.org/10.1007/s10614-015-9548-x)]
48. Silva ES, Hassani H. On the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recession. *International Economics* 2015 May;141:34-49. [doi: [10.1016/j.inteco.2014.11.003](https://doi.org/10.1016/j.inteco.2014.11.003)]
49. Hassani H, Heravi S, Brown G, Ayoubkhani D. Forecasting before, during, and after recession with singular spectrum analysis. *Journal of Applied Statistics* 2013 Oct;40(10):2290-2302. [doi: [10.1080/02664763.2013.810193](https://doi.org/10.1080/02664763.2013.810193)]
50. Hassani H, Heravi S, Zhigljavsky A. Forecasting UK Industrial Production with Multivariate Singular Spectrum Analysis. *J. Forecast* 2012 Apr 10;32(5):395-408. [doi: [10.1002/for.2244](https://doi.org/10.1002/for.2244)]
51. Miranian A, Abdollahzade M, Hassani H. Day-ahead electricity price analysis and forecasting by singular spectrum analysis. *IET Generation Transmission & Distribution* 2013;7(4):346.
52. Hassani H, Thomakos D. A review on singular spectrum analysis for economic and financial time series. *Statistics and Its Interface* 2010;3(3):377-397. [doi: [10.4310/SII.2010.v3.n3.a11](https://doi.org/10.4310/SII.2010.v3.n3.a11)]
53. Clifford GD, Tarassenko L. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. *IEEE Trans Biomed Eng* 2005 Apr;52(4):630-638. [doi: [10.1109/TBME.2005.844028](https://doi.org/10.1109/TBME.2005.844028)] [Medline: [15825865](https://pubmed.ncbi.nlm.nih.gov/15825865/)]
54. Thuraisingham RA. Use of SSA and MCSSA in the Analysis of Cardiac RR Time Series. *Journal of Computational Medicine* 2013;2013:1-7. [doi: [10.1155/2013/231459](https://doi.org/10.1155/2013/231459)]
55. Kalantari M, Yarmohammadi M, Hassani H. Singular Spectrum Analysis Based on L1-Norm. *Fluct. Noise Lett* 2016 Mar;15(01):1650009. [doi: [10.1142/S0219477516500097](https://doi.org/10.1142/S0219477516500097)]

56. Kalantari M, Yarmohammadi M, Hassani H, Silva ES. Time Series Imputation via L1 Norm-Based Singular Spectrum Analysis. *Fluct. Noise Lett* 2018 Jun;17(02):1850017. [doi: [10.1142/S0219477518500177](https://doi.org/10.1142/S0219477518500177)]
57. Hassani H, Kalantari M, Yarmohammadi M. An improved SSA forecasting result based on a filtered recurrent forecasting algorithm. *Comptes Rendus Mathematique* 2017 Sep;355(9):1026-1036. [doi: [10.1016/j.crma.2017.09.004](https://doi.org/10.1016/j.crma.2017.09.004)] [Medline: [25904163](https://pubmed.ncbi.nlm.nih.gov/25904163/)]
58. Hassani H, Kalantari M. A novel signal extraction approach for filtering and forecasting noisy exponential series. *Comptes Rendus Mathematique* 2018 May;356(5):563-570. [doi: [10.1016/j.crma.2018.03.006](https://doi.org/10.1016/j.crma.2018.03.006)] [Medline: [25904163](https://pubmed.ncbi.nlm.nih.gov/25904163/)]

Abbreviations

AC-SRC: adaptive control limits sequential ranks cumulative sum

ANS: autonomic nervous system

ARL: average run length

CUSUM: cumulative sum

ECG: electrocardiography

HR: heart rate

HRV: heart rate variability

I-SSA-CPD: lightweight SSA change-point detection

NSR: normal sinus rhythm

NSRB: Normal Sinus Rhythm Database

PVC: premature ventricular complex

RMSE: root mean square error

RRMSE: relative root mean square error

SA: sinoatrial

SRC: sequential ranks cumulative sum

SSA: singular spectrum analysis

SVD: singular value decomposition

Edited by G Eysenbach; submitted 09.04.18; peer-reviewed by H Hassani, IN Gomez, L Becker; comments to author 02.05.18; revised version received 15.05.18; accepted 09.06.18; published 30.01.19

Please cite as:

Lang M

Automatic Near Real-Time Outlier Detection and Correction in Cardiac Interbeat Interval Series for Heart Rate Variability Analysis: Singular Spectrum Analysis-Based Approach

JMIR Biomed Eng 2019;4(1):e10740

URL: <https://biomedeng.jmir.org/2019/1/e10740/>

doi: [10.2196/10740](https://doi.org/10.2196/10740)

PMID:

©Michael Lang. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 30.01.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <http://biomedeng.jmir.org/>, as well as this copyright and license information must be included.