2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-8791-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/BHI56158.2022.9926897

Spectrum Estimation of Heart Rate Variability Using Low-rank Matrix Completion

Lei Lu, Tingting Zhu, Yuan-Ting Zhang, and David A. Clifton

***Abstract*—Heart rate variability (HRV) is an important non-invasive parameter to assess the cardiac autonomic nervous system. In particular, spectrum matrices of HRV data have been widely used for physical and mental health monitoring. However, measurement uncertainties from data acquisition and physiological factors can easily affect the HRV spectrum and degrade outcomes of health monitoring. In this paper, we propose a new model for incomplete spectrum estimation of the HRV data based on matrix completion (MC). We show that our model performs efficiently when estimating missing entries for HRV spectra. Moreover, a refined model of matrix completion (RMC) is proposed that can be derived from correlation analysis of the HRV spectra. Two benchmark electrocardiography (ECG) datasets are retrieved and used to derive the HRV data, which are employed to evaluate the performance of our RMC method on the estimation of missing entries in the spectra. Furthermore, four different types of deep recurrent neural networks and the traditional MC method are used for a comparison study, and our RMC method obtains the least estimation error with different masking ratios. The experimental studies and comparison results demonstrate the advantages and robustness of our developed method for the estimation of incomplete HRV spectra.**

***Index Terms*—Heart rate variability, uncertainties, spectrum estimation, matrix completion.**

I. Introduction

Heart rate variability (HRV) is an important marker to assess autonomic nervous system (ANS) dynamics by measuring variations between consecutive heartbeats [1]. HRV variables have been widely used to evaluate anxiety disorder, depression, and psychotropic medication [2]. In particular, HRV indices in frequency domain describe power distributions with different frequency bands, which have been found to be reliable markers for assessing sympathetic (SNS) and parasympathetic nervous system (PNS) activities [1]. For example, the low frequency (LF) (0.04-0.15 Hz) of the HRV spectrum is generally mediated by sympathetic and parasympathetic actions; while the high frequency (HF) band (0.15-0.4 Hz) is mediated by PNS [3].

Generally, HRV data can be obtained by measuring variations of RR intervals (RRI) in electrocardiography (ECG) recordings, or it can be derived from inter-beat intervals (IBI) of photoplethysmography (PPG) signals [2]. However, measuring HRV using ECG or PPG signals is still challenging, because these data recordings are vulnerable to measurement noises and motion artifacts, which have subsequent influence on HRV data analysis [2], [4]. Many machine learning techniques have been developed to address the various uncertainties of HRV data, such as the Gaussian modelling [4], ensemble deep learning [5], and Bayesian deep learning [6]. However, deep neural networks (DNN) typically require expertise in hyperparameters tuning and need a large number of training samples.

Matrix completion (MC) is a method which enables recovery of missing entries from observed samples in an incomplete matrix [7], [8]. The MC technique uses the low-rank property of massive measured data to estimate missing values in data samples, and it has been used for uncertainty estimation in image processing, biology, and bioinformatics [8], [9]. However, the use of MC techniques for HRV spectrum analysis is largely unexplored. In this paper, we leverage the advantages of MC techniques to estimate uncertainties within the HRV spectrum data, which are modelled as missing values. In a further step, we develop a refined matrix completion (RMC) method by using representative information in the data matrix to recover missing values of HRV spectra; Two benchmark ECG datasets are used to investigate the performance and robustness of the developed RMC method for HRV spectrum estimation.

II. Method Development

Given a set of ECG recordings sampled from N subjects $\{s_p(t)\}_{p=1,2,\cdots,N}$, the HRV sequence can be obtained by measuring changes of consecutive RR intervals in the ECG recording; Then, a matrix $\mathbf{F} \in \mathbb{R}^{m\times n_{f_s}}$ can be formulated by spectra of HRV sequences, with m denoting the number of HRV sequences, and n_{f_s} indicating the number of truncated frequencies in the spectrum. Next, we introduce HRV spectrum estimation with the developed RMC technique.

A. Matrix Completion for Uncertainty Estimation

Given a matrix $\mathbf{X} \in \mathbb{R}^{n_1\times n_2}$ with some observed entries $(\mathbf{X})_{ij\in\Omega}$, with Ω as the set of observed elements in matrix $\mathbf{X}$, and $|\Omega| \ll n_1 n_2$. Assume matrix $\mathbf{X}$ has a rank of r, and $\min(n_1, n_2) \gg r$, the aim of MC technique is to recover

This work was supported in part by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), and in part by InnoHK Project Programme 3.2: Human Intelligence and AI Integration (HIAI) for the Prediction and Intervention of CVDs: Warning System at Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE). DAC is an Investigator in the Pandemic Sciences Institute, University of Oxford, Oxford, UK. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or of InnoHK – ITC.

L. Lu and T. Zhu are with the Department of Engineering Science, University of Oxford, Oxford, UK (Corresponding author: L. Lu; email: lei.lu@eng.ox.ac.uk).

Y. T. Zhang is with Hong Kong Center for Cerebrocardiovascular Health Engineering (COCHE), Hong Kong Science and Technology Park, Hong Kong SAR, China; and Department of Biomedical Engineering, City University of Hong Kong, Hong Kong SAR, China.

D. A. Clifton is with the Department of Engineering Science, University of Oxford, Oxford, UK; and Oxford Suzhou Centre for Advanced Research, Suzhou, China.

missing elements in the matrix. Formally, the MC technique is to solve the following problem [7], [8],

$$\begin{aligned} &\text{minimize} \quad \text{rank}(\mathbf{Y}) \\ &\text{subject to} \quad y_{ij} = x_{ij}, \quad \text{for } (i,j) \in \Omega \end{aligned} \tag{1}$$

As the optimisation of rank norm in Eq. (1) is an NP-hard problem, an alternative approximation can be used as,

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{Y}||_* \\ &\text{subject to} \quad y_{ij} = x_{ij}, \quad \text{for } (i,j) \in \Omega \end{aligned} \tag{2}$$

where, $||\mathbf{Y}||_*$ represents the nuclear norm of the matrix, and it is defined as the summation of singular values of the matrix.

By defining an operator $\mathcal{P}_\Omega$ to model the missing data, the optimisation in Eq. (2) can be rewritten as,

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{Y}||_* \\ &\text{subject to} \quad \mathcal{P}_\Omega(\mathbf{Y}) = \mathcal{P}_\Omega(\mathbf{X}), \end{aligned} \tag{3}$$

where, the operator $\{\mathcal{P}_\Omega(\mathbf{Y})\}_{i,j} = x_{i,j}$ if $(i,j) \in \Omega$; and 0 if $(i,j) \notin \Omega$.

Considering noises and uncertainties in data measurements, the estimation in Eq. (3) can be relaxed as,

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{Y}||_* \\ &\text{subject to} \quad ||\mathcal{P}_\Omega(\mathbf{Y}) - \mathcal{P}_\Omega(\mathbf{X})||_F \le \varepsilon, \end{aligned} \tag{4}$$

where, $\varepsilon > 0$ is a small value to control estimation error, and $||\cdot||_F$ is the Frobenius norm. An iterative solution to Eq. (4) can be obtained by the soft singular value thresholding [9].

Suppose we are interested in estimating a certain entries of the matrix, i.e., the HF or LF bands of the HRV spectrum; Then, the low rank approximation problem can be modified to approximate the matrix under Frobenius norm,

$$\begin{aligned} &\text{minimize} \quad ||\mathcal{P}_\Omega(\mathbf{Y}) - \mathcal{P}_\Omega(\mathbf{X})||_F \\ &\text{subject to} \quad f(\mathbf{Y}) \le 0, \end{aligned} \tag{5}$$

where, $f(\mathbf{Y})$ is a constraint for the estimated matrix, such as $f(\mathbf{Y}) = ||\mathbf{Y}||_* - \lambda$ is the nuclear norm constraint. The optimisation in Eq. (5) can be solved by the interest-zone matrix approximation as presented in [8].

B. Refined Matrix Completion for HRV Spectrum Estimation

Generally, HRV spectrum can be characterised and estimated by some specific models [10]. Next, we construct a refined matrix for spectrum estimation with identifying important information by the modelled signals.

Suppose we have interested entries to be estimated in the x_s spectrum, and it can be initially estimated as,

$$x_s^m(f) = \varphi(f, x_s), \tag{6}$$

where, $f \in [f_l, f_r]$ is the range of interested frequencies, x_s^m is the modelled spectrum using the operator $\varphi(\cdot)$, such as the autoregressive (AR) model or Gaussian process [10].

Then, the relevant information with respect to the x_s spectrum can be obtained as,

$$\mathcal{H}_K(x_s) \ := \ \max_{k \in [K]} \rho(x_s, x_k) \tag{7}$$

where, $\mathcal{H}_K(x_s)$ is the set of identified spectra for x_s, $[K] = 1, 2, \cdots, K$, $s \neq k$, is the set of indices for the identified spectra, and $\rho(x_s, x_k)$ is defined as,

$$\rho(x_s, x_k) \ = \ \psi(x_s^m, x_k^m) \tag{8}$$

where, $\psi(\cdot)$ calculates the relationship between paired spectra, which can be obtained by computing the correlation coefficient or distance between two spectra.

Next, a refined matrix $\mathbf{X}_{\mathcal{H}_K}$ can be obtained by the identified spectra, and the optimisation in Eq. (5) can be updated as,

$$\begin{aligned} &\text{minimize} \quad ||\mathcal{P}_{\tilde{\Omega}}(\mathbf{Y}_{\mathcal{H}_K}) - \mathcal{P}_{\tilde{\Omega}}(\mathbf{X}_{\mathcal{H}_K})||_F \\ &\text{subject to} \quad f(\mathbf{Y}_{\mathcal{H}_K}) \le 0, \end{aligned} \tag{9}$$

where, $\tilde{\Omega}$ is the set of observed entries in the refined matrix. Then, missing entries in the matrix can be estimated by solving the optimisation problem in Eq. (9) using the interest-zone matrix approximation technique [8].

III. Experimental Study and Results

HRV data usually can be derived from ECG signals, and two benchmark datasets from the PhysioNet Database [11] are used to investigate the model's performance of spectrum estimation for this study. The first ECG dataset is the MIT-BIH Normal Sinus Rhythm Database (NSRDB) [11]; and the second ECG dataset is the MIT-BIH Arrhythmia Database (ARDB) [11], [12]. The two datasets include both normal and abnormal sinus rhythms, which will be used to present a comprehensive investigation of the developed model.

A. ECG Signal Processing

The NSRDB includes 18 long-term ECG recordings collected from subjects with no significant arrhythmias, and the ECG recordings are sampled with 128 Hz; The ARDB consists of 48 ECG recordings with a sampling frequency of 360 Hz, and it is widely used for the investigation of cardiac arrhythmias.

We first identify R-peaks in the ECG recordings, calculate the RR intervals, and filter outliers of the intervals. Figure 1 shows the detected R peaks in the ECG recording, and the comparison of outlier removal. Then, the preprocessed data is divided into segments with a 5-min duration, which will be used to derive the HRV data. As the NSRDB has recordings with an approximately 24-hour sampling duration, it will produce a large matrix of data segments for the estimation. We therefore use ECG data with 1-hour duration for the analysis. Next, the HRV data is obtained by resampling the RR intervals with 4 Hz, and a 4^{th} order Butterworth filter with cutoff frequencies of 0.03 Hz and 0.9 Hz is used to preprocess the sequence [13].

B. Spectrum Estimation with the Developed RMC Method

The power spectral density (PSD) of the HRV data is obtained using Welch's algorithm with an overlap of 50% [14], and the HRV matrix can be obtained by combining all the spectra. Figure 2(a) illustrated the derived spectrum of HRV data with a particular focus on the frequency range from 0.004 Hz to 0.4 Hz. Then, we model the derived PSD spectrum using AR model for simplicity. As shown in Figure 2(a), the

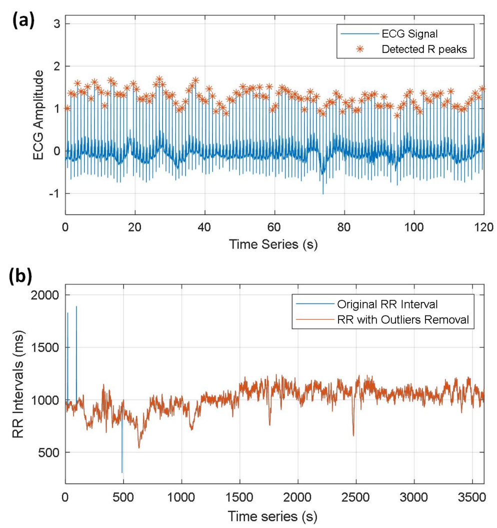

Fig. 1: ECG signal processing and RR intervals calculation. (a) R peaks detection. (b) Outlier removal from RR intervals.

AR model efficiently captures the underlying trend of the PSD spectrum in both the LF and HF parts.

To test the model's performance on estimating uncertainties in the PSD spectrum, we mask a certain number of elements in the sequence, and the remaining parts are used to derive the relevant important spectra in the matrix. Figure 2(b) shows the calculated correlation coefficients between the 20^{th} data segment and other spectra. We gradually increase the number of selected segments from 5 to 30, and find that the model obtains the optimal performance when a total number of 11 data segments are selected.

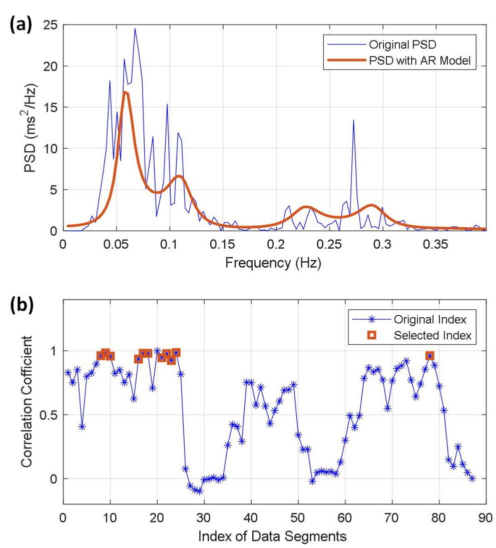

Fig. 2: HRV spectrum estimation and the identification of important data segments. (a) PSD estimation using AR model. (b) Correlation coefficients for the 20^{th} HRV spectrum.

It can be seen from Figure 2(b) that the model not only identifies neighbourhood samples for the 20^{th} data segment, it also recognises data segments with far distances as important information. Then, the refined matrix can be formulated by the identified important segments, and it is used to estimate the missing parts in the spectrum. Figure 3 shows a spectrum with 70% missing elements, which are estimated using our developed RMC method and other models. It can be seen from Figure 3(a) and (b) that both the MC and RMC methods efficiently estimate the missing values; in particular, the peak values around 0.27 Hz in the spectrum.

C. Comparison Study

Machine learning, and in particular deep learning, has demonstrated excellent performance on regression analysis of sequential data [15], [16]. As shown in Figure 3, four different types of DNN models are used to prediction missing elements in the HRV spectrum, including gated recurrent units (GRU), long short-term memory (LSTM), and hybrid models with combining LSTM and convolutional neural networks (CNN). All the recurrent neural networks have 120 hidden unites, the CNN models consist of layers with 32 and 64 kernels, and the kernel size is 3. The models are trained with a maximum epoch of 100, and an initial learning rate of 0.005.

It can be seen from Figure 3 that the four DNN models efficiently estimated missing elements in the spectrum. To present a comprehensive comparison between all the methods, we mask 50% and 70% elements of HRV spectrum for each data segment in the two datasets, i.e., NSRDB and ARDB, and then use all the models to estimate the spectra. The normalised root mean square error (NRMSE) is calculated between the estimated spectrum and original values [17]. Table I presents the performance comparison between the four DNN models, the original MC method, and our developed RMC method for the HRV spectrum estimation. The values are presented as mean $\pm$ standard deviation, and the smallest estimation error is bold-faced. It can be seen from Table I that both the MC and RMC methods outperform the DNN models on the two ECG datasets; in particular, our developed RMC method obtains the least error of HRV estimation spectrum.

TABLE I: Comparison of estimation errors (NRMSE) on the two ECG datasets with different masking ratios.

Model	NSRDB		ARDB	
	50%	70%	50%	70%
GRU	1.116 ± 0.234	1.348 ± 0.703	1.230 ± 0.444	1.117 ± 0.177
LSTM	1.142 ± 0.255	1.478 ± 0.792	1.264 ± 0.560	1.225 ± 0.352
CNN_3_LSTM	0.659 ± 0.137	0.910 ± 0.555	0.877 ± 0.494	0.859 ± 0.143
CNN_5_LSTM	0.705 ± 0.158	1.004 ± 0.628	0.964 ± 0.653	0.893 ± 0.155
MC	0.615 ± 0.805	0.743 ± 0.573	0.367 ± 0.256	0.616 ± 0.286
RMC	$\mathbf{0.285 \pm 0.278}$	$\mathbf{0.310 \pm 0.259}$	$\mathbf{0.326 \pm 0.235}$	$\mathbf{0.548 \pm 0.305}$

IV. Discussion

A matrix completion-based method was developed for HRV spectrum estimation in this paper. By leveraging the low-rank

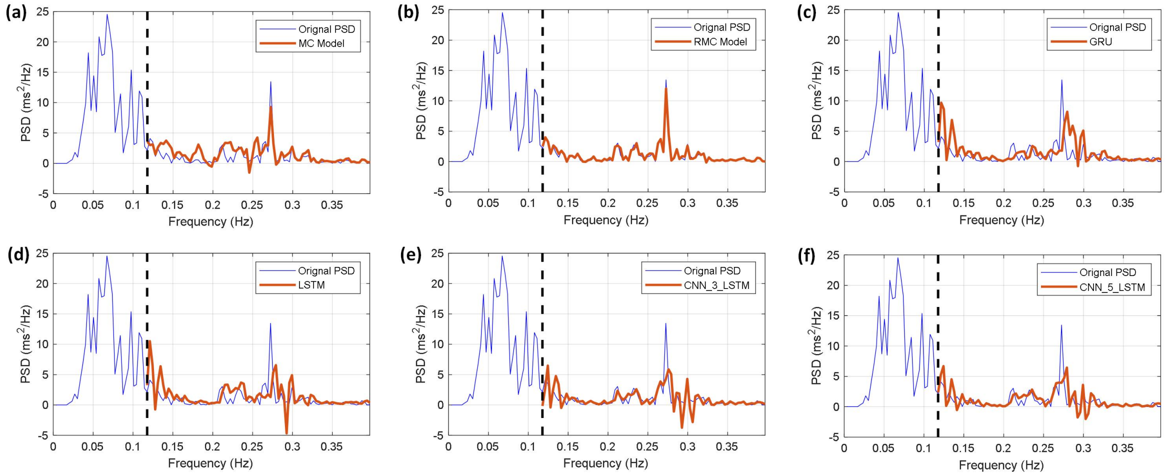

Fig. 3: Comparison study between the RMC method and other techniques. (a) Traditional MC method, (b) RMC method, (c) GRU model, (d) LSTM model, (e) LSTM and 3-layer CNN, and (f) LSTM and 5-layer CNN.

property of the HRV spectrum matrix, the developed RMC method demonstrates efficient performance on estimating missing entries of the spectrum. Comparing with four DNN models and the original MC method, we show that the developed RMC method has the best performance on the spectrum estimation, which indicates that statistical machine learning techniques may have comparable or superior performance with DNN models in some specific applications. More detailed performance of the developed RMC method and the evaluation of HRV variables will be investigated in a future study.

V. Conclusions

This study developed a new model for HRV spectrum estimation based on the low-rank MC method. A refined MC model was derived using modelled information of HRV spectra, and the developed method demonstrated efficient performance on estimating missing entries in HRV spectra, even with 70% elements masked. Comparing with four deep learning models and the traditional MC method, our developed refined MC model obtained the least NRMSE error on two benchmark ECG datasets with different masking ratios, and the results demonstrated the effectiveness and robustness of our developed model for HRV spectrum estimation.

References

[1] A. L. Callara, L. Sebastiani, N. Vanello, E. P. Scilingo, and A. Greco, "Parasympathetic-sympathetic causal interactions assessed by time-varying multivariate autoregressive modeling of electrodermal activity and heart-rate-variability," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 10, pp. 3019–3028, 2021.

[2] A. Aygun, H. Ghasemzadeh, and R. Jafari, "Robust interbeat interval and heart rate variability estimation method from various morphological features using wearable sensors," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2238–2250, 2019.

[3] M. B. Salazar, A. M. Hernández, M. Á. Mañanas, and C. C. Daza, "Modeling of heart rate variability and respiratory muscle activity in organophosphate poisoned patients," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2455–2463, 2019.

[4] O. Stegle, S. V. Fallert, D. J. C. MacKay, and S. Brage, "Gaussian process robust regression for noisy heart rate data," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2143–2151, 2008.

[5] Z. Chen, M. Wu, K. Gao, J. Wu, J. Ding, Z. Zeng, and X. Li, "A novel ensemble deep learning approach for sleep-wake detection using heart rate variability and acceleration," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 5, pp. 803–812, 2020.

[6] R. Harper and J. Southern, "A Bayesian deep learning framework for end-to-end prediction of emotion from heartbeat," *IEEE Transactions on Affective Computing*, pp. 1–1, 2020.

[7] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[8] G. Shabat and A. Averbuch, "Interest zone matrix approximation," *The Electronic Journal of Linear Algebra*, vol. 23, pp. 678–702, 2012.

[9] M. Yang, J. Wang *et al.*, "Feature and nuclear norm minimization for matrix completion," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[10] M. Daoud, P. Ravier, R. Harba, M. Jabloun, B. Yagoubi, and O. Buttelli, "HRV spectral estimation based on constrained Gaussian modeling in the nonstationary case," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 483–490, 2013.

[11] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[12] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[13] M. D. Pelaez-Coca, A. Hernando, M. T. Lozano, C. Sanchez, D. Izquierdo, and E. Gil, "Photoplethysmographic waveform and pulse rate variability analysis in hyperbaric environments," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1550–1560, 2020.

[14] C. Varon, J. Lázaro, J. Bolea, A. Hernando, J. Aguiló, E. Gil, S. Van Huffel, and R. Bailón, "Unconstrained estimation of HRV indices after removing respiratory influences from heart rate," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2386–2397, 2018.

[15] L. Lu, Y. Tan, D. Oetomo, I. Mareels, E. Zhao, and S. An, "On model-guided neural networks for system identification," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2019, pp. 610–616.

[16] L. Lu, Y. Tan, D. Oetomo, I. Mareels, and D. A. Clifton, "Weak monotonicity with trend analysis for unsupervised feature evaluation," *IEEE Transactions on Cybernetics*, 2022.

[17] S. Nousias, C. Tselios, D. Bitzas, A. S. Lalos, K. Moustakas, and 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, pp. 1–6.