Annals of Biomedical Engineering, Vol. 39, No. 2, February 2011 (© 2010) pp. 824–834
DOI: 10.1007/s10439-010-0204-2

A Robust Method to Estimate Instantaneous Heart Rate from Noisy Electrocardiogram Waveforms

ANDREI V. GRIBOK,[1,2] XIAOXIAO CHEN,[1] and JAQUES REIFMAN[1]

[1]Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702, USA; and [2]Nuclear Engineering Department of the University of Tennessee, Knoxville, TN 37996, USA

(*Received 6 July 2010; accepted 8 November 2010; published online 20 November 2010*)

Associate Editor Berj L. Bardakjian oversaw the review of this article.

Abstract—We propose a new algorithm for real-time estimation of instantaneous heart rate (HR) from noise-laden electrocardiogram (ECG) waveforms typical of unstructured, ambulatory field environments. The estimation of HR from ECG waveforms is an indirect measurement problem that requires differencing, which invariably amplifies high-frequency noise. We circumvented noise amplification by considering the estimation of HR as the solution of a weighted regularized least squares problem, which, in addition, directly provided analytically based confidence intervals (CIs) for the estimated HRs. To evaluate the performance of the proposed algorithm, we applied it to simulated data and to noise-laden ECG records that were collected during helicopter transport of trauma-injured patients to a trauma center. We compared the proposed algorithm with HR estimates produced by a widely used vital-sign travel monitor and a standard HR estimation technique, followed by postprocessing with Kalman filtering or spline smoothing. The simulation results indicated that our algorithm consistently produced more accurate HR estimates, with estimation errors as much as 67% smaller than those attained by the postprocessing methods, while the results with the field-collected data showed that the proposed algorithm produced much smoother and reliable HR estimates than those obtained by the vital-sign monitor. Moreover, the obtained CIs reflected the amount of noise in the ECG recording and could be used to statistically quantify uncertainties in the HR estimates. We conclude that the proposed method is robust to different types of noise and is particularly suitable for use in ambulatory environments where data quality is notoriously poor.

Keywords—Heart rate estimation, Least squares, Regularization, Confidence intervals.

Address correspondence to Jaques Reifman, Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702, USA. Electronic mail: jaques.reifman@us.army.mil

INTRODUCTION

The estimation of instantaneous heart rate (HR) is one of the most vexing problems in physiological measurements.[4,25] The difficulties arise from the exquisite sensitivity of such estimations to the smallest amount of noise present in electrocardiogram (ECG) waveform recordings from which the HRs are estimated.[5] Many sophisticated methods have been proposed to deal with this problem;[2,13,15] however, by and large, these are based on empirical approaches that fail to address the nature of such sensitivity.

The majority of HR estimation algorithms consist of two major steps: QRS detection and HR estimation. In this two-step approach, emphasis has been placed primarily on the first step, the detection of QRS complexes in the ECG signal. QRS detection generally involves linear and nonlinear transformations of the raw ECG to enhance the QRS complexes in the signal, and can be achieved by techniques that focus on the ECG amplitude, its first and second derivatives, or by using digital filters.[13] Unfortunately, with the exception of template-matching techniques, the most simple and easy-to-implement algorithms for detecting QRS complexes are inherently vulnerable to uncertainties in the determination of the exact location of the peak in the complex and to noise whose frequency spectra overlap with those of the complex, leading to the under- or overcounting of QRS events and, hence, to inaccurate estimations of beat occurrence times.[13] Thresholding algorithms used to detect the occurrence of heartbeats can partially alleviate the miscounting problem; however, these algorithms are based on time windows of limited length and may yield inaccurate estimates.[31]

After a time series of beat occurrence times has been determined, the next step is to estimate instantaneous

HRs by computing the difference between successive beat occurrence times. This step is generally followed by various postprocessing methods to smooth and reduce the resulting unphysiological HR variations. One class of such algorithms focuses mainly on removing large spikes in the resulting HRs caused by either ectopic beats or missed/false QRS detections,[3,26] necessarily deemphasizing noise (or error) in the exact determination of the beat occurrence times. Another class of algorithms focuses on improving the overall quality of the HR estimates by combining postprocessing methods, such as Kalman filter, with signal quality assessment techniques and additional physiological data.[11,17] Furthermore, other well-established smoothing techniques, such as spline smoothing,[28] could also be readily implemented to postprocess noisy HRs. Nevertheless, to the best of our knowledge, to date, no attempt has been made to elucidate the mathematical provenance of such unphysiological HR variations in the estimation of instantaneous HR. In this article, we provide the mathematical rationale for the sensitivity of HR estimates to measurement noise, which is ubiquitous in ambulatory environments.

In partly unstructured environments, such as in the case of transport of trauma patients to a trauma center, the recorded HRs are notoriously unreliable. For example, our automated *post hoc* analysis of vital-sign data of 898 trauma patients collected during helicopter transport to a Level I trauma center indicates that only 44% of the HRs are of sufficient quality to be used for automated decision support and closed-loop control.[31] And our analysis of vital-sign data collected from soldiers using field-wearable monitors during daily physical activity shows that such physiological data are also notoriously unreliable.[18] The percentage of reliable vital-sign data is expected to be even lower in more austere, uncontrolled environments, such as a battlefield or a mass-casualty event setting, where the challenges of obtaining accurate measurements are exacerbated by the unexpected and quickly changing physiological status of the casualties and other confounding factors. To address these challenges, our group developed physiological data qualification algorithms that automatically assess the reliability of the major vital signs.[7,18,31] While these algorithms have been shown to match the assessments made by human experts and significantly improve the accuracy of automated decision-making algorithms,[8,9,23] they have some shortcomings: they are not designed for real-time analysis, require the availability of redundant sensor measurements, and operate based on a fixed set of physiological types of measurements. Furthermore, to date, these algorithms have only been used to label physiological data as reliable or unreliable, without replacing the unreliable data with the improved estimates.

In this article, we present a new algorithm that resolves these shortcomings and provides a rigorous analysis of the HR estimation problem. We note that HRs are *indirectly* estimated and that their estimation can be cast as a solution to a least squares problem. We argue that errors in HR estimation are caused by the sensitivity of the least squares solution (or naïve differencing) to noise in the data, and further show that these errors can be eliminated by constraining the solution of the least squares problem using a well-known regularization technique with a weighting scheme,[27] i.e., weighted regularized least squares (WRLS), leading to more consistent and reliable HR estimates. The representation of HR estimation as a solution to a least squares problem also allows us to analytically compute statistical confidence intervals (CIs) on the estimated rates. We present results of simulated and field-collected data that demonstrate the performance of our algorithm on very noisy signals and compare its performance against well-established techniques, such as Kalman filtering and spline smoothing. We show that the proposed algorithm consistently provides superior performance to commonly observed types of noise in ECG records.

METHODS

QRS Detection Algorithm

This section describes the first step of the HR estimation algorithm: QRS detection. As mentioned earlier, this component of HR estimation algorithms has been well developed and studied. To process raw ECG waveforms for QRS complex identification, we modified and customized a well-known algorithm[20] to fit the needs of our perspective applications. Figure 1 shows a flow diagram of the algorithm along with the corresponding inputs and outputs of the four processing stages. In the first stage, we used a 5- to 25-Hz Butterworth band-pass filter to eliminate non-QRS-related frequencies, and in the second stage, we computed the difference between each two consecutive points of the entire data stream to amplify the sharp slopes of the QRS complex. After differencing, we squared the resulting signal to make the ECG samples positive and to amplify the high-frequency components. Finally, in the last stage, we used a low-pass filter to enhance the fiducial marks of the QRS complex and implemented a self-adaptive thresholding method to detect QRS peaks, reject noise, discriminate T-waves, and search back for missed QRS complexes if a detection was not made within a certain time interval.[20]

The *top* panel in Fig. 1 shows an example of a raw 0.5-min-long ECG segment from our trauma patient database described later in "Study Data" section. As can be seen, the raw ECG waveform was contaminated with low-frequency interference between 0.1 and 0.3 min, which could be attributed to a baseline drift due to respiration or loose electrode contact. The identification of QRS complexes in such a signal could be problematic for trained human experts and even more challenging for automated algorithms. Nevertheless, the *bottom* panel in Fig. 1 shows that the algorithm produced clearly identifiable peaks, denoted by solid circles, which were detected using our aforementioned self-adaptive thresholding method. Accordingly, the final output of the QRS detection algorithm of an ECG of length T is a time series of N monotonically increasing cumulative beat occurrence times (R_i), $0 < R_1 < R_2 < \cdots < R_i < \cdots < R_N < T$. (Notice that we selected this QRS detection algorithm due to its popularity and ease of use; however, any other QRS detection algorithm could be used in conjunction with our HR estimation algorithm described next.)

HR Estimation Algorithm

In contrast to the previous section, this section describes our unique contribution to the HR estimation problem. It addresses the second step of the HR estimation algorithm, where we estimate instantaneous HRs from a time series of cumulative beat occurrence times obtained by a QRS detection algorithm. Two important observations led to the conceptualization and development of this novel approach to HR estimation. The first was that a time series of monotonically increasing cumulative beat occurrence times needs to be differenced to obtain a sequence of R–R intervals (RRIs). Specifically, the sequence of RRIs is obtained by subtracting two successive cumulative beat occurrence times, $R_i - R_{i-1}$, i.e., $\text{RRI} = \{R_2 - R_1, R_3 - R_2, \ldots, R_N - R_{N-1}\}$. Notice that, by definition, the RRI time series represents

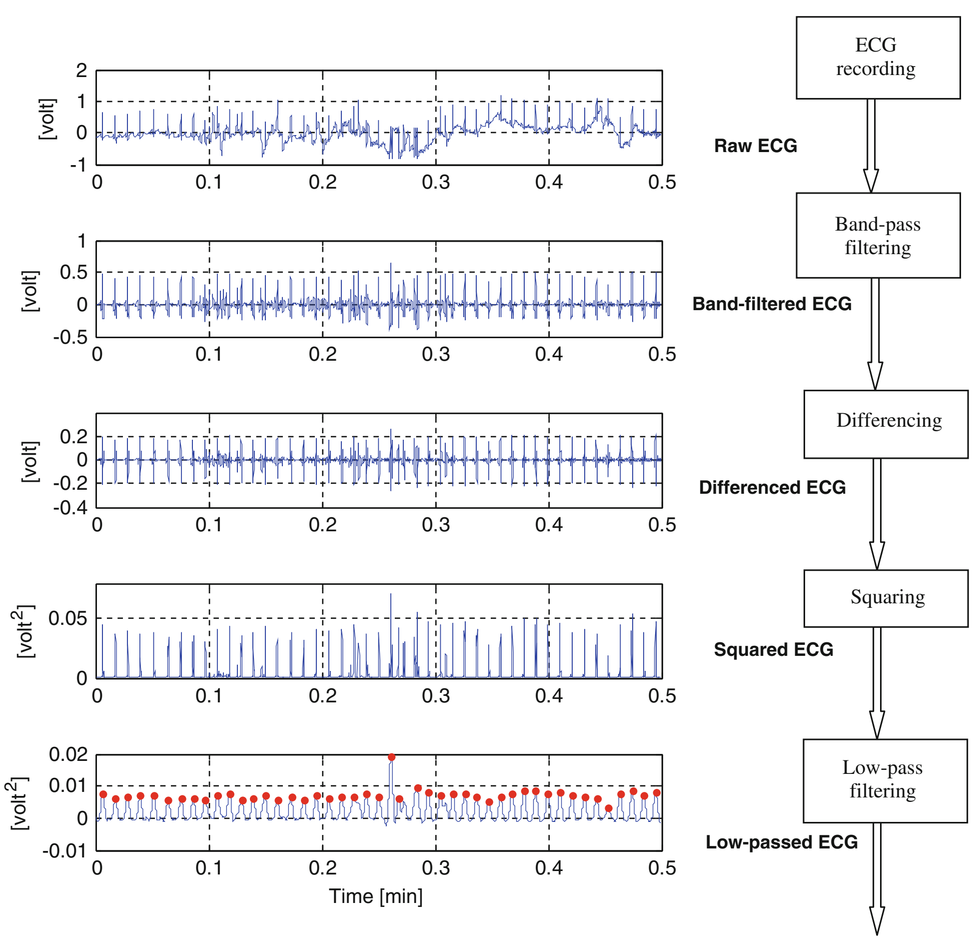

FIGURE 1. The four stages of the QRS detection algorithm along with corresponding sample inputs and outputs of each of the stages. The ECG fiducial points, or beat occurrence times, are marked with solid circles in the *bottom* panel.

the first-order difference of the cumulative beat occurrence times,

$$\mathrm{RRI} = \Delta R \quad (1)$$

where Δ denotes the first-order difference operator. Equation (1) reveals that the desired RRIs [or equivalently HRs, where $\mathrm{HR} = 60/\mathrm{RRI}$ in beats/min (bpm)] are *not a directly* measured quantity; rather, they are *indirectly* obtained by computing the difference of the directly observable cumulative beat occurrence times R. This observation that HR estimation is a differencing estimation problem has not been discussed in the literature so far and, *per se*, represents an important insight into the HR estimation problem.

The second observation was that computing the difference of an observable quantity containing some measurement noise (or errors) amplifies the noise, significantly contaminating the computed differences. This is because subtraction, or differencing, acts as a high-pass filter that amplifies high-frequency noise.[12] In our case, the right-hand side of Eq. (1) contains measurement noise because cumulative beat occurrence times cannot be determined with absolute certainty regardless of which QRS detection algorithm is used. In fact, the American National Standards Institute (ANSI) recommends a time window of 0.15 s for considering the synchronization between a QRS detection algorithm and a reference annotation.[1] In addition, and more importantly, any QRS detection algorithm eventually misses or overcounts a beat, adding more noise to the determination of the cumulative beat occurrence times. The sensitivity of such determinations to the slightest misidentification of peaks in ECG waveforms has been well documented,[5] and we argue that this sensitivity is significantly amplified in the computation of differences ΔR, which are inherent in the calculation of RRI. Therefore, we propose a different approach to estimate HR, which avoids noise amplification during differencing, thus removing the primary source of noise in RRI estimation. While all currently available techniques[4] attempt to deal with the effect of noise amplification, the proposed approach removes the very cause of such amplification.

In our approach, instead of directly taking the difference ΔR of the cumulative beat occurrence times as in Eq. (1), we reformulated the problem with R being represented as the integration of RRI

$$R = A \cdot \mathrm{RRI} + \varepsilon, \quad (2)$$

and estimated RRI as a solution to an ordinary least squares (OLS) problem[24]

$$\mathrm{RRI}_{\mathrm{OLS}} = \left[\left(A^{\mathrm{T}} \cdot A\right)^{-1} \cdot A^{\mathrm{T}}\right] \cdot R, \quad (3)$$

where R is an $N \times 1$ vector of measured cumulative beat occurrence times, RRI ($\mathrm{RRI}_{\mathrm{OLS}}$) is a $N \times 1$ vector of the corresponding "true" (OLS solution) RRI values, A denotes an $N \times N$ lower triangular integration matrix with all non-zero elements equal to one, ε represents a $N \times 1$ vector of measurement noise in R, and N is the total number of cumulative beat occurrence times. Note that Eq. (3) represents the solution of a general OLS estimation[6] (i.e., when A is not invertible) and the right-hand side is equivalent to the difference operation ΔR in Eq. (1) when A^{-1} is analytically obtainable, as in the studied case. Using Eq. (3), with RRI expressed in seconds, we computed $\mathrm{HR}_{\mathrm{OLS}} = 60/\mathrm{RRI}_{\mathrm{OLS}}$ in bpm.

Because the OLS solution is equivalent to naïve differencing, the resulting $\mathrm{RRI}_{\mathrm{OLS}}$ still contained amplified measurement noise. Thus, to obtain a smoothed estimation of RRI without noise amplification, we used the WRLS method,[22] which augments the least squares cost function $\|R - A \cdot \mathrm{RRI}\|^2$ with a weighting matrix W and a penalty term to constrain the variability of the solution. Thus, the goal is to obtain RRI that minimizes

$$\left(\|A \cdot W \cdot \mathrm{RRI}_{\mathrm{OLS}} - A \cdot W \cdot \mathrm{RRI}\|^2 + \lambda^2 \cdot \|L \cdot \mathrm{RRI}\|^2\right) \to \min, \quad (4)$$

where W denotes a diagonal $N \times N$ weighting matrix, where the elements are either zeros (represented by 10^{-5}) for spike-like outliers detected in $\mathrm{RRI}_{\mathrm{OLS}}$ (and $\mathrm{HR}_{\mathrm{OLS}}$) via an impulse rejection filter[26] or ones for non-outliers, L denotes a smoothing matrix that constrains high-frequency noise amplification in the RRI estimates and produces a smooth and consistent solution, and λ represents a positive regularization parameter, which controls the tradeoff between the fit to the data and the smoothness of the solution. A standard choice for L (and the one used here) is to use a $(N - 2) \times N$ matrix representing a second-order derivative.[22] Conversely, the regularization parameter λ is dependent on the signal-to-noise ratio in the data and can be selected by numerous methods, such as generalized cross-validation[14] or the discrepancy principle.[19] Because the signal-to-noise ratio may vary in different data sets, and to improve generalization, we customized λ for every patient. Specifically, starting with $\lambda = 0$ (i.e., no regularization), we incrementally increased it until the absolute time rate of change of the estimated HRs dropped below a specified threshold of 4.0 bpm/s, which represents the average absolute time rate of change of HRs estimated from clean ECG segments in our trauma patients database.[8]

The minimization of Eq. (4), representing the WRLS solution for RRI, can be analytically obtained[10]

$$\mathrm{RRI}_{\mathrm{WRLS}} = \left(W^{\mathrm{T}} \cdot A^{\mathrm{T}} \cdot A \cdot W + \lambda^2 \cdot L^{\mathrm{T}} \cdot L\right)^{-1} \cdot W^{\mathrm{T}} \cdot A^{\mathrm{T}} \cdot A \cdot W \cdot \mathrm{RRI}_{\mathrm{OLS}}. \quad (5)$$

Accordingly, we computed the WRLS solution for HR as $\mathrm{HR}_{\mathrm{WRLS}} = 60/\mathrm{RRI}_{\mathrm{WRLS}}$.

Confidence Interval Calculation

Finally, we computed the CI for the estimated $\mathrm{HR}_{\mathrm{WRLS}}$ through a standard formulation:[10]

$$\mathrm{CI} = \mathrm{HR}_{\mathrm{WRLS}} \pm t_{\alpha/2} \cdot \sqrt{\mathrm{Var}(\mathrm{HR}_{\mathrm{WRLS}})}, \quad (6)$$

where $t_{\alpha/2}$ denotes a percentile of a Student's t distribution with a significance level of α and $\mathrm{Var}(\mathrm{HR}_{\mathrm{WRLS}})$ represents the variance of $\mathrm{HR}_{\mathrm{WRLS}}$. The derivation of $\mathrm{Var}(\mathrm{HR}_{\mathrm{WRLS}})$ is described in the Appendix.

Study Data

Simulated Data

To validate the WRLS algorithm, we first tested and compared its performance using a pristine, i.e., a clean, noise-free, 2.5-min-long ECG segment selected from our trauma database (see "Field-Collected Data") from which we identified the cumulative beat occurrence times R and estimated (through naïve differencing) the "ground truth" instantaneous HRs. Subsequently, we simulated two types of noise in R: small time shifts and missed/added beats.[21] We simulated small time shifts, representing small errors in the detection of each peak in a QRS complex, by adding randomly selected noise sampled from a uniform distribution in the range $[-b\ b]$, with b set to 0, 0.05, 0.10, or 0.15 s. The largest b of 0.15 s was selected because the ANSI stipulates that a timing mismatch of 150 ms or less is considered insignificant.[1] For the second type of noise, we simulated both missed and added beats at random points in the beat occurrence time data record.[5] In this simulation, we substituted 0, 10, 20, or 30% of the original data record, where each time we randomly decided to substitute a record by deleting or adding a beat. In total, we simulated 16 combinations of noise conditions (including one noise-free condition). In addition, to separately investigate the performance of the algorithm on clusters of added beats and on clusters of missed beats, we simulated six more noise conditions, where in each condition we either added beats or missed beats in 10, 20, or 30% of the original data record. Note that we chose to introduce noise on R instead of on ECGs because the former allowed us to directly test the HR estimation step, as opposed to the detection of QRS complexes.

We simulated each of the 22 noise conditions (except for the noise-free one) 100 times, and computed the corresponding mean and standard deviation (SD) of the root mean squared error (RMSE) to assess the ability of four different methods (naïve differencing (i.e., OLS), Kalman filtering, spline smoothing, and WRLS) to reproduce the ground-truth HRs. More specifically, for each method, we first resampled the estimated and the ground-truth HRs to 1Hz via linear interpolation, and then calculated the RMSE of the resampled rates by taking the square root of the mean squared differences of these two HRs. We applied Wilcoxon signed-rank tests to determine whether the RMSEs were statistically significantly different from those obtained with the WRLS algorithm. The Wilcoxon signed-rank test is a nonparametric statistical hypothesis test for evaluating two related samples or repeated measurements on a single sample.[30] It can be used as an alternative to the paired Student's t-test when the population cannot be assumed to be normally distributed.

Field-Collected Data

To validate the WRLS algorithm using field-collected data, we tested the estimated instantaneous HRs using ECG recordings selected from our database containing 898 trauma patients.[8] The physiological time-series data in this database were collected from injured patients during transport via helicopter ambulance from the scene of injury to the Level I unit at the Memorial Hermann Hospital in Houston, TX. The time-series variables were measured by Propaq 206EL vital-sign monitors (Welch Allyn; Skaneateles Falls, NY), downloaded to an attached personal digital assistant, and ultimately stored in our database. The physiological data include ECG waveforms, sampled at 182 Hz, the corresponding monitor-computed HR, and other vital-sign data described elsewhere.[8] Patient attribute data, such as demographics, were also collected via chart review. Data were collected and analyzed with the approval of the local and the U.S. Army's human subjects Institutional Review Board, Fort Detrick, MD. The duration of the ECG recordings varied for different patients, with an average length of about 25 min/patient. The Propaq processed the ECG with a 0.5–40 Hz bandpass filter. Visual inspection of the ECG data revealed that they were contaminated with the various types of noise reported in the literature.[13]

RESULTS

Performance of the Algorithm on Simulated Data

We estimated HRs in the simulated data, i.e., the cumulative beat occurrence times identified from pristine ECG segments with superimposed noise (see "Study Data" section), using four methods: naïve differencing (i.e., OLS), Kalman filtering, spline smoothing, and WRLS. The Kalman filtering[17] and the spline smoothing[28] were used as postprocessing methods for HRs computed through naïve differencing. For a fair comparison, and to eliminate outliers, we used a similar weighting scheme in the Kalman filter as the one used in the WRLS algorithm. Accordingly, the Kalman filter measurement noise covariance matrix M (which was a 1×1 matrix in our case) was computed as $M = M_0 \cdot \exp\left(1/w^2 - 1\right)$, where M_0 was selected using the same criterion as the one for λ in Eq. (4) and w corresponds to the elements of the weighting matrix W, with $w = 1$ for non-outliers and $w = 10^{-5}$ for outliers. Following Li et al.[17], we set the Kalman filter 1×1 state noise covariance matrix Q to 0.1. Similarly, before we applied spline smoothing[28] to the HRs computed through naïve differencing, we identified outliers and substituted them through linear interpolation. We determined the smoothing parameter using the same criterion used to select λ.

Table 1 summarizes the results (mean (SD), except for the noise-free condition in the first row) of 16 combinations of noise conditions and shows that the WRLS algorithm significantly depressed noise and consistently yielded smaller RMSEs than each of the other three methods. The WRLS results were considerably better than those obtained with naïve differencing, in particular for larger noise levels. For example, for the highest noise level (missed/added beats of 30% and uniform noise of [−0.15 0.15] s), the RMSE for the naïve differencing was 81.79 (29.90) bpm while that for the WRLS algorithm was more than one order of magnitude lower at 8.03 (1.42) bpm. With respect to the two postprocessing methods, the WRLS algorithm outperformed the Kalman filtering and the spline smoothing in 14 out of the 16 noise conditions, where in each of these 14 conditions, the p-values of the Wilcoxon signed-rank tests were below the 0.001 level of significance. For the two exceptions where the WRLS algorithm did not have the smallest RMSE (the first two rows in the table), the largest difference between the WRLS's error and that obtained with the spline smoothing was merely 0.04 bpm.

Table 2 summarizes the simulated results [mean (SD)], where we separately added beats or missed beats in the data record, and shows that the WRLS algorithm significantly depressed noise and consistently yielded smaller RMSEs than naïve differencing. With respect to the postprocessing methods, for added beats, the WRLS algorithm outperformed both the Kalman filtering and the spline smoothing methods for each of the three noise levels, with p-values for the Wilcoxon signed-rank tests below the 0.001 level of significance. As for missed beats, the WRLS algorithm had consistently smaller RMSEs than the Kalman filtering method but marginally larger errors than the spline smoothing method, with the largest difference of 0.02 bpm being clinically insignificant.

Performance of the Algorithm on Field-Collected ECG Waveforms

To compare and contrast the performance of the WRLS algorithm on field-collected ECG, we selected

TABLE 1. Comparison of the mean (SD) root mean squared error (RMSE) of HRs estimated by the four methods for the 16 combinations of simulated noise conditions.

Missed/added beats	Uniform noise level (s)	Naïve differencing RMSE (bpm)	Kalman filtering RMSE (bpm)	Spline smoothing RMSE (bpm)	Weighted regularized least squares (WRLS) RMSE (bpm)
0%	[0.00 0.00]	N/A	0.27	0.19	0.21
	[−0.05 0.05]	4.20 (0.26)*	2.21 (0.12)*	1.01 (0.11)†	1.05 (0.12)
	[−0.10 0.10]	8.74 (0.60)*	3.57 (0.17)*	2.07 (0.17)*	1.40 (0.16)
	[−0.15 0.15]	14.16 (1.15)*	5.19 (0.33)*	4.21 (0.43)*	1.71 (0.21)
10%	[0.00 0.00]	44.72 (19.85)*	2.48 (0.47)*	1.21 (0.39)*	1.04 (0.30)
	[−0.05 0.05]	46.68 (20.26)*	3.57 (0.39)*	2.70 (0.62)*	2.09 (0.37)
	[−0.10 0.10]	49.17 (21.06)*	5.12 (0.67)*	4.97 (1.03)*	3.13 (0.52)
	[−0.15 0.15]	50.21 (21.55)*	6.67 (0.92)*	7.36 (1.24)*	4.66 (0.81)
20%	[0.00 0.00]	69.13 (24.20)*	3.24 (0.35)*	2.24 (0.60)*	1.85 (0.44)
	[−0.05 0.05]	67.11 (22.22)*	4.74 (0.62)*	4.75 (1.18)*	3.35 (0.66)
	[−0.10 0.10]	66.62 (20.90)*	7.05 (1.19)*	8.21 (1.62)*	5.14 (1.01)
	[−0.15 0.15]	68.19 (22.73)*	8.47 (1.86)*	10.39 (2.20)*	6.67 (1.20)
30%	[0.00 0.00]	80.25 (22.94)*	3.92 (0.52)*	3.37 (0.92)*	2.75 (0.65)
	[−0.05 0.05]	81.42 (24.96)*	5.91 (1.10)*	6.83 (1.78)*	4.66 (0.98)
	[−0.10 0.10]	80.81 (22.03)*	8.34 (1.93)*	10.68 (2.50)*	6.90 (1.35)
	[−0.15 0.15]	81.79 (29.90)*	10.17 (2.37)*	13.52 (3.00)*	8.03 (1.42)

SD: standard deviation; HR: heart rate; N/A: not applicable, as naïve differencing was used to compute the ground truth HRs based on noise-free condition; *p-value <0.001; †p-value <0.01.

TABLE 2. Comparison of the mean (SD) root mean squared error (RMSE) of HRs estimated by the four methods for the six simulated noise conditions with either added or missed beats.

Noise type	Noise level (%)	Naïve differencing RMSE (bpm)	Kalman filtering RMSE (bpm)	Spline smoothing RMSE (bpm)	Weighted regularized least squares (WRLS) RMSE (bpm)
Added beats	10	67.16 (33.12)*	2.80 (0.21)*	1.80 (0.38)*	1.48 (0.26)
	20	106.05 (31.95)*	4.21 (0.43)*	4.05 (0.71)*	3.06 (0.47)
	30	134.93 (35.90)*	8.82 (1.45)*	10.89 (2.03)*	7.31 (1.13)
Missed beats	10	16.18 (0.78)*	1.02 (0.33)*	0.41 (0.06)*	0.43 (0.07)
	20	24.52 (0.89)*	1.76 (0.41)*	0.61 (0.10)*	0.63 (0.11)
	30	32.05 (1.33)*	2.36 (0.50)*	0.83 (0.18)‡	0.84 (0.18)

SD: standard deviation; HR: heart rate; *p-value <0.001, ‡p-value >0.05.

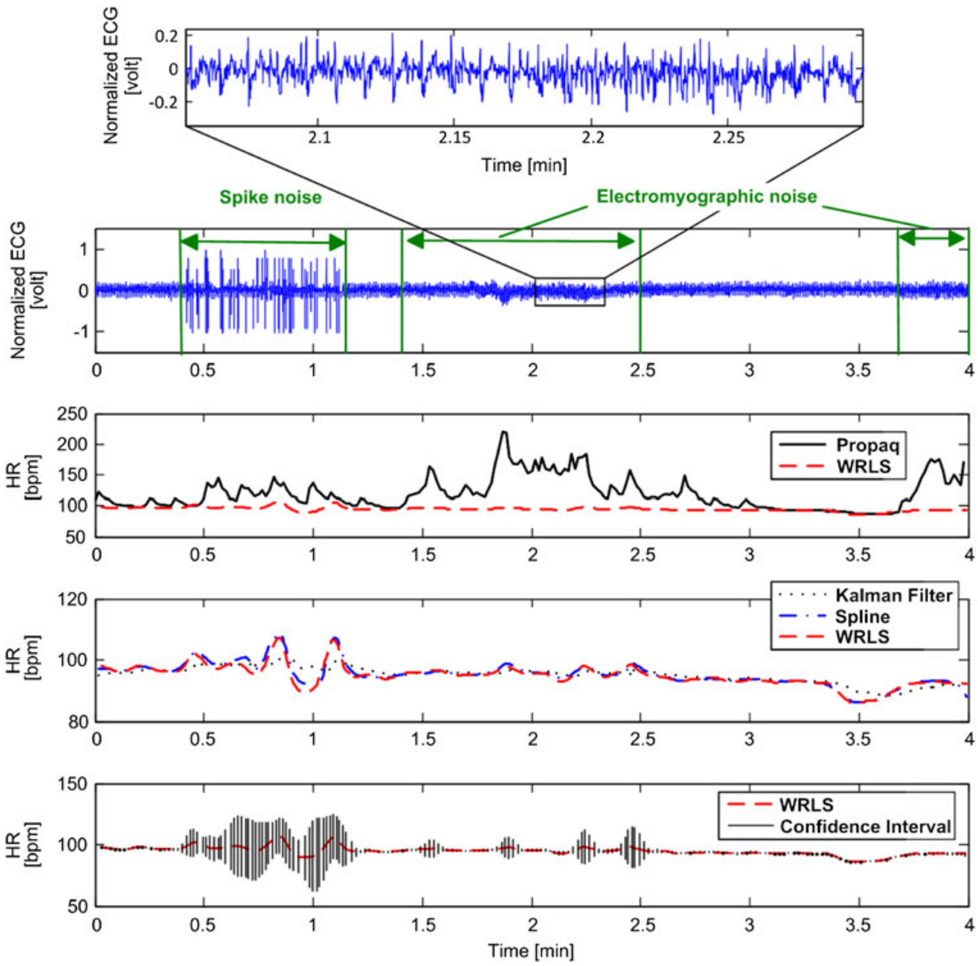

FIGURE 2. Example of noisy ECG signals collected from a patient during transport to a trauma center and the corresponding heart rates [HRs; in bpm] estimated by four distinct methods. The top panel depicts an ECG segment contaminated with spike noise between 0.4 and 1.2 min and electromyographic noise between 1.4–2.5 and 3.7–4.0 min. The second and third panels compare HRs estimated by the proposed weighted regularized least squares (WRLS) algorithm against those obtained by Propaq, Kalman filtering, and spline smoothing. The bottom panel illustrates the WRLS-estimated HRs along with the corresponding 95% confidence intervals.

two different representative patients from our trauma database whose ECGs were contaminated with commonly observed noise artifacts. Figure 2 shows 4 min of a normalized ECG record (*top* panel and inset displaying a smaller scale) for one patient and the associated HR estimates (*middle* two panels) calculated from a Propaq 206EL vital-sign monitor, Kalman filtering, spline smoothing, and the proposed WRLS algorithm. The ECG in Fig. 2 was contaminated with spike noise and electromyographic (EMG) noise. Such spike noise, consisting of high-frequency components resembling those of QRS complexes, is common in our trauma patient database, and EMG noise due to muscle contractions is ubiquitous in ECG signals.[13] Because EMG noise partially overlaps with the QRS spectra in the frequency domain, most HR monitors, including Propaq, use some kind of EMG noise filtering. However, the complete removal of this kind of contaminant is usually not possible. In Fig. 2, the largest amount of EMG noise was observed between 1.4–2.5 and 3.7–4.0 min, outside of which the ECG was relatively clean, except for the spike-noise-contaminated segment between 0.4 and 1.2 min. The ECG contamination was reflected in the Propaq-calculated HRs (*second* panel), which showed very large variations between 1.4–2.5 and 3.7–4.0 min, changing from 100 to 200 bpm within a few seconds, and smaller but marked variations between 0.5 and 1.2 min. Such rapid HR excursions are usually unphysiological and indicate the inability of the algorithm to provide continuously smooth point estimates. It is not clear why the Propaq-calculated HRs also showed marked variations between 2.5 and 2.8 min, when the ECG recording was relatively clean. In contrast, the WRLS algorithm provided smooth and consistent estimates for the whole recording, indicating its ability to handle both types of noise. Note that during the periods of clean ECG segments, the two algorithms yielded relatively similar HR estimates. The third panel in Fig. 2 compares HR estimates obtained by the WRLS algorithm with those from the Kalman filtering and the spline smoothing methods. While overall there was good agreement among the three HR estimates, the Kalman filter tended to fit a flat line no matter whether the "intrinsic" HR was changing or not, and, in comparison with our algorithm, the spline smoothing exhibited small overestimations, for example, at 0.5, 0.7, and 0.9 min.

Figure 3 shows similar comparisons for three additional types of commonly found noise in ECG recordings: motion artifacts, electrode contact noise, and instrumentation saturation noise. Motion artifacts are transient (but not step) baseline variations caused by changes in the electrode-skin impedance due to electrode motion[13] and are commonly observed during patient transport. Electrode contact noise is generally caused by a loose contact between an electrode and a patient's skin, and saturation noise is usually caused by

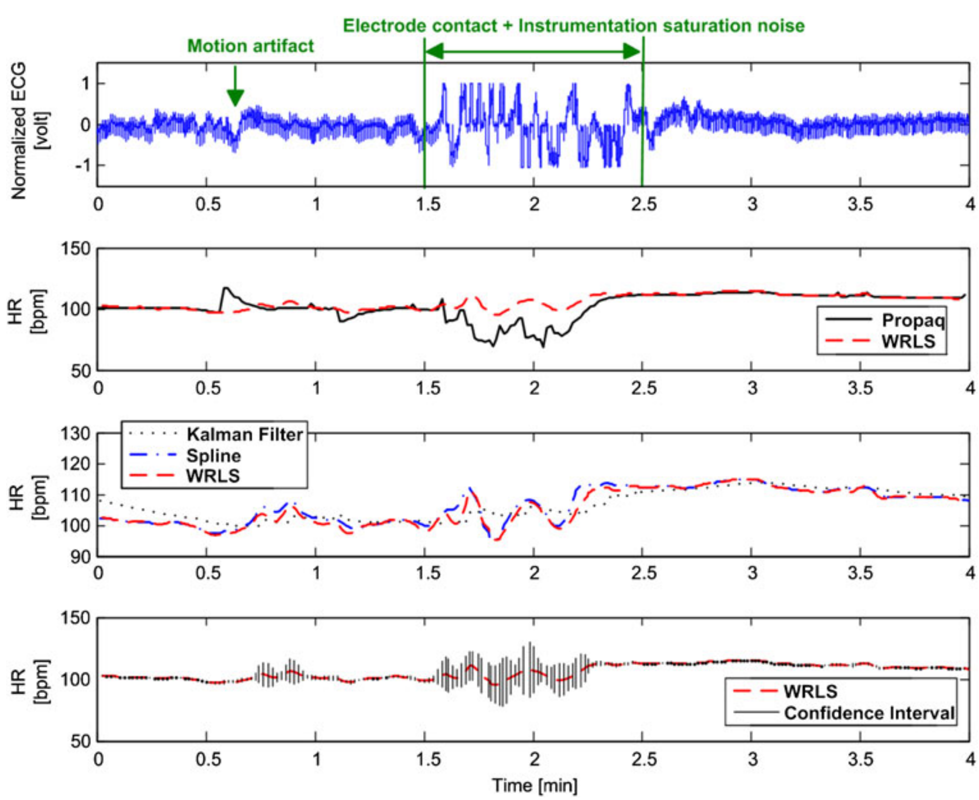

FIGURE 3. Example of noisy ECG signals collected from a patient during transport to a trauma center and the corresponding heart rates [HRs; in bpm] estimated by four distinct methods. The top panel depicts an ECG segment contaminated with electrode contact and instrumentation saturation noise between 1.5 and 2.5 min and motion artifact between 0.6 and 0.7 min. The second and third panels compare HRs estimated by the proposed weighted regularized least squares (WRLS) algorithm against those obtained by Propaq, Kalman filtering, and spline smoothing. The bottom panel illustrates the WRLS-estimated HRs along with the corresponding 95% confidence intervals.

amplification of artifacts exceeding the data range.[13] The latter two types of noise significantly affected the ECG between 1.5 and 2.5 min, causing Propaq's estimates to change from around 100 bpm to as low as 70 bpm in just a few seconds. Between 0.6 and 0.7 min of the recording, the Propaq-estimated HRs displayed a ''bump'' above the 100 bpm rate, reflecting the effects of the ECG motion artifact in its estimates. In contrast, the WRLS algorithm produced smooth and consistent estimates. Notice that the ECG was relatively clean before 0.5 min and after 2.5 min of the recording, during which periods the Propaq's and the WRLS's estimates practically coincided with each other, indicating that during the clean ECG segments both algorithms worked correctly. When compared with the WRLS algorithm, the Kalman filtering and the spline smoothing-estimated HRs showed a similar pattern (*third* panel), although the differences were more pronounced than those in Fig. 2. In particular, during the clean ECG recording before 0.5 min the Kalman filtered HRs displayed a ~5% deviation from those obtained with the other two methods, and at ~1.7 min the spline smoothing overestimated the WRLS by ~3% and the Kalman filter by ~8%.

Confidence Intervals

The bottom panels in Figs. 2 and 3 illustrate the WRLS estimated HRs and their corresponding 95% CIs. The width of the CIs should reflect the uncertainty in the HR estimates, with smaller width indicating lower uncertainty, and vice versa. Accordingly, ECG segments containing larger amounts of noise artifacts should yield wider CIs. For example, the estimated CIs in the *bottom* panel in Fig. 2 were, as expected, large between 0.4–1.2 and 1.4–2.5 min, indicating uncertainties in the HR estimates associated with spike and EMG noises. Outside of these regions, the CIs were narrow, reflecting small uncertainties in the HR estimates because of the good quality of the ECG signal. Figure 3 shows a similar pattern; the CIs were wider for noisy regions in the ECG and narrower for good-quality ones.

DISCUSSIONS AND CONCLUSIONS

This article provides a formal mathematical formulation to compute instantaneous HR in real time from noise-laden ECG signals typical of field environments. We noted that HR estimation is an indirect measurement problem that requires the computation of differences between successive beat occurrence times, which invariably amplifies noise (i.e., the uncertainties in the determination of the beat occurrence times). We circumvented noise amplification by casting the estimation of HRs, from beat occurrence times, as the solution of a weighted regularized least squares problem, thereby eliminating the need for *post hoc* smoothing of HR signals corrupted with amplified noise. Hence, the proposed formulation is fundamentally different from existing techniques[3,4,11,17,26] in that while these solutions attempt to smooth noise-amplified HRs with poor signal-to-noise ratios, the proposed algorithm avoids noise amplification altogether.

A unique advantage of the proposed WRLS algorithm is that the resulting solution is optimal in the least squares sense and, hence, it yields a closer estimate to the true (unknown) solution than the standard least squares approach.[16] Therefore, in principle, the method of regularized least squares can improve generalization with respect to different data sets. However, regularization is inherently biased, as it trades off a solution with a larger variance and a smaller bias for one with a smaller variance but a slightly larger bias, where the bias is controlled by the selection of the smoothing matrix L and the regularization parameter λ. Hence, in practice, regularization can also limit generalization. To avoid this potential problem, we set L as a second-order derivative, which is the most general choice for imposing smoothness in the estimates of R–R intervals,[22] and customized λ for every patient to allow the estimated HRs to vary as much as possible within the ''normal'' range (i.e., ±4.0 bpm/s). We also introduced a weighting scheme to eliminate the effects of outliers, which could otherwise significantly bias the results even with the use of regularization.

Another advantage of the proposed algorithm is the ability to infer a measure of uncertainty of the estimated HRs in the form of CIs. This is achieved by casting HR estimation as a solution to a least squares problem, directly yielding analytic, statistically based CIs for the estimated rates that account for noise in the ECG, a feature that is lacking in any prevailing algorithm. This is in contrast with our previous attempts to qualify physiological measurements based on empirical rules with no statistical justification.[7,18,31] Moreover, our method is generic, requiring only the availability of a single waveform signal. This independence of other physiological measurements promotes usability across different waveform signals, e.g., respiratory waveforms, and portability across vital-sign-monitoring devices and decision-support systems.

To validate our WRLS algorithm, we assessed its performance with both simulated and field-collected data against established HR estimation methods. For the simulated data, the WRLS algorithm yielded estimation errors (i.e., RMSEs) that were more than one order of magnitude smaller than those obtained through naïve differencing (i.e., OLS). When we postprocessed these HRs with two different methods (Kalman filter and spline) to smooth the estimated rates, we obtained significantly improved results. However, the attained RMSEs were still higher by as much as 67% than those obtained with the WRLS algorithm, which consistently yielded superior results. This supports our assertion that rather than attempting to smooth noise-amplified HRs with poor signal-to-noise ratios, one should avoid noise amplification altogether. When we applied the WRLS algorithm to the 898 patients in our trauma database, it reliably produced improved HR estimates that showed observable differences when compared with those provided by the monitoring device (i.e., Propaq) and the two postprocessing algorithms. In the presence of noise frequently observed in field-collected ECG, the Propaq often yielded HRs that exhibited very rapid excursions, suggesting that the estimated HRs were unphysiological. In contrast, the Kalman filter failed to express any variability in its smoothing, as it systematically fitted a flat line regardless the amount of noise in the naïve differencing HR estimates. These estimates could be improved, however, by jointly employing signal quality indices and using additional vital-sign information.[17] Conversely, spline smoothing frequently exhibited larger variations in the HR estimates, which we attribute to noise amplification. In addition, the CIs estimated by the WRLS algorithm correctly reflected the uncertainties in the HR estimates with respect to noise contamination in the ECG recordings.

One potential limitation of the proposed method is the inability to track true, sudden large changes in HR. Because we imposed a constraint of 4.0 bpm/s on the maximum time rate of change of "normal" HRs, the proposed method could compromise the accuracy of the instantaneous HR estimates if the true HR changes were larger than 4.0 bpm/s. In this case, the method would correctly detect the onset of elevations or reductions in HR without time-lags; however, it would produce a delay in reaching the actual elevated or reduced HR value. Such a delay would be proportional to the difference between the true absolute time rate of change of the HR and the imposed constraint of 4.0 bpm/s. Naturally, this limitation could be alleviated by arbitrarily increasing the constraint; however, setting the constraint too high would also diminish the accuracy of the algorithm by including excessive noise in the estimated HRs.

As biosensors become ubiquitous in everyday life, it is important that we continue improving algorithms for the real-time estimation of vital signs. For both civilian and military applications, it is particularly important to infer reliable values for HRs—arguably one of the most-used vital signs—collected from austere, unstructured environments, such as a battlefield, during the transport of trauma patients, in-home care of elderly patients, and in the monitoring of active individuals during physical activity, where the original ECG data are prone to be contaminated with noise artifacts. The study proposed here is a step in this direction, as it allows for more consistent estimation of HRs and of other ECG features, such as HR variability, which are highly sensitive to ECG noise.

APPENDIX

Because $\mathrm{HR}_{\mathrm{WRLS}}$ is reciprocal to $\mathrm{RRI}_{\mathrm{WRLS}}$, we approximated $\mathrm{Var}(\mathrm{HR}_{\mathrm{WRLS}})$ using the mean and variance of $\mathrm{RRI}_{\mathrm{WRLS}}$ through a Taylor expansion, i.e.,

$$\mathrm{Var}(\mathrm{HR}_{\mathrm{WRLS}}) \approx \left(\frac{60}{[\mathrm{mean}(\mathrm{RRI}_{\mathrm{WRLS}})]^2}\right)^2 \cdot \mathrm{Var}(\mathrm{RRI}_{\mathrm{WRLS}}), \quad \text{(A1)}$$

where $\mathrm{Var}(\mathrm{RRI}_{\mathrm{WRLS}})$ was estimated as the diagonal of the covariance of $\mathrm{RRI}_{\mathrm{WRLS}}$, $\mathrm{Cov}(\mathrm{RRI}_{\mathrm{WRLS}})$, which was calculated from Eqs. (2), (3), and (5) as

$$\mathrm{Cov}(\mathrm{RRI}_{\mathrm{WRLS}}) = C \cdot B \cdot \mathrm{Cov}(\varepsilon) \cdot B^{\mathrm{T}} \cdot C^{\mathrm{T}}, \quad \text{(A2)}$$

where $B = \left(A^{\mathrm{T}} \cdot A\right)^{-1} \cdot A^{\mathrm{T}}$, $C = \left(W^{\mathrm{T}} \cdot A^{\mathrm{T}} \cdot A \cdot W + \lambda^2 \cdot L^{\mathrm{T}} \cdot L\right)^{-1} \cdot W^{\mathrm{T}} \cdot A^{\mathrm{T}} \cdot A \cdot W$, and $\mathrm{Cov}(\varepsilon)$ denotes the covariance of the measurement noise, which was estimated as a diagonal matrix with elements equal to half of the square of the residual between $\mathrm{RRI}_{\mathrm{OLS}}$ and $\mathrm{RRI}_{\mathrm{WRLS}}$.[29]

ACKNOWLEDGMENTS

We thank Y. Lu, S. Rajaraman, and C. Meng for providing stimulating feedback. This study was supported by a War Supplemental competitive grant managed by the Combat Casualty Care Directorate of the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD.

DISCLAIMER

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

REFERENCES

[1]ANSI/AAMIEC57. Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. American National Standards Institute, 1998.

[2]Arzeno, N. M., Z. D. Deng, and C. S. Poon. Analysis of first-derivative based QRS detection algorithms. *IEEE Trans. Biomed. Eng*. 55(2):478–484, 2008.

[3]Bassil, H. E., J. H. Dripps, and P. M. Grant. Detection and correction of outliers in foetal heart rate time series. *Electron. Lett*. 28(4):382–383, 1992.

[4]Berntson, G. G., T. J. Bigger, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone, *et al*. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34(6):623–648, 1997.

[5]Berntson, G. G., and J. R. Stowell. ECG artifacts and heart period variability: don't miss a beat! *Psychophysiology* 35(1):127–132, 1998.

[6]Björck, A. Numerical Methods for Least Squares Problems. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1996.

[7]Chen, L., T. McKenna, A. T. Reisner, and J. Reifman. Algorithms to qualify respiratory data collected during the transport of trauma patients. *Physiol. Meas*. 27(9):797–816, 2006.

[8]Chen, L., T. M. McKenna, A. T. Reisner, A. Gribok, and J. Reifman. Decision tool for the early diagnosis of trauma patient hypovolemia. *J. Biomed. Inform*. 41(3):469–478, 2008.

[9]Chen, L., A. T. Reisner, A. Gribok, T. M. McKenna, and J. Reifman. Can we improve the clinical utility of respiratory rate as a monitored vital sign? *Shock* 31(6):574–580, 2009.

[10]Draper, N. R., and H. Smith. Applied Regression Analysis. New York, NY: Wiley, 1966.

[11]Ebrahim, M. H., J. M. Feldman, and I. Bar-Kana. A robust sensor fusion method for heart rate estimation. *J. Clin. Monit*. 13(6):385–393, 1997.

[12]Engl, H. W., M. Hanke, and A. Neubauer. Regularization of Inverse Problems. New York, NY: Springer, 2000.

[13]Friesen, G. M., T. C. Jannett, M. A. Jadallah, S. L. Yate, S. R. Quint, and H. T. Nagle. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Trans. Biomed. Eng*. 37(1):85–98, 1990.

[14]Golub, G. H., M. Heath, and G. Wahba. Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics* 21(2):215–223, 1979.

[15]Hamilton, P. S., and W. J. Tompkins. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Trans. Biomed. Eng*. 33(12): 1157–1165, 1986.

[16]Hoerl, A. E., and R. W. Kennard. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12(1):55–67, 1970.

[17]Li, Q., R. G. Mark, and G. D. Clifford. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiol. Meas*. 29:15–32, 2008.

[18]Liu, J., T. M. McKenna, A. Gribok, B. A. Beidleman, W. J. Tharion, and J. Reifman. A fuzzy logic algorithm to assign confidence levels to heart and respiratory rate time series. *Physiol. Meas*. 29(1):81–94, 2008.

[19]Morozov, V. A. Regularization Methods for Ill-Posed Problems. Florida: CRC Press, 1993.

[20]Pan, J., and W. J. Tompkins. A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng*. 32(3):230–236, 1985.

[21]Pu, Y., and R. P. Patterson. Comparison of R-wave detection errors of four wireless heart rate belts in the presence of noise. *Physiol. Meas*. 24:913–924, 2003.

[22]Ramsay, J. O., and B. W. Silverman. Functional Data Analysis. New York, NY: Springer, 2005.

[23]Reisner, A. T., L. Chen, T. M. McKenna, and J. Reifman. Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. *J. Trauma* 65(4):915–923, 2008.

[24]Tarantola, A. Inverse Problem Theory. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2005.

[25]Task Force of the European Society of Cardiology; North American Society of Pacing Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation* 93:1043–1065, 1996.

[26]Thuraisingham, R. A. Preprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals. *Comput. Methods Programs Biomed*. 83(1):78–82, 2006.

[27]Tikhonov, A. N., and V. Y. Arsenin. Solution of Ill-Posed Problems. Washington: Winston & Sons, 1977.

[28]Wahba, G. Spline Models for Observational Data. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1990.

[29]White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4):817–838, 1980.

[30]Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull*. 1(6):80–83, 1945.

[31]Yu, C., Z. Liu, T. McKenna, A. T. Reisner, and J. Reifman. A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. *J. Am. Med. Inform. Assoc*. 13(3):309–320, 2006.