



Research article

Refined matrix completion for spectrum estimation of heart rate variability

Lei Lu^{1,2,*}, Tingting Zhu², Ying Tan³, Jiandong Zhou^{4,5,6}, Jenny Yang², Lei Clifton⁷, Yuan-Ting Zhang⁸ and David A. Clifton^{2,9}

¹ School of Life Course & Population Sciences, King's College London, London WC2R 2LS, UK

² Department of Engineering Science, University of Oxford, Oxford OX1 2JD, UK

³ Department of Mechanical Engineering, The University of Melbourne, Parkville 3010, Australia

⁴ Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

⁵ School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

⁶ Department of Pharmacology and Pharmacy, The University of Hong Kong, Hong Kong, China

⁷ Nuffield Department of Clinical Medicine, Experimental Medicine Division, University of Oxford, Oxford, UK

⁸ Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, China

⁹ Oxford Suzhou Centre for Advanced Research, Suzhou, China

* **Correspondence:** Email: lei.lu@eng.ox.ac.uk; Tel: +442078365454.

Abstract: Heart rate variability (HRV) is an important metric in cardiovascular health monitoring. Spectral analysis of HRV provides essential insights into the functioning of the cardiac autonomic nervous system. However, data artefacts could degrade signal quality, potentially leading to unreliable assessments of cardiac activities. In this study, we introduced a novel approach for estimating uncertainties in HRV spectrum based on matrix completion. The proposed method utilises the low-rank characteristic of HRV spectrum matrix to efficiently estimate data uncertainties. In addition, we developed a refined matrix completion technique to enhance the estimation accuracy and computational cost. Benchmarking on five public datasets, our model shows effectiveness and reliability in estimating uncertainties in HRV spectrum, and has superior performance against five deep learning models. The results underscore the potential of our developed matrix completion-based statistical machine learning model in providing reliable HRV spectrum uncertainty estimation.

Keywords: heart rate variability; spectrum estimation; matrix completion; uncertainty; HRV modelling

1. Introduction

Heart rate variability (HRV) is the measure of fluctuations between successive heartbeats, reflecting the interplay between the sympathetic and parasympathetic branches of the autonomic nervous system (ANS) [1]. These fluctuations are critical markers of the physiological factors influencing heart rhythm, serving as indicators of an individual's health status [2]. Studies have shown that higher HRV levels are indicative of robust health, suggesting a well-balanced autonomic regulation. In contrast, reduced HRV is associated with an imbalance in autonomic function, leaning towards either diminished parasympathetic activity or heightened sympathetic drive, which correlates with an increased risk of adverse health outcomes [3, 4]. The relevance of HRV extends to its application in diagnosing and monitoring cardiovascular conditions, which are leading contributors to global mortality [5]. This underscores the importance of HRV as an important parameter in the evaluation of both physiological well-being and the presence of pathological conditions, particularly those related to cardiovascular health.

HRV data analysis can be implemented through various signal processing techniques. In the time domain, HRV is assessed by calculating statistical measures such as the average of normal-to-normal (NN) intervals, the root mean square of successive differences between NN intervals (RMSSD), and the standard deviation of NN intervals (SDNN) [6]. These metrics provide insights into the heart rhythm and its variability over time. On the other hand, frequency domain analysis of HRV offers a perspective on the distribution of power across different frequency bands within the HRV data, serving as a more detailed reflection of autonomic nervous system activity. Spectral analysis has proven effective in distinguishing between sympathetic and parasympathetic nervous system influences on heart rate variability [7]. Specifically, the low-frequency (LF) component of the HRV spectrum is influenced by both sympathetic and parasympathetic activities, whereas the high-frequency (HF) component is predominantly a result of parasympathetic activity. Consequently, the LF/HF ratio is commonly utilised as a measure of the balance between sympathetic and parasympathetic influences, offering valuable insights into autonomic nervous system dynamics [8].

The most commonly used method to derive HRV data is via electrocardiogram (ECG) signals. With appropriate QRS detectors, R-peaks in the ECG morphology can be identified, and then the HRV data can be obtained by computing the RR intervals [9]. Other than using ECGs, some studies suggest using photoplethysmography (PPG) signals to derive the HRV data, where PPG is a non-invasive technique that utilises optical principles to obtain pulse waves from the microcirculation in peripheral tissue [10]. In particular, the utilisation of PPG signals for detecting pulse-wave related HRV, known as pulse rate variability (PRV), has attracted considerable attention in recent years [11]. This is primarily attributed to the convenience and affordability of acquiring PPG signals through wearable devices.

While electrocardiogram (ECG) or photoplethysmography (PPG) signals are commonly employed for HRV analysis, ensuring the reliability of HRV data extracted from these signals presents significant challenges. These difficulties are largely due to uncertainties inherent in data collection and computational methodologies, which may compromise data integrity. In particular, motion artefacts represent a substantial challenge, often overlapping in frequency with legitimate HRV signals (0.5 to 5 Hz) as they typically manifest within a 0.01 to 10 Hz range [12]. This overlap complicates the task of artefact removal without inadvertently altering the PPG signal. Moreover, the processes involved in HRV data analysis, such as the identification of R-peaks, interpolation of RR intervals, and data resampling, introduce additional complexity. These steps are crucial for accurate HRV measurement but

can potentially distort the HRV data, highlighting the intricate balance required in processing and analysing HRV signals.

A variety of computational methods have been devised to mitigate uncertainties or inaccuracies in HRV data analysis, including adaptive filtering, blind source separation, and advanced deep learning techniques [13–15]. Despite the sophistication of these approaches, they often overlook the unique properties of the HRV spectrum's frequency bands, which limits their effectiveness in accurately estimating uncertainties. For instance, the low-frequency (LF) component of the HRV spectrum is influenced by both sympathetic and parasympathetic nervous system activities, whereas the high-frequency (HF) component is primarily regulated by the parasympathetic nervous system (PNS) [8]. Research has shown that measurement noise impacts these frequency bands; for example, the LF component exhibits substantial sensitivity to increasing noise levels, while the impact is not obvious in the HF component of the HRV spectrum [16]. This differential impact highlights the necessity of accounting for the distinct characteristics of HRV frequency bands in the development and application of computational techniques for HRV data analysis.

In this study, we introduce a novel approach for estimating uncertainties in the HRV spectrum by leveraging a refined low-rank matrix completion (MC) technique, which incorporates the distinct characteristics of HRV frequency components to improve the precision of uncertainty estimation. The concept of low-rank MC emerges as a powerful strategy for reconstructing missing or imprecise entries in observed data, grounded in the principle that many real-world datasets exhibit a low-rank structure due to their inherent low-dimensional characteristics or underlying patterns [17, 18]. The application of low-rank properties to data is widely recognised in various domains, indicating the existence of underlying trends or dimensions within extensive datasets [19]. The MC method has demonstrated its efficacy in various fields, including enhancing image quality in compressive sampling photoacoustic microscopy [20], optimising signal channel selection in electroencephalogram analysis [21], and facilitating the identification of protein-protein interactions [22]. While there have been initial explorations of MC techniques in analysing HRV data [23, 24], its potential in healthcare remains largely untapped, with a significant gap in detailed model development and comprehensive validation of its performance in medical applications.

By leveraging the low-rank characteristic of HRV data, we demonstrated the effective inference of uncertainties within the HRV spectrum using partial entries from the data matrix. This approach is further enhanced by a detailed analysis of HRV data characteristics, particularly focusing on the low-frequency (LF) and high-frequency (HF) components of the spectrum. Drawing on these insights, we developed a refined matrix completion (RMC) framework specifically designed to improve the estimation of uncertainties. This framework prioritises critical data entries within the HRV matrix, thereby enhancing the method's efficiency and robustness in estimating uncertainties. To assess the efficacy of our proposed RMC method in HRV spectrum estimation, we conducted comprehensive experimental studies across five benchmark datasets. Our method was evaluated against both the traditional MC approach and five regression-based machine learning models. The findings from these experiments and comparative analyses underscore the superior performance and robustness of our RMC method in HRV spectrum estimation. The key contributions of this study include:

- The introduction of an innovative method for estimating uncertainties in the HRV spectrum, utilising the matrix completion technique to exploit the low-rank nature of the HRV data.

- The development of an advanced matrix completion framework that enhances estimation accuracy and computational efficiency by focusing on the modelled spectrum of HRV data.
- The implementation of extensive experimental validation on five renowned benchmark datasets to evaluate the effectiveness of our model in HRV spectrum estimation.
- A series of comparative analyses with various neural network models and the traditional MC method, illustrating the significant advantages of our approach in the context of uncertainty estimation.

The remainder of this paper is organised as follows. Section 2 formulates the problem of HRV spectrum estimation. Section 3 develops the MC method and refinement. Section 4 performs the analysis of HRV data. The results of spectrum estimation and comparison study are presented in Section 5 and, finally, conclusions are drawn in Section 6.

2. Problem formulation

2.1. Notations

In this paper, we denote the set of real numbers by \mathbb{R} and the set of natural numbers by \mathcal{N} . A matrix is represented by a bold uppercase letter \mathbf{A} , which consists of elements $a_{i,j}$ for $i = 1, \dots, n$ and $j = 1, \dots, p$, and belongs to the space $\mathbb{R}^{n \times p}$, indicating it has n rows and p columns. The Frobenius norm of matrix \mathbf{A} is expressed as $\|\mathbf{A}\|_F = \left(\sum_{i=1}^n \sum_{j=1}^p |a_{ij}|^2\right)^{1/2}$, encapsulating the square root of the sum of the absolute squares of its elements. The nuclear norm of matrix \mathbf{A} is defined as $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$, where $\sigma_i(\mathbf{A})$ represents the i^{th} singular value of \mathbf{A} . The 2-norm of \mathbf{A} is denoted as $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$, with σ_{\max} being its largest singular value.

2.2. Problem formulation

Assuming that $z(t) \in \mathbb{R}$, for $t \in [0, T_s]$, represents the preprocessed HRV data over a time span of T_s , its frequency components can be delineated as follows:

$$P(f, \tilde{f}) = \mathcal{F}(z(t)), \quad (2.1)$$

where $\tilde{f} = [f_l, f_r]$ specifies the frequency range of interest within the HRV data, with f_l and f_r denoting the left and right boundaries, respectively. Here, $\mathcal{F}(\cdot)$ symbolises the spectrum operation tailored to the selected frequency range \tilde{f} , and $P(f, \tilde{f})$ represents the power spectral density (PSD) function of the HRV data, where various frequency-related HRV metrics can be extracted. More precisely,

$$h_k(f, \tilde{f}_k) = \phi_k(\tilde{f}_k, P(f, \tilde{f})), \quad (2.2)$$

where $k = 1, 2, \dots, N_k$ indicates different types of HRV variables, such as the power of LF or HF, or the ratio between the LF and HF. The nonlinear mapping $\phi_k(\cdot, \cdot)$ computes frequency variable h_k from the spectrum $P(\cdot, \cdot)$ for the k^{th} type of HRV variables in the \tilde{f}_k frequency range.

It is important to note that the accuracy of HRV measurements is influenced by various uncertainties, including noise and missing values. Additional factors, such as respiration rate, circadian rhythms, and

physiological states, can also affect the data quality, therefore impacting the frequency analysis of HRV data. Consequently, these uncertainties pose challenges in precisely estimating the HRV spectrum.

Given data collected from N_s subjects, we construct a matrix $\mathbf{H} = \{h_{i,k}\}_{i=1,\dots,N_s,k=1,\dots,N_k} \in \mathbb{R}^{N_s \times N_k}$, derived from the measurements, where each element is defined by

$$h_{i,k}(f, \tilde{f}_k) = \phi_k(\tilde{f}_k, P_i(f, \tilde{f})), \quad (2.3)$$

where P_i is the PSD of the i^{th} subject.

The PSD often reveals predominant frequency components, with many of the non-dominant frequencies having small amplitudes that are considered negligible, resulting in a sparse matrix. In addition, the HRV spectrum matrix is derived from the time series of intervals between heartbeats. As the ECG waveform has similar repeating patterns [25], the different spectral components derived from ECG data would not be completely independent. This leads to the rank of the matrix being lower than the dimensions of the matrix, suggesting the low-rank characteristic of the HRV spectrum matrix. This observation underpins our motivation to apply low-rank MC for approximating the matrix and estimating uncertainties. This motivates us to introduce the RMC approach aiming at more effectively estimating the uncertainties within the spectrum. This step represents an advancement in our methodology, leveraging the foundational concept of low-rank MC while incorporating adjustments to enhance the precision and efficiency of uncertainty estimation in HRV spectrum analysis.

3. Methodology

This section delves into the application of low-rank MC technique for handling sparse matrices. The core principle revolves around identifying an effective approximation for the matrix that aligns with certain predefined cost functions. One common approach involves seeking an optimal approximation that minimises the matrix rank. The section starts with a revisit of low-rank MC, followed by the matrix approximation with the interested zone, and then the introduction of the RMC approach.

3.1. Low-rank matrix completion

Assuming matrix $\mathbf{H} \in \mathbb{R}^{N_s \times N_k}$ is sparse, i.e., only a limited number of entries in the matrix are non-zero, we denote the set Ω as the collection of non-zero entries of the matrix \mathbf{H} , which is represented as

$$\Omega = \{(i, k) \mid h_{i,k} \neq 0, i = 1, \dots, N_s, k = 1, \dots, N_k\}. \quad (3.1)$$

The spectra \mathbf{H} , derived from segmenting HRV data sequences, often contain redundant information due to the recursive slicing approach used to generate these segments. This redundancy suggests that not all information within the matrix is equally valuable for analysis. To address this, we employ a low-rank matrix completion technique, which is particularly adept at preserving essential information while minimising redundancy [17, 26]. It can be formulated by approximating matrix \mathbf{H} in terms of the smallest rank, i.e.,

$$\begin{aligned} \min \quad & \text{rank}(\mathbf{X}) \\ \text{s.t.} \quad & x_{i,k} = h_{i,k}, \quad \text{for } (i, k) \in \Omega, \end{aligned} \quad (3.2)$$

where matrix $\mathbf{X} = \{x_{i,k}\}_{i=1,\dots,N_s,k=1,\dots,N_k} \in \mathbb{R}^{N_s \times N_k}$ is an optimal approximation matrix that can keep the majority information of \mathbf{H} .

The optimisation of rank-minimization problem in Eq (3.2) is generally computationally intractable. Instead of solving the original problem of the rank norm, a convex relaxation based on the nuclear norm is given by [17],

$$\begin{aligned} \min \quad & \|\mathbf{X}\|_* \\ \text{s.t.} \quad & x_{i,k} = h_{i,k}, \quad \text{for } (i,k) \in \Omega, \end{aligned} \quad (3.3)$$

where, $\|\cdot\|_*$ indicates the nuclear norm of matrix \mathbf{X} , which is defined as the summation of its singular values. In particular, it is demonstrated that a smaller nuclear norm indicates the lower rank of the matrix [17]. However, it is generally too harsh to force the estimation of all observed entries in the matrix, so alternatively, a more practical approximation of Eq (3.3) can be formulated by relaxing the constraint [27, 28].

$$\begin{aligned} \min \quad & \|\mathbf{X}\|_* \\ \text{s.t.} \quad & \|\mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{X}) - \mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{H})\|_F \leq \delta, \end{aligned} \quad (3.4)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, δ is a small value to constrain the estimation, and $\mathcal{A}_{\Omega, \mathbf{H}}(\cdot)$ is a projection operator that is defined as

$$\mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{X}) = \begin{cases} x_{i,k} = h_{i,k} & (i,k) \in \Omega \\ 0 & \text{otherwise.} \end{cases} \quad (3.5)$$

It is noted that Eq (3.4) can be formulated as a penalised least-squares convex program [29], which can be solved by the singular value thresholding (SVT) technique as described in [30].

3.2. Matrix approximation with interested zone

The method previously outlined operates under the assumption that the observed entries, including those missing, are random variables that follow a uniform distribution [17]. This assumption does not fully account for the effects of measurement uncertainties on different frequency components of the data, which often deviate from a uniform distribution pattern [16]. Therefore, the traditional low-rank MC techniques mentioned above could not be applied directly. Alternatively, the interest-zone matrix-approximation (IZMA) method [26] is proposed to solve the low-rank MC problem when some columns/sub-columns or rows/sub-rows are missing. More specifically, the IZMA tries to solve the following optimisation problem,

$$\begin{aligned} \min \quad & \|\mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{X}) - \mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{H})\|_F \\ \text{s.t.} \quad & p(\mathbf{X}) \leq 0, \end{aligned} \quad (3.6)$$

where the nonlinear mapping $p(\cdot) : \mathbb{R}^{N_s \times N_k} \rightarrow \mathbb{R}^n$ is a smooth function for some $n \in \mathcal{N}$. For example, $p(\mathbf{X})$ can be some constraints with respect to the estimated matrix, such as $p(\mathbf{X}) = \|\mathbf{X}\|_2 - \lambda, \lambda > 0$. Another example of this nonlinear mapping is $p(\mathbf{X}) = \|\mathbf{X}\|_* - \lambda$ for some $\lambda > 0$, as described in [26]. The optimisation problem in Eq (3.6) can be solved by the following steps.

It first approximates the following optimisation for the constrained full matrix,

$$\begin{aligned} \min \quad & \|\mathbf{X} - \mathbf{H}\|_F \\ \text{s.t.} \quad & p(\mathbf{X}) \leq 0. \end{aligned} \quad (3.7)$$

The solution to the optimisation in Eq (3.7) with the spectral norm constraint can be estimated by

$$\begin{cases} \tilde{\mathbf{X}} = \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^*, \\ \tilde{\mathbf{\Sigma}} = \text{diag}(\tilde{\sigma}_i), \end{cases} \quad (3.8)$$

where matrices \mathbf{U} and \mathbf{V} are calculated by $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, with $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$. Next, the diagonal matrix is obtained as $\tilde{\sigma}_i = \min(\sigma_i, \lambda)$, $i = 1, 2, \dots, k$, and the approximation $\tilde{\mathbf{X}}$ can be obtained by the data reconstruction.

Then, the estimation of matrix \mathbf{H} with entries in the interested zone can be updated as follows,

$$\mathcal{H}(\mathbf{X}) := \mathcal{I}(\tilde{\mathbf{X}}) - \mathcal{A}_{\Omega, \mathbf{H}}(\tilde{\mathbf{X}}) + \mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{H}), \quad (3.9)$$

where \mathcal{I} is an identity operator, and $\mathcal{H}(\cdot)$ denotes the updating process, which combines the known entries in matrix \mathbf{H} and the estimated entries in matrix $\tilde{\mathbf{X}}$.

Next, the estimated matrix is replaced with $\mathcal{H}(\mathbf{X})$ and the iteration is repeated starting from Eq (3.7). The estimation is evaluated by $\epsilon(\mathbf{X}) \triangleq \|\mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{X}) - \mathcal{A}_{\Omega, \mathbf{H}}(\mathbf{H})\|_F$, and missing entries in matrix \mathbf{H} can be approximated until $\epsilon(\mathbf{X})$ converges. For convenience, we denote the whole iteration process of the constrained approximation as $\mathcal{T}(\mathbf{H}, \lambda)$, which will be used in the next section for the HRV spectrum estimation.

It is noted that the matrix approximation in Eq (3.6) is different from the low-rank MC as presented in Eq (3.2), where the low-rank MC requires fixing the known entries for the estimation, while the matrix approximation in Eq (3.6) minimises the objective function with Frobenius norm to the constraint function. In particular, the technique of matrix approximation with an interested zone can also be extended as matrix completion, which develops a new iteration process separate from the traditional SVT technique; more details of the iteration process can be found in [26] and will also be discussed in Section 3.3.

The process of approximating missing entries in this context is driven by the data itself through an iterative approach. It is important to acknowledge that in HRV data analysis, certain mathematical models can be leveraged to represent the spectrum effectively. One such model, the Gaussian model [31], has been highlighted for its capability to accurately characterise HRV spectra and elucidate the relationships between various spectrum components. In the subsequent subsection, we will apply the Gaussian model to refine the RMC process by constructing a new matrix that possesses significantly reduced dimensions. This approach is anticipated to not only enhance the efficiency of the estimation process, but also to improve its overall performance by providing a more focused and computationally manageable framework for addressing missing data within the HRV spectrum analysis.

3.3. Model-based refined MC for HRV spectrum estimation

The HRV data spectrum is characterised by distinct frequency bands, each with unique physiological implications. As illustrated in Figure 1, the HF band captures the influence of respiration on heart rate, a phenomenon known as respiratory sinus arrhythmia (RSA); conversely, the LF band encompasses oscillations associated with blood pressure regulation and vasomotor tone, often referred to as Mayer waves. These distinctions provide a foundation for modelling the HRV spectrum, allowing for a nuanced understanding of the underlying physiological processes. With these considerations, we model the spectrum of HRV data as follows.

We assume the matrix $\mathbf{S} = \{s_{i,\ell}\}_{i=1,\dots,n_r, \ell=1,\dots,n_{fr}} \in \mathbb{R}^{n_r \times n_{fr}}$ consists of n_r spectra that are derived from the HRV data, where $\mathbf{s}_i = [s_{i,1}, \dots, s_{i,n_{fr}}]^T \in \mathbb{R}^{n_{fr}}$ is a vector of the spectrum. In particular, for each $\mathbf{s}_i \in \mathbf{S}$, it can be modelled with Gaussian functions [31], and the entry $s_{i,\ell}$ can be represented as,

$$s_{i,\ell}^m(f, A_{i,\ell}, \sigma_{i,\ell}, f_{i,\ell}) = \frac{A_{i,\ell}}{\sqrt{2\pi}\sigma_{i,\ell}} \exp\left(-\frac{(f-f_{i,\ell})^2}{2\sigma_{i,\ell}^2}\right), \quad (3.10)$$

where $f \in [f_{l,i}, f_{r,i}]$ indicates the interval of the frequency band of interests, i.e., LF or HF spectrum. Consequently, $A_{i,\ell} > 0$ is the weight of amplitude, $f_{i,\ell} > 0$ is the peak position of frequencies in the spectrum, and $\sigma_{i,\ell} > 0$ is its standard deviation for this frequency. For simplicity of notations, $s_{i,\ell}^m$ is used to represent $s_{i,\ell}^m(f, A_{i,\ell}, \sigma_{i,\ell}, f_{i,\ell})$ when no confusion arises, and $\tilde{\mathbf{s}}_i^m = \{s_{i,\ell}^m\}_{\ell=1,\dots,n_{fr}}$ indicates the modelled spectrum.

It is noted that the HRV spectrum can be characterised with different frequency components, and they can be simulated by changing the frequency intervals of the modelling, after which the complete spectrum can be obtained by combining these different frequency components. Next, we use the modelled information to define a new matrix for the estimation of missing entries. For the convenience of presentation, it is assumed that some entries of matrix \mathbf{S} are missing, and the set $\mathcal{S}_m = \{\mathbf{s}_j\}_{j=1,\dots,n_q}$ contains the rows in \mathbf{S} with missing values.

For any $\mathbf{s}_j \in \mathcal{S}_m$, we model the spectrum and obtain $\tilde{\mathbf{s}}_j^m$ with Eq (3.10), and then calculate the correlation between $\tilde{\mathbf{s}}_j^m$ and $\tilde{\mathbf{s}}_q^m$ spectra, where $\tilde{\mathbf{s}}_q^m$ is the modelled information of spectrum in \mathbf{S} , and the correlation is computed as follows,

$$\gamma(s_{j,\ell_t}^m, s_{q,\ell_t}^m) = \frac{\sum_{\ell_t=1}^{N_1} (s_{j,\ell_t}^m - \bar{s}_j^m)(s_{q,\ell_t}^m - \bar{s}_q^m)}{\sqrt{\sum_{\ell_t=1}^{N_1} (s_{j,\ell_t}^m - \bar{s}_j^m)^2} \sqrt{\sum_{\ell_t=1}^{N_1} (s_{q,\ell_t}^m - \bar{s}_q^m)^2}}, \quad (3.11)$$

where $\bar{s}_j^m = \frac{1}{N_1} \sum_{\ell_t=1}^{N_1} s_{j,\ell_t}^m$ and $\bar{s}_q^m = \frac{1}{N_1} \sum_{\ell_t=1}^{N_1} s_{q,\ell_t}^m$ are means of the two spectra, respectively, $\ell_t = 1, \dots, N_1$, $N_1 \leq n_{fr}$ is the index of known components of the spectrum, and $\gamma_{j,q} = \gamma(s_{j,\ell_t}^m, s_{q,\ell_t}^m)$ indicates the relationship between the two spectra.

After calculating $\gamma_{j,q}$ for the \mathbf{s}_j spectrum, a new vector $\mathbf{t}_j = [\gamma_{j,1} \ \dots \ \gamma_{j,n_r-1}]$ can be obtained, we sort \mathbf{t}_j and select the K highest-ranked elements for \mathbf{s}_j , and then a set that contains the identified indices can be formed as,

$$\mathcal{G}_K(\mathbf{s}_j) = \{g_k \mid t_{j,g_k} \in \mathbf{t}_j, g_k \in \mathcal{J}, k = 1, \dots, K\}, \quad (3.12)$$

where $\mathcal{J} = 1, \dots, n_r - 1$ denotes the index of spectrum in matrix \mathbf{S} .

Next, a refined matrix $\mathbf{S}_{\mathcal{G}_K} \in \mathbb{R}^{K \times n_{fr}}$ can be formulated by combining \mathbf{s}_j and the identified HRV spectra. Consequently, the optimisation in Eq (3.6) can be updated as

$$\begin{aligned} \min \quad & \|\mathcal{A}_{\tilde{\Omega}, \mathbf{S}_{\mathcal{G}_K}}(\mathbf{X}_{\mathcal{G}_K}) - \mathcal{A}_{\tilde{\Omega}, \mathbf{S}_{\mathcal{G}_K}}(\mathbf{S}_{\mathcal{G}_K})\|_F \\ \text{s.t.} \quad & p(\mathbf{X}_{\mathcal{G}_K}) \leq 0, \end{aligned} \quad (3.13)$$

where $\tilde{\Omega}$ is the set of observed entries in the refined matrix, $\mathcal{A}_{\tilde{\Omega}, \mathbf{S}_{\mathcal{G}_K}}(\cdot)$ is the updated operator to model the missing values, $\mathbf{X}_{\mathcal{G}_K}$ indicates the estimated matrix, and the spectral norm $p(\mathbf{X}_{\mathcal{G}_K}) = \|\mathbf{X}_{\mathcal{G}_K}\|_2 - \lambda$ can be used for the constraint.

For the estimation of missing entries in the refined matrix, we first approximate the following constrained full matrix,

$$\begin{aligned} \min \quad & \|\mathbf{X}_{\mathcal{G}_K} - \mathbf{S}_{\mathcal{G}_K}\|_F \\ \text{s.t.} \quad & p(\mathbf{X}_{\mathcal{G}_K}) \leq 0, \end{aligned} \quad (3.14)$$

where the soft imputing method as described in Eq (3.8) can be used to solve the estimation in Eq (3.14). Then, by updating the parameter λ in the constraint function, missing entries in the refined matrix can be estimated using the iteration process as described in Section 3.2. Algorithm 1 presents the pseudo code of the proposed RMC method. The hyperparameters tolerance e_{tol} and λ_{tol} are set as 10^{-8} and 10 according to the configuration in [26].

As an illustration, Figure 1 demonstrates the concept of estimating HRV spectrum using the RMC technique, where the matrix is constructed with m spectra that are derived from HRV data segments. The x -axis of the matrix indicates the frequency band of interests, and y -axis shows the indices of data segments. Suppose the HF band of s_3 spectrum is affected by measurement uncertainties, which will be estimated using the LF band. By using the RMC model, we model the spectra in the matrix, such as the Mayer wave and the RSA component and use the LF spectrum as a reference to identify relevant elements in the matrix. Then, these identified spectra, such as vectors s_2 , s_5 , and s_m , along with s_3 form a new matrix, where the HF band of s_3 can be estimated using the RMC method described in the above iteration procedures.

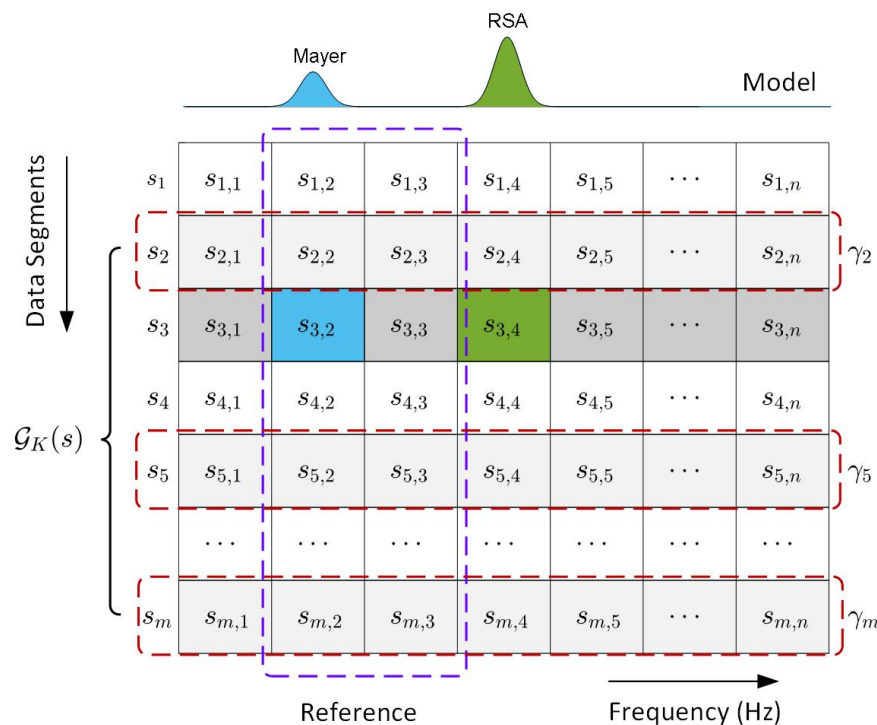


Figure 1. Illustration of MC and RMC using model information for HRV spectrum estimation.

Algorithm 1: Estimation of HRV spectrum using our developed RMC method.

Input : Matrix $\mathbf{S} \in \mathbb{R}^{n_r \times n_{fr}}$, $\mathcal{A}_{\Omega, \mathbf{H}}$, \mathcal{S}_m ,

tolerance e_{tol} and λ_{tol} .

Output : Estimated spectrum matrix \mathbf{X} .

```

1 Initialisation:  $\mathbf{s}_j \in \mathcal{S}_m$ ,  $\lambda \leftarrow 0$ ,  $\lambda_{prev} \leftarrow 10^6$ ;
2 for  $i \leftarrow 1$  to  $n_r - 1$  do
3   | Compute  $s_i^m(f)$  by modelling  $s_i(f) \in \mathbf{S}$ ;
4   | Calculate  $\gamma(s_{i, \ell_i}^m, s_{j, \ell_j}^m)$  between the spectra;
5 end
6 Sort  $\gamma(s_{i, \ell_i}^m, s_{j, \ell_j}^m)$  and obtain  $\mathcal{G}_K(\mathbf{s}_j)$ ;
7 Update  $\mathbf{S} \leftarrow \mathbf{S}_{\mathcal{G}_K}$ , and  $\mathcal{A}_{\Omega, \mathbf{H}} \leftarrow \mathcal{A}_{\Omega, \mathcal{G}_K}$ ;
8 Assign  $\mathbf{S} \leftarrow \mathcal{A}_{\Omega, \mathbf{H}} \mathbf{S}$ ,  $\lambda_{min} \leftarrow 0$ ,  $\lambda_{max} \leftarrow \|\mathbf{S}\|_*$ ;
9 while  $\epsilon(\mathbf{X}) > e_{tol}$  or  $|\lambda - \lambda_{prev}| > \lambda_{tol}$  do
10  | Update  $\lambda_{prev} \leftarrow \lambda$ ;
11  | Compute and update  $\lambda \leftarrow (\lambda_{min} + \lambda_{max})/2$ ;
12  | Estimate and update  $\mathbf{X} \leftarrow \mathcal{T}(\mathbf{S}, \lambda)$ ;
13  | Compute the estimation error  $\epsilon(\mathbf{X})$ ;
14  | if  $\epsilon(\mathbf{X}) > e_{tol}$  then
15  |   | Update  $\lambda_{min} \leftarrow \lambda$ ;
16  | else
17  |   | Update  $\lambda_{max} \leftarrow \lambda$ ;
18  | end
19 end
20 Output the estimated matrix  $\mathbf{X}$ .
```

3.4. Evaluation of estimation performance

The developed RMC method for HRV spectrum estimation is used to analyse a variety of datasets, as discussed in the next section. To evaluate the model performance on spectrum estimation across different subjects and datasets, we use the normalised root mean square error (NRMSE) as the performance indicator. For the estimation of missing entries in the \mathbf{s}_j spectrum, the NRMSE index can be calculated as follows [23],

$$\text{NRMSE} = \sqrt{\frac{\frac{1}{N_f} \sum_{k=1}^{N_f} (\hat{s}_{j,k} - s_{j,k})^2}{\frac{1}{N_f - 1} \sum_{k=1}^{N_f} (s_{j,k} - \bar{s}_j)^2}}, \quad (3.15)$$

where, $s_{j,k} \in \mathbf{S}$ is the original spectrum, $j = 1, 2, \dots, n_r$ is the index of the spectrum, and $k = 1, 2, \dots, N_f$ indicates the data points of missing entries in the spectrum. $\hat{s}_{j,k}$ is the estimated spectrum, and $\bar{s}_j = \frac{1}{N_f} \sum_{k=1}^{N_f} s_{j,k}$ is the mean value of the spectrum. The calculation of this index normalises the root mean square error with the standard deviation of the spectrum, which enables to evaluate the model performance across different ECG datasets using an indicator with the same scale.

4. Data analysis

4.1. Datasets

Generally, HRV data can be derived from ECG signals, and in the current study, we retrieved five widely used benchmark ECG datasets to evaluate the performance of our developed RMC model. These datasets are summarised as follows.

- (i) Combined measurement of ECG, breathing and seismocardiograms (*CEBSDB*) [32]: This dataset consists of ECG recordings sampled from 20 healthy volunteers. After excluding a low-quality recording ('m018') due to the issue of electrode contact, the ECG signals collected from 19 subjects were used for the study, which have a sampling duration of approximately 50 minutes.
- (ii) Supraventricular arrhythmia database (*SADB*) [33]: This dataset includes 78 ECG recordings sampled from subjects with supraventricular arrhythmias, and each signal has a recording time duration of 30 minutes.
- (iii) St Petersburg INCART 12-lead Arrhythmia Database (*SPIADB*) [34]: This dataset consists of 75 ECG recordings with a duration of 30 minutes, which were collected from patients undergoing tests for coronary artery disease.
- (iv) European ST-T Database (*ESTDB*) [35]: This database is widely used for the evaluation of ST and T-wave changes in the ECG morphologies, and includes 90 annotated excerpts of ambulatory ECG recordings with two-hour sampling durations.
- (v) Apnea-ECG Database (*APEDB*) [36]: This data consists of 70 ECG recordings with a set of apnea annotations, and each of the recordings has an approximate duration of 7 to 10 hours.

These datasets include ECG recordings with diverse characteristics, such as data sampled from healthy volunteers, patients with arrhythmias, coronary artery disease, ST and T-wave changes, and apnea. Therefore, the five ECG datasets will generate different types of HRV data, and in combination allow a comprehensive evaluation of our developed method for spectrum estimation. It is noted that the ECG signals in these datasets have varied lengths of sampling durations, and we use the first hour of the signal only in instances of a long-time sampling duration.

4.2. Signal processing

To generate HRV data for this study, we first identify R-peaks from these ECG recordings by analysing the QRS complex. Figure 2(a) shows the identified R-peaks in the ECG recording using the QRS detector as developed in [37]. After identifying the R-peaks, a sequence of RR intervals can be calculated from timestamps of the consecutive peak values. Next, we implement outlier detection in the derived RR intervals using an open-source benchmarked toolbox [38], where the outliers are defined as data points that are too close together, or large RR intervals caused by gaps, or artefact annotations of the signal. Figure 2(c) illustrates the detected outliers, which are processed by removal and interpolated to generate the preprocessed sequence.

We note that although a variety of techniques have been developed to detect the R-peaks [39, 40], it is still challenging to accurately identify the peak values for different types of ECG recordings. For

example, as shown in Figure 2(a), the R-peaks are correctly identified for the ECG signal that is sampled from a healthy subject. However, as shown in Figure 2(b), the R peaks in waves *A* and *B* are wrongly identified for the ECG signal that is collected from an abnormal subject. The comparison of R-peak detection indicates that it may not always be reliable to use the technique to identify peak values for different types of ECGs.

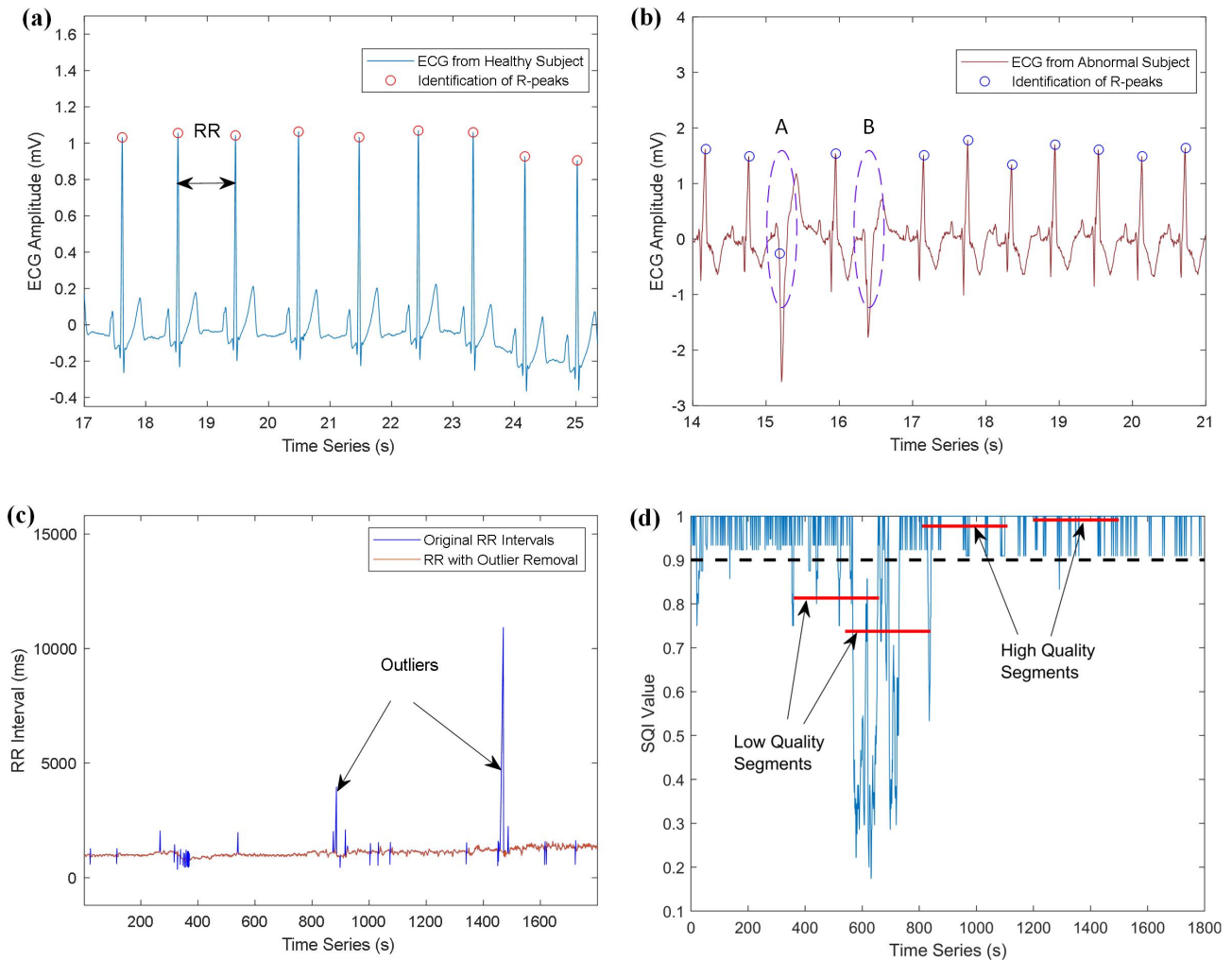


Figure 2. ECG signal processing and signal quality analysis of HRV data. (a) R-peaks identification for ECG recordings sampled from a healthy subject. (b) R-peaks identification for ECG recording sampled from an abnormal subject. (c) Outlier detection in RR intervals. (d) Signal quality analysis for the HRV data.

To obtain high-quality HRV data, this study utilises two widely recognised detectors for identifying peak values: the ‘jqrs’ detector [37] and the ‘wqrs’ detector [40]. Subsequently, a signal quality index (SQI) is computed for each detected R-peak by contrasting annotations from both detectors [41]. Following this, the average SQI value of R-peaks within 5-minute HRV data segments is calculated, serving as a quality indicator for each segment. For instance, a segment is deemed high-quality if its

SQI indicator exceeds 0.9; otherwise, it is considered low-quality. As illustrated in Figure 2(d), two data segments exhibit SQI values of 0.813 and 0.737, categorising them as low-quality. By calculating the average SQI values, the low-quality and high-quality data segments can be efficiently identified, and the low-quality segments will not be used for further analysis.

4.3. Calculation of the PSD spectrum

The preprocessed HRV data is then resampled with a frequency of 4 Hz, and a 4th order Butterworth filter is used to remove noises with the pass band of 0.03 to 0.9 Hz [42]. The filtered data is used to derive HRV spectrum in the frequency domain, and Welch's algorithm with an overlap of 50% is used for the calculation [42]. Next, the HRV spectrum matrix can be obtained by stacking all the calculated spectra. Figure 3(a) shows the three-dimensional plots of PSD spectra that are derived from different data segments. It can be seen from Figure 3(a) that the PSD spectra have similar patterns in the waveforms, i.e., LF and HF components, indicating the low-rank property of the HRV matrix derived from the PSD spectra.

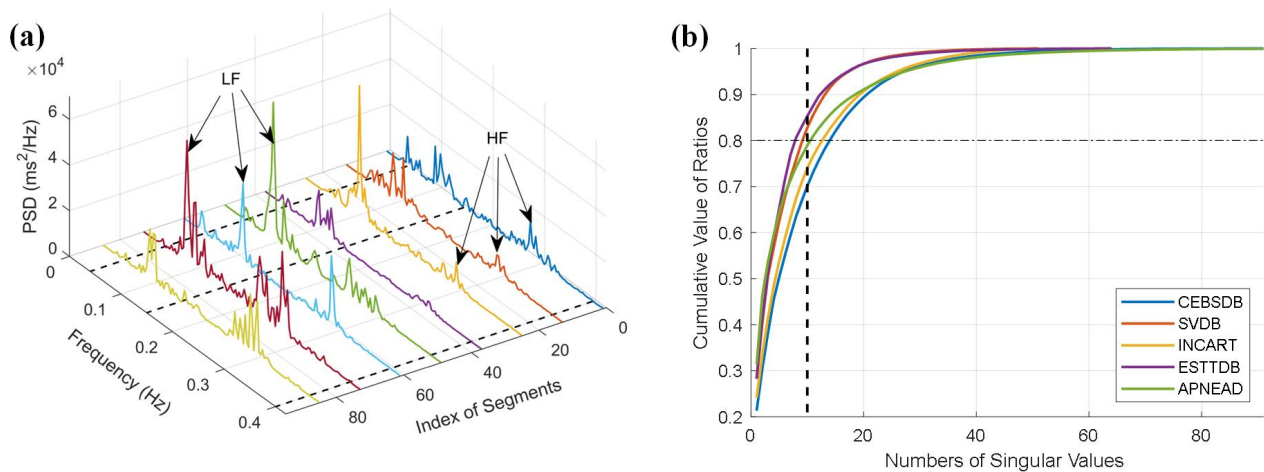


Figure 3. HRV spectrum and the cumulative ratios. (a) PSD spectra of different data segments. (b) Distributions of cumulative ratios between singular values of the nuclear norm.

In a further step, we calculate the singular values of the derived HRV matrix, and compute the ratio between singular values and the nuclear norm of the spectrum matrix, which is used as an indicator of the rank property of the matrix [23]. As shown in Figure 3(b), we demonstrate the low-rank property of the HRV data by calculating the cumulative ratios of singular values. It can be seen from Figure 3(b) that the summation of the top ten singular values accounts for approximately 80% of the nuclear norm, indicating the low-rank properties of the HRV spectrum matrices that are derived from these five ECG datasets.

5. Results and discussion

5.1. Uncertainty estimation for PSD spectrum

The derived HRV matrix is then used to evaluate the performance of our developed model for uncertainty estimation. We note that the uncertainties or distortions of HRV spectrum can result from many different factors, such as noises, artefacts, missing values, and a variety of computational approaches (e.g., QRS detection and data interpolation). Without loss of generality, we simulate the uncertainties as missing values in the HRV spectrum, and evaluate the performance of our developed method for the uncertainties estimation. As an example in Figure 4, we show the estimation of uncertainty in HRV spectrum, which is represented as missing values and is mostly in the HF of the spectrum. Then, we use the rest of the spectrum (i.e., LF band) to estimate the HF band. It can be seen from Figure 4 that the estimated HF spectrum matches well with the original PSD spectrum, which efficiently estimated the waveform of the HF in the spectrum, but the estimation with the MC model introduced large variations around the frequency of 0.25 Hz.

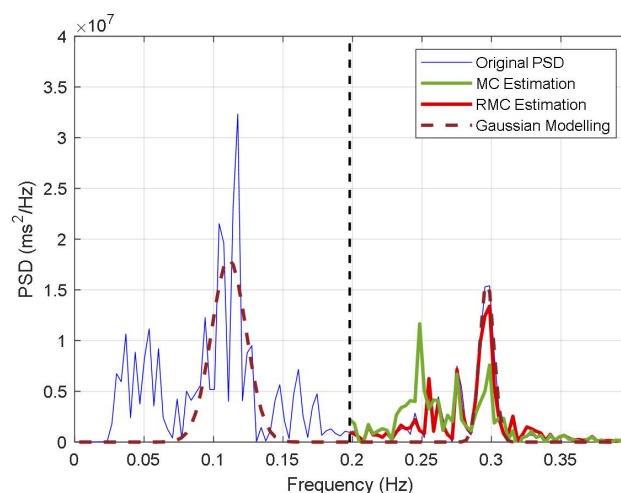


Figure 4. Comparison of HRV spectrum estimation using different methods.

We use the developed RMC method to estimate uncertainties in the spectrum. We simulate the spectrum with the model described in Eq (3.10), and then use the modelled signal to identify relevant data segments in the HRV matrix. As an illustration in Figure 4, we show the signal modelling for the PSD spectrum of 31st data segment of the matrix. It can be seen that the modelled signal efficiently represents the characteristics of the spectrum, such as the LF and HF.

Next, we calculate correlation coefficients between different data segments using the modelled signals, and select the data with high coefficients to reconstruct the HRV matrix. Figure 5(a) shows the calculated coefficients for segments in the original data matrix, and Figure 5(b) illustrates estimation errors for different combinations of selected data segments. It can be seen from Figure 5(b) that the RMC model obtains the least estimation error of 0.222 when using nine data segments, and it is smaller than the estimation error of 0.830 using the traditional MC method. As shown in Figure 5(a), we highlight the selected data segments with square markers. It can be seen from Figure 5(a) that the selected important data elements include five neighbours of the target data segment; this is consistent

with the rational approach that missing values can be generally imputed using nearest neighbours. We also note that the model identifies four segments as important data, which are not the neighbours of the target segment, indicating that some data with far distances in the sequence may also provide valuable information for the estimation of missing entries.

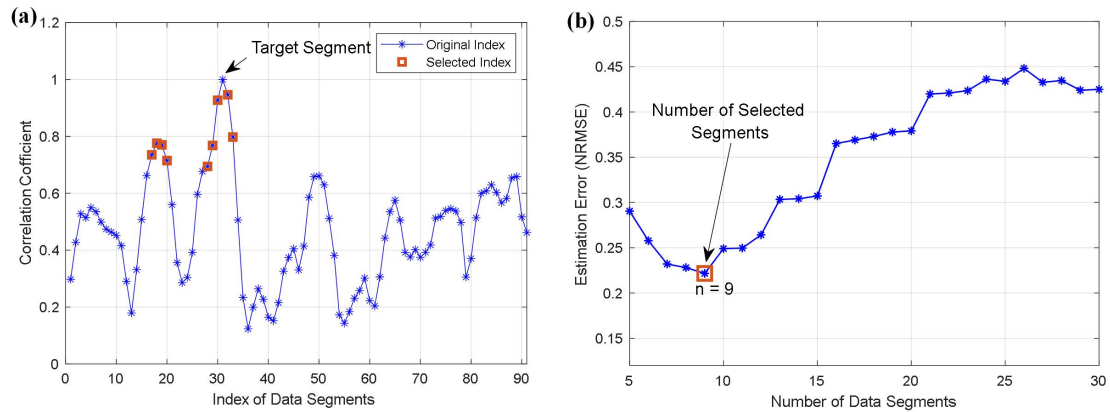


Figure 5. HRV spectrum estimation using matrix completion. (a) The correlation coefficients of different HRV segments. (b) Estimation errors for different combinations of data segments.

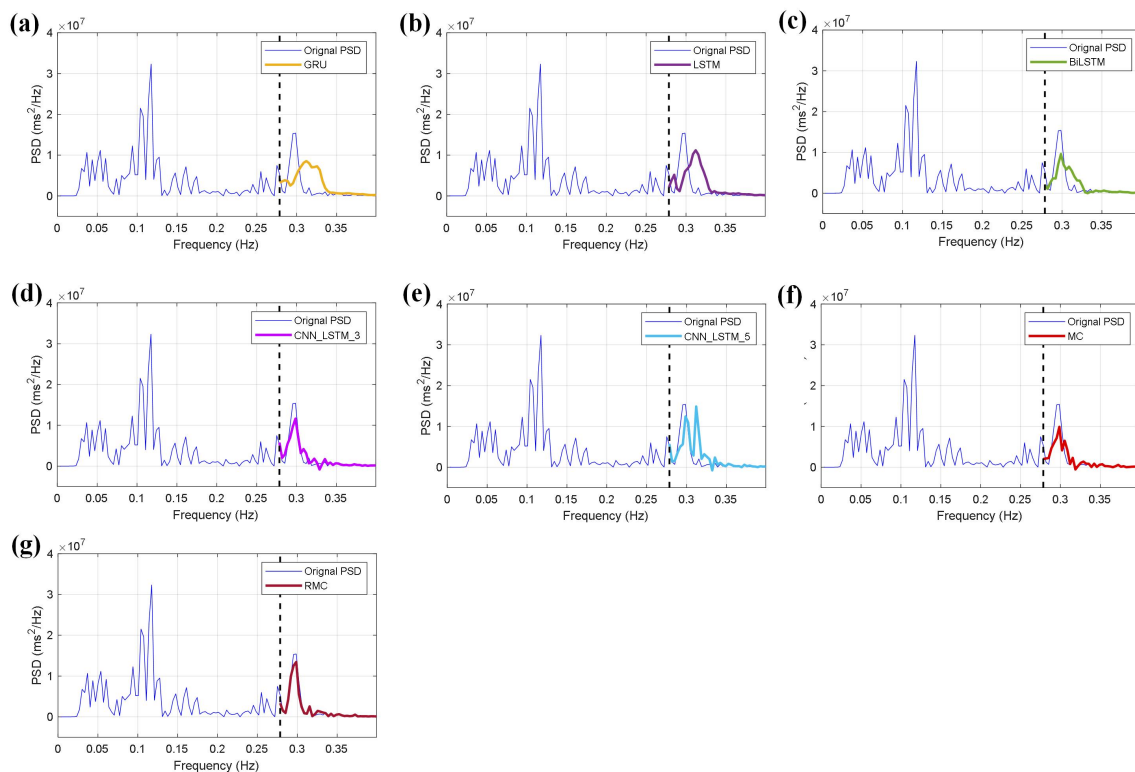


Figure 6. Uncertainty estimation of HRV spectrum using different models with the masking ratio of 30%: (a) The GRU model, (b) the LSTM model, (c) the BiLSTM model, (d) the CNN_LSTM_3 model, (e) the CNN_LSTM_5 model, (f) the MC method, and (g) the RMC method.

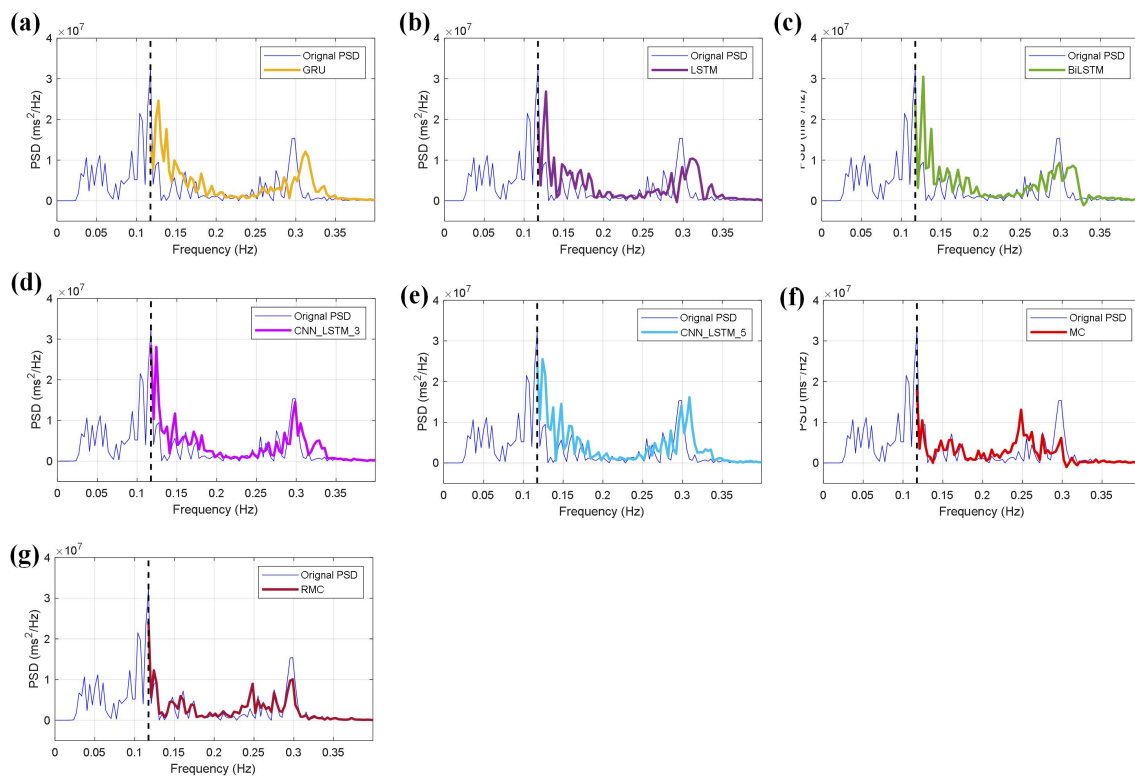


Figure 7. Uncertainty estimation of HRV spectrum using different models with the masking ratio of 70%: (a) The GRU model, (b) the LSTM model, (c) the BiLSTM model, (d) the CNN_LSTM.3 model, (e) the CNN_LSTM.5 model, (f) the MC method, and (g) the RMC method.

In a further step, we evaluate the model performance on uncertainty estimation in different conditions, and we apply different masking ratios on the HRV spectrum. As shown in Figures 6 and 7, our developed RMC method efficiently estimates missing values for the PSD spectra with masking ratios of 30% (Figure 6(g)), and 70% (Figure 7(g)), which have estimation errors of 0.221 and 0.426 separately. The two experiments show the robustness of our developed RMC method for uncertainty estimation in the HRV spectrum.

5.2. Comparison methods and results

The current study focused on estimating uncertainties of the HRV spectrum, and we compared our developed methods with different types of regression machine learning models; in particular, the deep recurrent neural networks have been demonstrated with excellent performance on prediction or regression analysis. To be specific, we used the gated recurrent units (GRU), the long short-term memory (LSTM) model, and the bidirectional LSTM (BiLSTM) for the comparison study. We note that convolutional neural networks (CNN) have shown efficiency in extracting features from data sequences [43]. Therefore, other than the GRU, LSTM, and BiLSTM, we used two hybrid models for the comparison study.

For the three types of recurrent neural networks, i.e., the GRU, LSTM, and BiLSTM, they have 120 hidden units for the regression analysis [44]. For the two types of hybrid models, the hyperparameters are

tuned with trial-and-error search; the first one (CNN_LSTM_3) uses three 1D-CNN layers with 32, 64, and 64 filters, each filter with a size of 3 [45]. Each convolutional layer is followed by a rectified linear unit (ReLU) function, and a batch normalisation layer, which is then flattened and connected with a LSTM layer for the regression analysis. The second type of hybrid models (CNN_LSTM_5) consists of five CNN layers, which have 32, 32, 64, 64, and 64 filters, and also uses a LSTM layer for the regression analysis. The deep learning models are trained with Adam optimizer with an initial learning rate of 0.005, the maximum epoch size is 100, and the batch size is set as 20. For the prediction analysis, we use data points from previous steps, i.e., 10 steps [46], to predict the current value in the data sequence.

The estimation results of these regression models are demonstrated in Figures 6 and 7, which correspond to the masking ratios of 30% and 70%, respectively. As a comparison study, we also include the estimation results from the traditional MC method. It can be seen from Figures 6 and 7 that these machine learning models have small errors on the spectrum estimation; in particular, the BiLSTM and hybrid models efficiently estimate the peak values around 0.3 Hz for the two masking ratios. Compared with the deep learning models and the MC method, our developed RMC method has more efficient estimation performance, which obtains the least estimation errors of 0.221 and 0.426 for the 30% and 70% masking ratios, respectively.

5.3. Statistical results

We implement the estimation of missing entries for all ECG recordings in the five benchmark datasets using our developed RMC method and the regression models. Tables 1–3 show the mean value and standard deviation of estimation errors for the five benchmark datasets with 30%, 50%, and 70% masking ratios. The minimum estimation error for each dataset is bold-faced. It can be seen from Tables 1–3 that, compared with the five machine learning models and the traditional MC method, our developed RMC method obtains the least estimation error for all of the five ECG datasets, and the increased masks ratios occurred alongside an increase in the estimation errors, which have the values of 0.286 ± 0.194 , 0.312 ± 0.203 , and 0.379 ± 0.228 for the CEBSDB with the three masking ratios.

Table 1. Comparison of estimation errors for different methods on benchmark datasets with 30% masking ratio.

Datasets	GRU	LSTM	BiLSTM	CNN_LSTM_3	CNN_LSTM_5	MC	RMC
CEBSDB	1.379 ± 1.112	1.346 ± 0.970	0.671 ± 0.355	0.672 ± 0.459	0.770 ± 0.692	0.393 ± 0.291	0.286 ± 0.194
SADB	1.389 ± 1.751	1.318 ± 1.280	0.733 ± 0.901	0.746 ± 0.902	0.902 ± 1.808	0.332 ± 0.253	0.304 ± 0.199
SPIADB	1.241 ± 0.776	1.211 ± 0.609	0.647 ± 0.289	0.655 ± 0.375	0.709 ± 0.378	0.361 ± 0.539	0.305 ± 0.327
ESTDB	1.317 ± 1.096	1.297 ± 1.043	0.687 ± 0.430	0.691 ± 0.456	0.753 ± 0.549	0.494 ± 0.464	0.329 ± 0.251
APEADB	1.283 ± 0.740	1.268 ± 0.646	0.644 ± 0.263	0.663 ± 0.337	0.727 ± 0.392	0.426 ± 0.357	0.290 ± 0.228

Table 2. Comparison of estimation errors for different methods on benchmark datasets with 50% masking ratio.

Datasets	GRU	LSTM	BiLSTM	CNN_LSTM_3	CNN_LSTM_5	MC	RMC
CEBSDB	1.061 ± 0.234	1.091 ± 0.275	0.621 ± 0.163	0.619 ± 0.164	0.663 ± 0.175	0.432 ± 0.192	0.312 ± 0.203
SADB	1.067 ± 0.149	1.093 ± 0.190	0.670 ± 0.153	0.673 ± 0.129	0.714 ± 0.136	0.386 ± 0.254	0.338 ± 0.226
SPIADB	1.076 ± 0.168	1.097 ± 0.201	0.657 ± 0.150	0.662 ± 0.115	0.706 ± 0.128	0.388 ± 0.529	0.310 ± 0.290
ESTDB	1.113 ± 0.707	1.138 ± 0.692	0.680 ± 0.306	0.704 ± 0.427	0.745 ± 0.443	0.590 ± 0.446	0.336 ± 0.276
APEADB	1.059 ± 0.292	1.085 ± 0.334	0.642 ± 0.216	0.648 ± 0.168	0.687 ± 0.175	0.547 ± 0.422	0.312 ± 0.256

Table 3. Comparison of estimation errors for different methods on benchmark datasets with 70% masking ratio.

Datasets	GRU	LSTM	BiLSTM	CNN_LSTM_3	CNN_LSTM_5	MC	RMC
CEBSDB	1.350 ± 0.591	1.446 ± 0.676	0.926 ± 0.563	0.906 ± 0.451	1.030 ± 0.622	0.633 ± 0.202	0.379 ± 0.228
SADB	1.131 ± 0.361	1.247 ± 0.422	0.983 ± 0.652	0.819 ± 0.238	0.873 ± 0.284	0.539 ± 0.318	0.397 ± 0.262
SPIADB	1.179 ± 0.398	1.325 ± 0.544	1.019 ± 0.674	0.828 ± 0.230	0.890 ± 0.308	0.502 ± 0.406	0.338 ± 0.247
ESTDB	1.193 ± 0.418	1.335 ± 0.543	1.028 ± 0.743	0.849 ± 0.304	0.903 ± 0.369	0.728 ± 0.356	0.337 ± 0.252
APEADB	1.144 ± 0.368	1.266 ± 0.470	0.958 ± 0.689	0.811 ± 0.256	0.864 ± 0.315	0.694 ± 0.409	0.315 ± 0.246

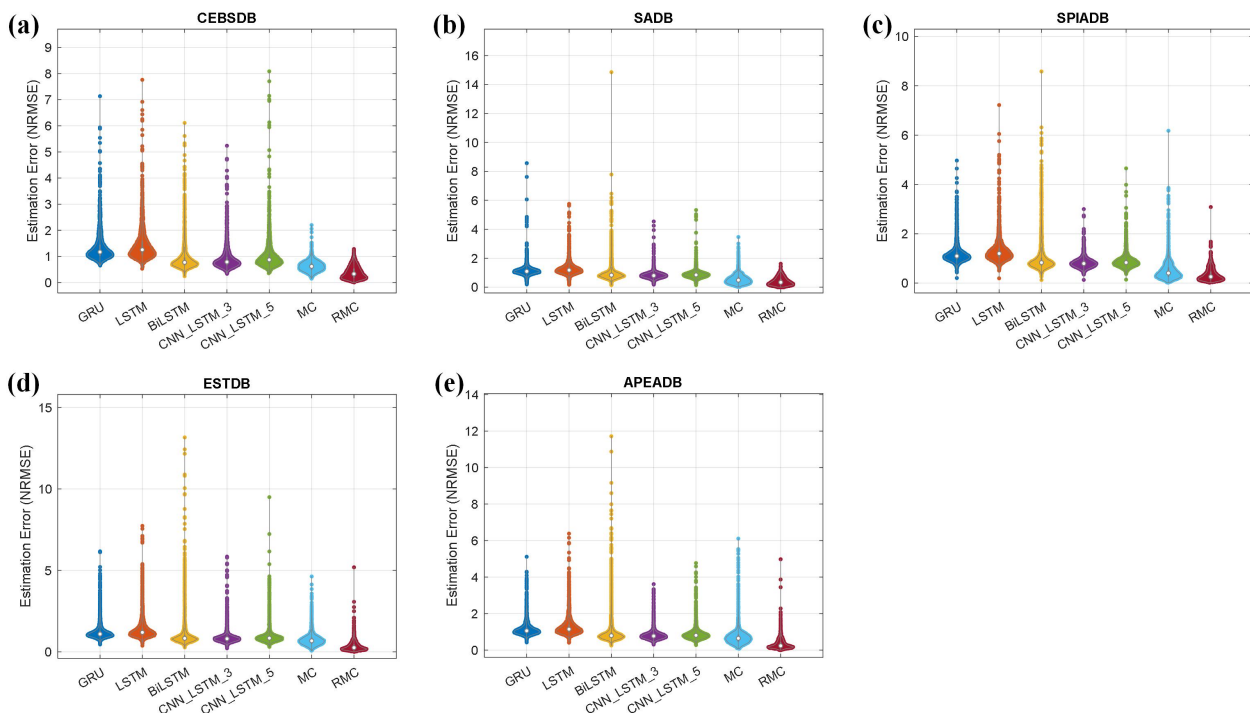


Figure 8. Distributions of estimation errors of different methods on five ECG benchmark datasets, (a) CEBSDB, (b) SADB, (c) SPIADB, (d) ESTDB, and (e) APEADB.

Additionally, we provide violin plots as shown in Figure 8 to demonstrate the distributions of estimation errors for the seven models. It can be seen from Figure 8 that our developed RMC method obtains the least estimation errors with median values of 0.330, 0.337, 0.260, 0.259, and 0.235 for the five benchmark datasets. In particular, we note that compared with the five machine learning models and the traditional MC method, the RMC method has much smaller variations of the estimation errors for all the ECG datasets, which indicate the stability and robustness of our developed RMC method for the spectrum estimation.

5.4. Comparison of computational costs

We calculate the computation cost for missing entry estimation using the RMC method and the other six methods. All of the estimation tasks are implemented with Matlab R2022a, Intel(R) Core(TM) i7-1165G7@2.8 GHz, and 32 GB RAM. We calculate the mean value and standard deviation of computation time for all HRV data segments in the benchmark dataset. Table 4 presents the comparison of computation cost for all the estimation methods. We note that the five deep learning models have an increasing trend of computation cost with the increasing model complexity. From Table 4 it can be seen that among these deep learning models, the GRU model has the least computation time with 1.117 ± 0.460 s. Compared with the deep learning models and the MC method, our developed RMC method has the least computation time with 0.105 ± 0.096 s. The results demonstrate the efficiency of our developed RMC method for HRV spectrum estimation.

Table 4. Comparison of computation time for different estimation methods.

Methods	Computation cost (s)
GRU	1.117 ± 0.460
LSTM	1.357 ± 0.704
BiLSTM	2.298 ± 2.091
CNN_LSTM_3	2.484 ± 2.218
CNN_LSTM_5	2.889 ± 0.909
MC	0.669 ± 0.340
RMC	0.105 ± 0.096

5.5. Discussion

Spectrum metrics of HRV data are important indicators for monitoring physiological and pathological conditions. However, HRV data are sensitive to a variety of uncertainties in the measurements, such as motion artefacts and missing values. Although numerous methods have been developed to address uncertainties of HRV analysis, it is difficult to remove artefacts without affecting the signal waveforms. We propose to estimate uncertainties of HRV data from a new view in the spectrum analysis; in particular, we consider the characteristics of LF and HF bands of HRV spectrum. With extensive experimental studies, we show that the PSD data demonstrate similarities between spectrum waveforms, indicating the low-rank property of the spectrum matrix, which enables us to use the MC technique for the uncertainty estimation.

A variety of computational techniques have been developed for missing value estimation, for example, different types of interpolation methods [47] and machine learning models [48], including linear interpolation, k-nearest neighbour (kNN), multiple imputation by chained equation (MICE), neural networks, and random forest. However, we note that deep neural networks (DNNs), particularly with recent advancements, have shown superior performance in estimating sequential data [49]. Therefore, we compared our proposed RMC method with different types of DNNs. We also acknowledge more advanced predictive models that have been developed recently, such as Transformer [50], Informer [51], and TimesNet [52], which have not been used for HRV uncertainty estimation, motivating us to implement more comprehensive comparisons in our future study.

It is a challenging task of estimating missing entries for sparse matrices. We note the performance of low-rank MC method in handling sparse matrices, which has been demonstrated in previous literature [53–55]. In this study, we proposed a refined version of the traditional MC method, and our developed RMC method was used for HRV uncertainty estimation. The results indicated that compared with different types of DNN models, our RMC method demonstrated promising performance across all benchmark datasets. The improved model performance can be attributed to the incorporation of specific characteristics of the HRV spectrum into our modelling approach. For instance, the LF of the HRV spectrum exhibits a peaked Mayer waveform, while the HF displays a dominant RSA waveform. These peaked waveforms, observed within a limited number of data points, can pose challenges for DNN models in uncertainty estimation, and therefore lead to reduced performance.

There are many different types of uncertainties in the analysis of HRV data, such as measurement noises and artefacts in data acquisition and computational uncertainties in data processing (e.g., data interpolation, and data resampling). Without loss of generality, we simulate the uncertainties of HRV spectrum as missing values, which can also be used to indicate the distorted entries affected by motion artefacts or data interpolation. We show the advantages and robustness of our developed RMC method for uncertainty estimation on a range of benchmark datasets. To perform a fair comparison, the NRMSE indicator was used to evaluate the performance of different models on five ECG datasets; we will explore other metrics for evaluating different types of uncertainties in our future studies. In addition, we note that HRV spectrum is derived from NN intervals of ECG signals. However, NN interval outliers, caused by missed or false beat detections, could impact the HRV spectrum. Our next step research aims to thoroughly investigate these uncertainties and develop robust MC with other advanced techniques, such as ℓ_p minimisation [27] and adaptive weighting strategy [56] to mitigate the effects of outliers.

6. Conclusions

This paper presents a new approach to address the challenges of uncertainty estimation in the HRV spectrum using the low-rank MC method. In particular, we propose a new RMC method that takes into account the characteristics of frequency components in the HRV spectrum for more accurate uncertainty modelling. We evaluate the performance of our developed RMC method on five benchmark datasets with varying masking ratios. To assess its effectiveness, we compare our method with five deep regression models and the traditional MC method. The results demonstrate that our RMC method achieves the lowest estimation error while maintaining the minimal computational cost, which highlights the advantages and efficiency of our developed model for HRV spectrum estimation.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

DAC was supported by the Pandemic Sciences Institute at the University of Oxford; the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC); an NIHR Research Professorship; a Royal Academy of Engineering Research Chair; and the InnoHK Hong Kong Centre for Centre for Cerebro-cardiovascular Engineering (COCHE). TZ was supported by the Royal Academy of Engineering under the Research Fellowship scheme. This research was funded in whole, or in part, by the Wellcome Trust (217650/Z/19/Z). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. L. Lu, T. Zhu, D. Morelli, A. Creagh, Z. Liu, J. Yang, et al., Uncertainties in the analysis of heart rate variability: A systematic review, *IEEE Rev. Biomed. Eng.*, **17** (2023), 180–196. <https://doi.org/10.1109/RBME.2023.3271595>
2. S. Roy, D. P. Goswami, A. Sengupta, Geometry of the Poincaré plot can segregate the two arms of autonomic nervous system—a hypothesis, *Med. Hypotheses*, **138** (2020), 109574. <https://doi.org/10.1016/j.mehy.2020.109574>
3. L. Vazquez, J. D. Blood, J. Wu, T. M. Chaplin, R. E. Hommer, H. J. Rutherford, et al., High frequency heart-rate variability predicts adolescent depressive symptoms, particularly anhedonia, across one year, *J. Affect. Disorders*, **196** (2016), 243–247. <https://doi.org/10.1016/j.jad.2016.02.040>
4. S. Hillebrand, K. B. Gast, R. de Mutsert, C. A. Swenne, J. W. Jukema, S. Middeldorp, et al., Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: Meta-analysis and dose–response meta-regression, *Europace*, **15** (2013), 742–749. <https://doi.org/10.1093/europace/eus341>
5. N. Townsend, D. Kazakiewicz, F. L. Wright, A. Timmis, R. Huculeci, A. Torbica, et al., Epidemiology of cardiovascular disease in Europe, *Nat. Rev. Cardiol.*, **19** (2022), 133–143. <https://doi.org/10.1038/s41569-021-00607-3>
6. T. Xiang, N. Ji, D. A. Clifton, L. Lu, Y. T. Zhang, Interactive effects of HRV and P-QRS-T on the power density spectra of ECG signals, *IEEE J. Biomed. Health. Inf.*, **25** (2021), 4163–4174. <https://doi.org/10.1109/JBHI.2021.3100425>

7. A. L. Callara, L. Sebastiani, N. Vanello, E. P. Scilingo, A. Greco, Parasympathetic-sympathetic causal interactions assessed by time-varying multivariate autoregressive modeling of electrodermal activity and heart-rate-variability, *IEEE Trans. Biomed. Eng.*, **68** (2021), 3019–3028. <https://doi.org/10.1109/TBME.2021.3060867>
8. L. Rodríguez-Liñares, D. Simpson, Spectral estimation of HRV in signals with gaps, *Biomed. Signal Process. Control*, **52** (2019), 187–197. <https://doi.org/10.1016/j.bspc.2019.04.006>
9. T. W. Bae, K. K. Kwon, ECG PQRST complex detector and heart rate variability analysis using temporal characteristics of fiducial points, *Biomed. Signal Process. Control*, **66** (2021), 102291. <https://doi.org/10.1016/j.bspc.2020.102291>
10. F. Massie, S. Vits, A. Khachatryan, B. Van Pee, J. Verbraecken, J. Bergmann, Central sleep apnea detection by means of finger photoplethysmography, *IEEE J. Transl. Eng. Health Med.*, **11** (2023), 126–136. <https://doi.org/10.1109/JTEHM.2023.3236393>
11. B. E. Ajtay, S. Béres, L. Hejjel, The oscillating pulse arrival time as a physiological explanation regarding the difference between ECG-and photoplethysmogram-derived heart rate variability parameters, *Biomed. Signal Process. Control*, **79** (2023), 104033. <https://doi.org/10.1016/j.bspc.2022.104033>
12. J. Lee, M. Kim, H. K. Park, I. Y. Kim, Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths, *Sensors*, **20** (2020), 1493. <https://doi.org/10.3390/s20051493>
13. F. Xiong, D. Chen, Z. Chen, S. Dai, Cancellation of motion artifacts in ambulatory ECG signals using TD-LMS adaptive filtering techniques, *J. Visual Commun. Image Represent.*, **58** (2019), 606–618. <https://doi.org/10.1016/j.jvcir.2018.12.030>
14. P. Guyot, P. Voiriot, E. H. Djermoune, S. Papelier, C. Lessard, M. Felices, et al., R-peak detection in Holter ECG signals using non-negative matrix factorization, in *2018 Computing in Cardiology Conference (CinC)*, IEEE, (2018), 1–4. <https://doi.org/10.22489/CinC.2018.123>
15. Y. I. Jang, J. Y. Sim, J. R. Yang, N. K. Kwon, Improving heart rate variability information consistency in Doppler cardiogram using signal reconstruction system with deep learning for contact-free heartbeat monitoring, *Biomed. Signal Process. Control*, **76** (2022), 103691. <https://doi.org/10.1016/j.bspc.2022.103691>
16. G. S. Lee, M. L. Chen, G. Y. Wang, Evoked response of heart rate variability using short-duration white noise, *Auton. Neurosci.*, **155** (2010), 94–97. <https://doi.org/10.1016/j.autneu.2009.12.008>
17. E. J. Candès, B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.*, **9** (2009), 717–772. <https://doi.org/10.1007/s10208-009-9045-5>
18. X. P. Li, L. Huang, H. C. So, B. Zhao, A survey on matrix completion: Perspective of signal processing, preprint, arXiv:1901.10885.
19. L. Lu, Y. Tan, M. Klaić, M. P. Galea, F. Khan, A. Oliver, et al., Evaluating rehabilitation progress using motion features identified by machine learning, *IEEE Trans. Biomed. Eng.*, **68** (2020), 1417–1428. <https://doi.org/10.1109/TBME.2020.3036095>

20. T. Liu, M. Sun, J. Meng, Z. Wu, Y. Shen, N. Feng, Compressive sampling photoacoustic microscope system based on low rank matrix completion, *Biomed. Signal Process. Control*, **26** (2016), 58–63. <https://doi.org/10.1016/j.bspc.2015.12.008>
21. L. Xu, M. E. Chavez-Echeagaray, V. Berisha, Unsupervised EEG channel selection based on nonnegative matrix factorization, *Biomed. Signal Process. Control*, **76** (2022), 103700. <https://doi.org/10.1016/j.bspc.2022.103700>
22. E. Nasiri, K. Berahmand, M. Rostami, M. Dabiri, A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding, *Comput. Biol. Med.*, **137** (2021), 104772. <https://doi.org/10.1016/j.compbiomed.2021.104772>
23. S. Nousias, C. Tselios, D. Bitzas, A. S. Lalos, K. Moustakas, I. Chatzigiannakis, Uncertainty management for wearable IoT wristband sensors using Laplacian-based matrix completion, in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, IEEE, (2018), 1–6. <https://doi.org/10.1109/CAMAD.2018.8515001>
24. L. Lu, T. Zhu, Y. T. Zhang, D. A. Clifton, Spectrum estimation of heart rate variability using low-rank matrix completion, in *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, IEEE, (2022), 01–04. <https://doi.org/10.1109/BHI56158.2022.9926897>
25. W. S. Chen, L. Hsieh, S. Y. Yuan, High performance data compression method with pattern matching for biomedical ecg and arterial pulse waveforms, *Comput. Methods Programs Biomed.*, **74** (2004), 11–27. [https://doi.org/10.1016/S0169-2607\(03\)00022-1](https://doi.org/10.1016/S0169-2607(03)00022-1)
26. G. Shabat, A. Averbuch, Interest zone matrix approximation, *Electron. J. Linear Algebra*, **23** (2012), 678–702.
27. W. J. Zeng, H. C. So, Outlier-robust matrix completion via ℓ_p -minimization, *IEEE Trans. Signal Process.*, **66** (2017), 1125–1140. <https://doi.org/10.1109/TSP.2017.2784361>
28. T. Hastie, R. Tibshirani, M. Wainwright, Statistical learning with sparsity, *Monogr. Stat. Appl. Probab.*, **143** (2015), 8.
29. R. Mazumder, T. Hastie, R. Tibshirani, Spectral regularization algorithms for learning large incomplete matrices, *J. Mach. Learn. Res.*, **11** (2010), 2287–2322.
30. T. Hastie, R. Mazumder, J. D. Lee, R. Zadeh, Matrix completion and low-rank SVD via fast alternating least squares, *J. Mach. Learn. Res.*, **16** (2015), 3367–3402.
31. M. Daoud, P. Ravier, R. Harba, M. Jabloun, B. Yagoubi, O. Buttelli, HRV spectral estimation based on constrained Gaussian modeling in the nonstationary case, *Biomed. Signal Process. Control*, **8** (2013), 483–490. <https://doi.org/10.1016/j.bspc.2013.04.007>
32. M. A. García-González, A. Argelagós-Palau, M. Fernández-Chimeno, J. Ramos-Castro, A comparison of heartbeat detectors for the seismocardiogram, in *Computing in Cardiology 2013*, IEEE, (2013), 461–464.
33. S. Greenwald, *Improved Detection and Classification of Arrhythmias in Noise-corrupted Electrocardiograms Using Contextual Information*, Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology, 1990.

34. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, *Circulation*, **101** (2000), e215–e220. <https://doi.org/10.1161/01.CIR.101.23.e215>
35. A. Taddei, G. Distanti, M. Emdin, P. Pisani, G. Moody, C. Zeelenberg, et al., The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography, *Eur. Heart J.*, **13** (1992), 1164–1172. <https://doi.org/10.1093/oxfordjournals.eurheartj.a060332>
36. T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, J. H. Peter, The Apnea-ECG database, in *Computers in Cardiology 2000*, IEEE, (2000), 255–258. <https://doi.org/10.1109/CIC.2000.898505>
37. A. E. Johnson, J. Behar, F. Andreotti, G. D. Clifford, J. Oster, R-peak estimation using multimodal lead switching, in *Computing in Cardiology 2014*, IEEE, (2014), 281–284.
38. A. N. Vest, G. Da Poian, Q. Li, C. Liu, S. Nemati, A. J. Shah, et al., An open source benchmarked toolbox for cardiovascular waveform and interval analysis, *Physiol. Meas.*, **39** (2018), 105004. <https://doi.org/10.1088/1361-6579/aae021>
39. J. Rahul, M. Sora, L. D. Sharma, A novel and lightweight P, QRS, and T peaks detector using adaptive thresholding and template waveform, *Comput. Biol. Med.*, **132** (2021), 104307. <https://doi.org/10.1016/j.compbiomed.2021.104307>
40. W. Zong, G. Moody, D. Jiang, A robust open-source algorithm to detect onset and duration of QRS complexes, in *Computers in Cardiology 2003*, IEEE, (2003), 737–740. <https://doi.org/10.1109/CIC.2003.1291261>
41. J. Behar, J. Oster, Q. Li, G. D. Clifford, ECG signal quality during arrhythmia and its application to false alarm reduction, *IEEE Trans. Biomed. Eng.*, **60** (2013), 1660–1666. <https://doi.org/10.1109/TBME.2013.2240452>
42. C. Varon, J. Lázaro, J. Bolea, A. Hernando, J. Aguiló, E. Gil, et al., Unconstrained estimation of HRV indices after removing respiratory influences from heart rate, *IEEE J. Biomed. Health. Inf.*, **23** (2018), 2386–2397. <https://doi.org/10.1109/JBHI.2018.2884644>
43. L. Lu, Y. Tan, D. Oetomo, I. Mareels, E. Zhao, S. An, On model-guided neural networks for system identification, in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, (2019), 610–616. <https://doi.org/10.1109/SSCI44817.2019.9002703>
44. Y. Hu, H. Wang, Y. Zhang, B. Wen, Frequency prediction model combining isfr model and LSTM network, *Int. J. Electr. Power Energy Syst.*, **139** (2022), 108001. <https://doi.org/10.1016/j.ijepes.2022.108001>
45. A. Ameri, M. A. Akhaee, E. Scheme, K. Englehart, Regression convolutional neural network for improved simultaneous EMG control, *J. Neural Eng.*, **16** (2019), 036015. <https://doi.org/10.1088/1741-2552/ab0e2e>
46. A. Veloz, R. Salas, H. Allende-Cid, H. Allende, C. Moraga, Identification of lags in nonlinear autoregressive time series using a flexible fuzzy model, *Neural Process. Lett.*, **43** (2016), 641–666. <https://doi.org/10.1007/s11063-015-9438-1>

47. M. Benchekroun, B. Chevallier, V. Zalc, D. Istrate, D. Lenne, N. Vera, The impact of missing data on heart rate variability features: A comparative study of interpolation methods for ambulatory health monitoring, *IRBM*, **44** (2023), 100776. <https://doi.org/10.1016/j.irbm.2023.100776>
48. M. Alkhodari, H. F. Jelinek, S. Saleem, L. J. Hadjileontiadis, A. H. Khandoker, Revisiting left ventricular ejection fraction levels: A circadian heart rate variability-based approach, *IEEE Access*, **9** (2021), 130111–130126. <https://doi.org/10.1109/ACCESS.2021.3114029>
49. L. Lu, T. Zhu, A. H. Ribeiro, L. Clifton, E. Zhao, J. Zhou, et al., Decoding 2.3 million ECGs: Interpretable deep learning for advancing cardiovascular diagnosis and mortality risk stratification, *Eur. Heart J., Digit. Health.*, **5** (2024), 247–259. <https://doi.org/10.1093/ehjdh/ztae014>
50. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, in *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., 2017.
51. H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, et al., Informer: Beyond efficient transformer for long sequence time-series forecasting, in *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI Press, (2021), 11106–11115. <https://doi.org/10.1609/aaai.v35i12.17325>
52. H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, M. Long, Timesnet: Temporal 2D-variation modeling for general time series analysis, in *The Eleventh International Conference on Learning Representations*, 2023.
53. J. Fan, T. W. Chow, Matrix completion by least-square, low-rank, and sparse self-representations, *Pattern Recognit.*, **71** (2017), 290–305. <https://doi.org/10.1016/j.patcog.2017.05.013>
54. J. Zhao, L. Zhao, Low-rank and sparse matrices fitting algorithm for low-rank representation, *Comput. Math. Appl.*, **79** (2020), 407–425. <https://doi.org/10.1016/j.camwa.2019.07.012>
55. A. Waters, A. Sankaranarayanan, R. Baraniuk, Sparcs: Recovering low-rank and sparse matrices from compressive measurements, in *Advances in Neural Information Processing Systems 24*, Curran Associates, Inc., 2011.
56. F. Nie, Z. Li, Z. Hu, R. Wang, X. Li, Robust matrix completion with column outliers, *IEEE Trans. Cybern.*, **52** (2021), 12042–12055. <https://doi.org/10.1109/TCYB.2021.3072896>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)