

# Characterising an ECG signal using statistical modelling: a feasibility study

T. Bodisco<sup>1</sup>      J. D’Netto<sup>2</sup>      N. Kelson<sup>3</sup>      J. Banks<sup>4</sup>  
R. Hayward<sup>5</sup>      T. Parker<sup>6</sup>

(Received 18 December 2013; revised 25 February 2014)

## Abstract

For clinical use, in electrocardiogram (ECG) signal analysis it is important to detect not only the centre of the P wave, the QRS complex and the T wave, but also the time intervals, such as the ST segment. Much research focused entirely on QRS complex detection, via methods such as wavelet transforms, spline fitting and neural networks. However, drawbacks include the false classification of a severe noise spike as a QRS complex, possibly requiring manual editing, or the omission of information contained in other regions of the ECG signal. While some attempts were made to develop algorithms to detect additional signal characteristics, such as P and T waves, the reported success rates are subject to change from person-to-person and beat-to-beat. To

---

<http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/7818> gives this article, © Austral. Mathematical Soc. 2014. Published March 26, 2014, as part of the Proceedings of the 11th Biennial Engineering Mathematics and Applications Conference. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

address this variability we propose the use of Markov-chain Monte Carlo statistical modelling to extract the key features of an ECG signal and we report on a feasibility study to investigate the utility of the approach. The modelling approach is examined with reference to a realistic computer generated ECG signal, where details such as wave morphology and noise levels are variable.

## Contents

<b>1 Introduction</b>	<b>C33</b>
<b>2 The conceptual model</b>	<b>C36</b>
<b>3 Algorithm implementation</b>	<b>C38</b>
<b>4 Results</b>	<b>C40</b>
<b>5 Discussion and conclusions</b>	<b>C40</b>
<b>References</b>	<b>C43</b>

## 1 Introduction

The electrocardiogram (ECG) signal (generated by measuring the trans-thoracic electrical activity of the heart) provides a non-invasive means of examining the physiological function of the heart [3]. Clinically relevant features of the ECG signal include the relative time between different features, the amplitude of different waves and the shape of some features [3]. Each segment of the ECG signal is split into categorised components: the P wave; the Q, R and S waves, grouped together as the QRS complex; and the T wave; all shown in Figure 1.

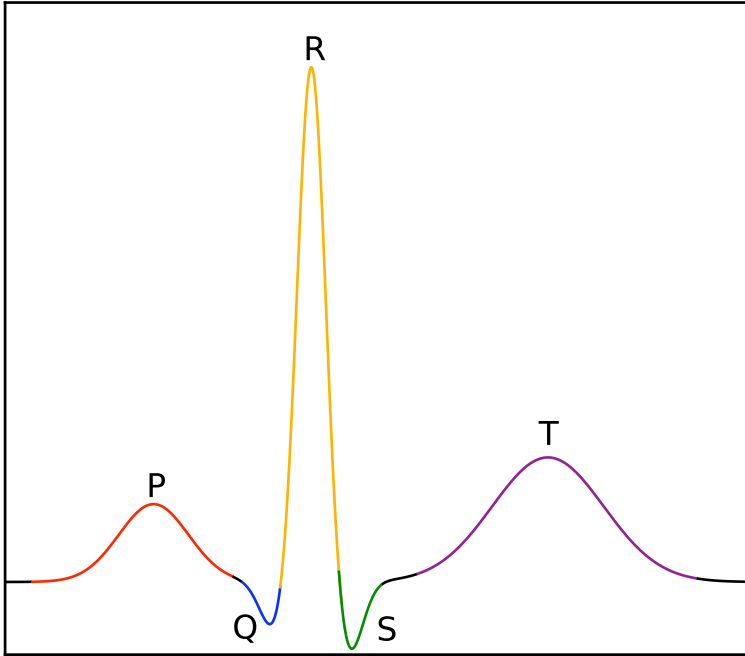


Figure 1: An idealised representation of an ECG signal showing the major components.

This article tests the feasibility of extracting clinically relevant information from an ECG signal, that is typical of a healthy adult, using Markov-chain Monte Carlo (MCMC). Specifically, by fitting a statistical model to the ECG signal, this study aims to obtain the turning points and corresponding amplitudes of the significant features. The results identify the amplitude, amplitude time, onset time and offset time of each wave present in a typical healthy adult ECG signal.

Automated ECG analysis algorithms have existed for a long time, and they apply different methods with different strengths and weaknesses. For example, the Pan–Tompkins algorithm [9] is a first derivative threshold QRS detector with adaptive thresholds and time restrictions. Part of the algorithm attempts

to remove T waves falsely identified as QRS complexes. In the original study [9], the worst case result reported was 12.54% failure; however, the average was 0.675% failure and the best was 0% failure. Despite its age, it is still used in current research [4] and is often used in commercial applications. Recently, Lin et al. [4] used the Pan–Tompkins algorithm for QRS detection and then searched specific time intervals with reference to the previously located QRS complexes for the detection of P and T waves [4]. Neither of these algorithms are capable of detecting all the key features of a typical ECG signal.

The quadratic spline wavelet approach taken by Niknazar et al. [8] performs extremely well for QRS complex detection, with a reported average 0.16% error for the subset of the MIT–BIH arrhythmia database records chosen [7]. Niknazar et al. [8] proposed a method for the detection of the P and T waves, and provided an illustrated example of the method, but did not show any results for isolating P and T waves. However, the authors did admit to deficient capabilities in the detection of P and T waves in “severe noisy cases” [8]. Similarly, Zong et al. [13, 12] developed an algorithm for determining the onset and offset of the QRS complex, which were then fed into their algorithm to detect the end of the T wave. Details on the algorithm used for the QRS complex detection are referenced in an earlier paper [12]. This QRS detection algorithm has a reported failure rate of 0.35%. This approach was tested on the PTB diagnostic ECG database [7] which has 549 records. A key strength of this approach is that it is able to detect the QT interval—which is clinically significant [3].

Work by Edla et al. [1] used an MCMC method for ECG beat classification. Beat classification is when, given the location of all of the features in the ECG signal, the condition of the subject is identified. In contrast, feature classification is when, given the raw waveform output from an ECG lead, the location of the features are identified.

The method of feature detection proposed here differs from previous work in that well defined models of the ECG signal are fitted to the data, which allows the features of a typical healthy adult ECG signal to be identified. We present

a feasibility study using computer simulated data [5]. Data presentation is manipulated by controlling model parameters (three per P, Q, R, S and T wave that control amplitude, time offset and wave width). The simulator also has heart rate, baseline and noise parameters. This ECG signal synthesiser allows the production of multiple morphological characteristics for different beats inside an ECG signal, and is therefore appropriate for a feasibility study and to explore the strengths and weaknesses of our model.

## 2 The conceptual model

In our approach, an empirical model was developed to fit data. This model consists of a set of piece-wise continuous periodic functions that are controlled to begin at a peak and finish at a trough (thereby ensuring that the overall function is continuous). Given how well QRS complex detectors are established, the data window for analysis is between the peaks of the QRS complexes of the beats immediately before and after the one of interest. Therefore, the model starts and finishes at the peak of a QRS complex. To control for noise and minor abnormalities, the data is modelled as a Normal distribution with a time  $t$  varying mean  $\mu$  defined in Table 1, and some precision factor  $\tau$ ,

$$s(t) = y \sim N[\mu(t) + DC, \tau].$$

In our model:

- $s$  is the ECG signal and  $y$  is the model representation of the ECG signal;
- $t$  is time and  $T$  is the time related to the top of the QRS complex in the beat following that of interest;
- $DC$  is the DC voltage offset;
- $\delta_{1,\dots,17}$  are the turning points in an ECG heart beat;
- $\alpha_{1,\dots,12}$  are related to the amplitudes of the cosine waves;

Table 1: Form of mean  $\mu(t)$  during different time intervals.

mean $\mu(t)$	time interval
$\alpha_1 \cos\left(\frac{\pi}{\delta_1}t\right) + \beta$	$[0, \delta_1)$
$\alpha_2 \left[1 - \cos\left(\frac{\pi}{\delta_2 - \delta_1}t(t - \delta_1)\right)\right] + \beta - \alpha_1$	$[\delta_1, \delta_2)$
$\alpha_3 \left[1 - \cos\left(\frac{\pi}{\delta_3 - \delta_2}(t - \delta_2)\right)\right] + \beta - \alpha_1 + 2\alpha_2$	$[\delta_2, \delta_3)$
$\frac{\beta - \alpha_1 + 2\alpha_2 + 2\alpha_3}{2} \left[1 + \cos\left(\frac{\pi}{\delta_4 - \delta_3}(t - \delta_3)\right)\right]$	$[\delta_3, \delta_4)$
0	$[\delta_4, \delta_5)$
$\alpha_4 \left[1 - \cos\left(\frac{\pi}{\delta_6 - \delta_5}(t - \delta_5)\right)\right]$	$[\delta_5, \delta_6)$
$\alpha_4 \left[1 + \cos\left(\frac{\pi}{\delta_7 - \delta_6}(t - \delta_6)\right)\right]$	$[\delta_6, \delta_7)$
$\alpha_5 \left[\cos\left(\frac{\pi}{\delta_8 - \delta_7}(t - \delta_7)\right) - 1\right]$	$[\delta_7, \delta_8)$
$\alpha_6 \left[1 - \cos\left(\frac{\pi}{\delta_9 - \delta_8}(t - \delta_8)\right)\right] - 2\alpha_5$	$[\delta_8, \delta_9)$
$\alpha_7 \left[\cos\left(\frac{\pi}{\delta_{10} - \delta_9}(t - \delta_9)\right) - 1\right] + 2\alpha_6 - 2\alpha_5$	$[\delta_9, \delta_{10})$
$\alpha_8 \left[1 - \cos\left(\frac{\pi}{\delta_{11} - \delta_{10}}(t - \delta_{10})\right)\right] - 2\alpha_7 + 2\alpha_6 - 2\alpha_5$	$[\delta_{10}, \delta_{11})$
$\alpha_9 \left[1 - \cos\left(\frac{\pi}{\delta_{12} - \delta_{11}}(t - \delta_{11})\right)\right] + 2\alpha_8 - \alpha_7 + 2\alpha_6 - 2\alpha_5$	$[\delta_{11}, \delta_{12})$
$(\alpha_9 + \alpha_8 - \alpha_7 + \alpha_6 - \alpha_5) \left[1 + \cos\left(\frac{\pi}{\delta_{13} - \delta_{12}}(t - \delta_{12})\right)\right]$	$[\delta_{12}, \delta_{13})$
0	$[\delta_{13}, \delta_{14})$
$\alpha_{10} \left[1 - \cos\left(\frac{\pi}{\delta_{15} - \delta_{14}}(t - \delta_{14})\right)\right]$	$[\delta_{14}, \delta_{15})$
$\alpha_{10} \left[1 + \cos\left(\frac{\pi}{\delta_{16} - \delta_{15}}(t - \delta_{15})\right)\right]$	$[\delta_{15}, \delta_{16})$
$\alpha_{11} \left[\cos\left(\frac{\pi}{\delta_{17} - \delta_{16}}(t - \delta_{16})\right) - 1\right]$	$[\delta_{16}, \delta_{17})$
$\alpha_{12} \left[1 - \cos\left(\frac{\pi}{T - \delta_{17}}(t - \delta_{17})\right)\right] - 2\alpha_{11}$	$[\delta_{17}, T)$

- $\beta$  is an offset to compensate for starting at the top of a QRS complex.

### 3 Algorithm implementation

The model parameter posterior distributions are resolved within a pseudo-Bayesian framework using the Metropolis–Hastings algorithm—the end point of this research is to develop a personal monitor for continuous cardiac monitoring and therefore the algorithm needs to be computationally optimised. An example of a computational optimisation is that in place of uninformative priors (such as a Normal prior with a very low precision), a computationally faster uniform prior between zero and infinity is utilised—all model parameters are constrained to be greater than zero, with the exception of the DC offset, to ensure the shape of the model is preserved. The DC offset model parameter is resolved from the data only and has no prior distribution. Although creating strong dependence between the model parameters, each of the change point parameters,  $\delta_{1,\dots,17}$ , are constrained to their order, that is  $\delta_i \sim \text{Unif}(\delta_{i-1}, \delta_{i+1})$ , where  $\delta_0 = 0$  and  $\delta_{18} = T$ . The precision model parameter,  $\tau$ , has an uninformative Gamma prior,  $\tau \sim \Gamma(0.01, 0.01)$ .

In this implementation a random-walk approach to candidate value selection was taken using Normal proposals where the mean value is taken as the current parameter value. The probability of acceptance is

$$\alpha = \min \left( 1, \frac{R(\theta^*)}{R(\theta^{m-1})} \frac{p(\theta^* | \cdot)}{p(\theta^{m-1} | \cdot)} \right), \quad (1)$$

where  $m = 1, \dots, M$  denotes the  $m$ th MCMC simulation from the chain and

$$p(\theta | \cdot) = p(\theta)p(y | \theta; \cdot),$$

where  $\cdot$  denotes the full set of parameters, omitting the parameter of interest,  $\theta$ . Gelman et al. [6, 2] provide more details regarding the use of the Metropolis–Hastings algorithm. The proposal ratio,  $R(\theta^*)/R(\theta^{m-1})$ , is assumed to be unity in all cases.

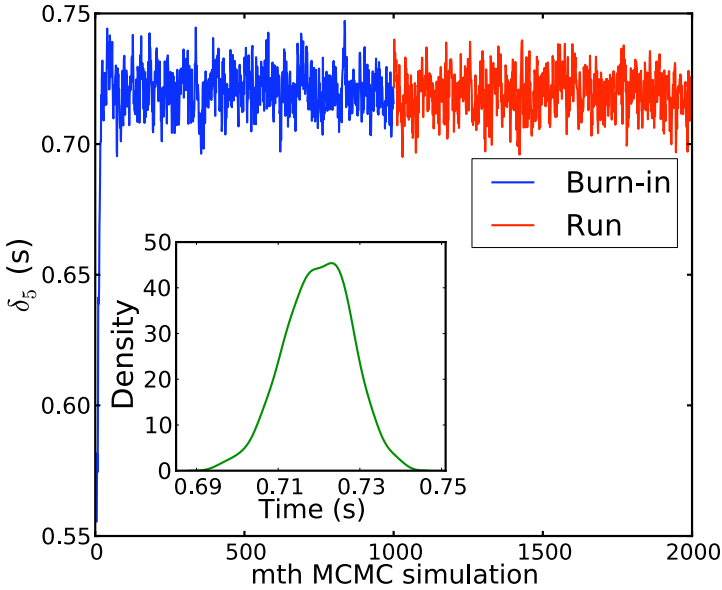


Figure 2: Convergence behaviour and kernel density estimate of  $\delta_5$ .

The MCMC simulation is allowed to run for 2000 iterations—the first 1000 iterations are the ‘burn-in’ and are discarded. From the remaining 1000 iterations the saved model parameters are used to generate the posterior distributions of each model parameter from a kernel density estimate. For the feasibility study, 1000 iterations are used to ensure that all of the parameters have converged. It is expected that in later versions as few as 500 iterations may be used. In order to save time, a single value, taken as the median of the 1000 saved values, is used for interpretation; however, in a clinical application the simulations will be checked for convergence. In the initial setup, the starting values for each turning point,  $\delta_{1,\dots,17}$ , are evenly spaced. Figure 2 illustrates typical convergence behaviour as an MCMC simulation progresses. It is seen that the simulation is exploring the parameter space and that it converges to a solution during the burn-in period.



## 4 Results

A signal with no noise or deviation from a so-called normal heart beat was the first test for a proof-of-concept. The overall model fit is shown in Figure 3(a). It is seen that the model explains the data well and that the turning points and amplitudes are all resolved accurately. This simulation took approximately 1.4 minutes to run on a standard desktop computer (Core i7 DELL Optiplex 9010). The algorithm was coded in C++ using the Boost libraries for random number generation (Mersenne twister).

Matching with the worst case scenario available in the MIT-BIH database on PhysioNet [7], the ECG signal shown in Figure 3(b) is a typical healthy adult ECG signal with 6 dB of white noise. Figure 3(b) shows that the simulation still resolves all of the change points and fits the features of the ECG signal well. Hence, this method of feature extraction is robust against noise.

For further proof-of-concept, Figure 4 shows the model fit to an ECG signal with a delayed T wave and an enlarged, delayed T wave. A delayed T wave (lengthened QT interval) can indicate ventricular tachyarrhythmias [11]; an enlarged T wave can indicate hyperkalemia (too much potassium in the blood), a small T wave can indicate the opposite [3]. Again, it is seen that the simulation accurately explains the data.

## 5 Discussion and conclusions

The work completed so far has obtained a proof-of-concept. In order for this methodology to be robust to variation present in individuals, due to phenomena such as the mean electrical vector and cardiac axis [3], the approach will require multiple QRS complex models and terms to allow for P, QRS or T wave inversion. Moreover, even in healthy individuals there can be great variation in the presentation of an ECG signal [3]. In order to handle this variation, it is proposed that future implementations have decision variables

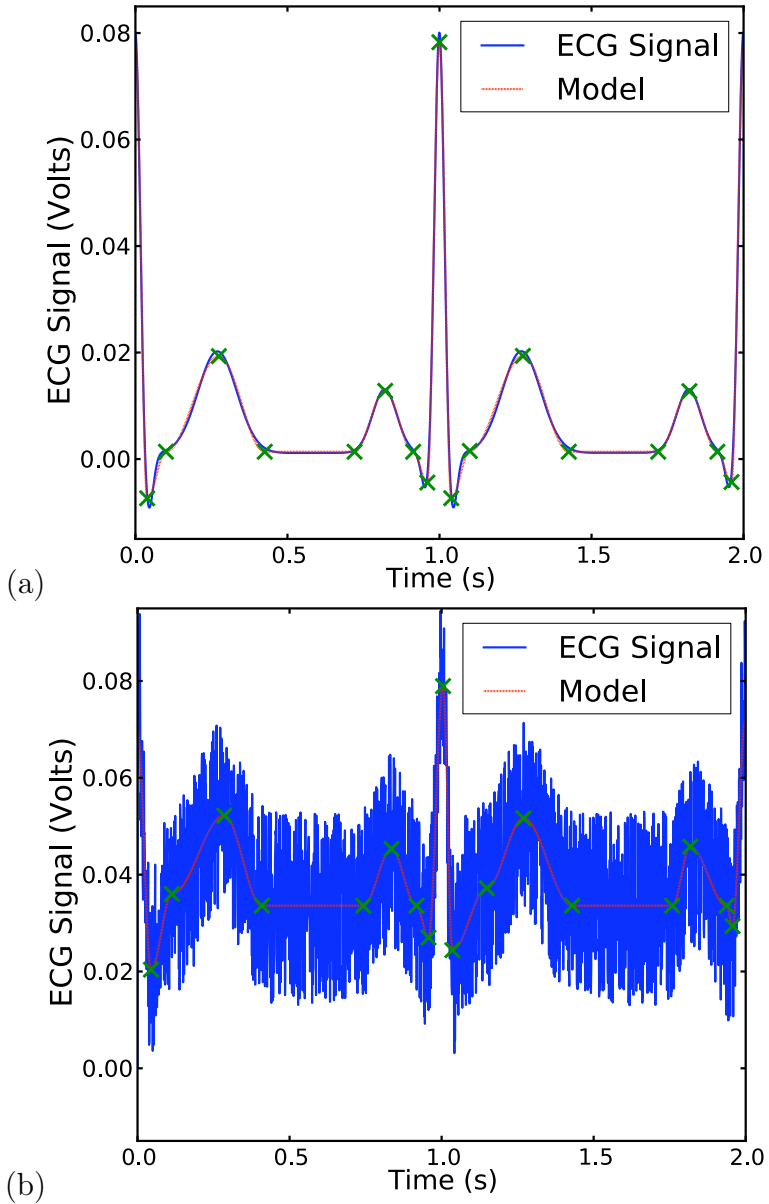


Figure 3: (a) Model fit superimposed onto a typical healthy adult ECG signal  
(b) with 6 dB of white noise

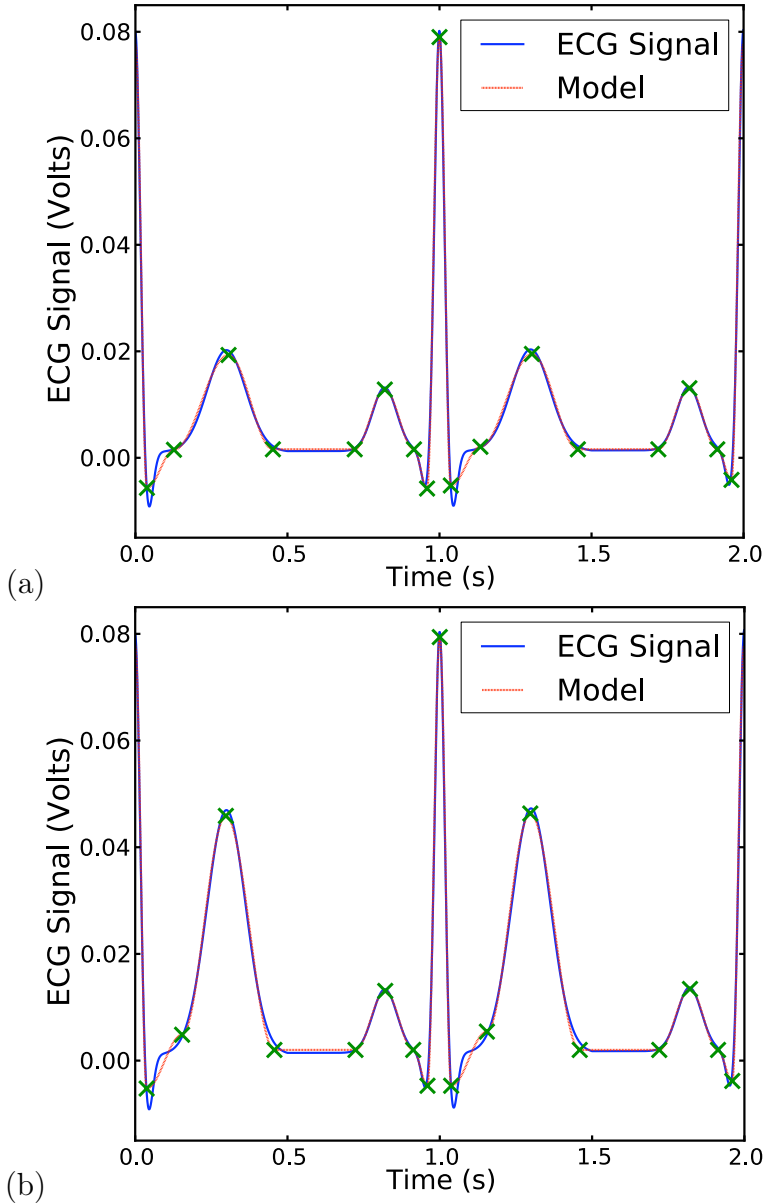


Figure 4: Model fit superimposed onto an ECG signal (a) with a delayed T wave; (b) with a delayed and enlarged T wave.

within the model so that the converged model not only fits the data and therefore provides important clinical information, such as the relative time of the QT interval, but can also flag known cardiac conditions. An example of a potential model decision is to model the ST segment as a line with a gradient and its own offset from the baseline present in the beat. This would enable the model to properly capture all of the information present in the ST segment of the beat—this has clinical relevance in the diagnosis of myocardial infarction [10]. Similarly, a line segment between the S and T waves could also reduce error in the model fit to the trough of the S wave.

Once the modelling work is established for the majority of ‘normal’ and the common abnormal presentations of an ECG signal, it is proposed to extend this work to run the analysis on a field programmable gate array (FPGA). This will allow the analysis of an ECG beat in a few seconds and therefore allow this methodology to be used in a portable, personal device. Applications of this include: clinical use in hospitals and surgeries, monitoring the health of workers in hazardous situations, such as underground mining; and patient care for stay-at-home patients.

## References

- [1] S. Edla, N. Kovvali, and A. Papandreou-Suppappola. Sequential Markov chain Monte Carlo filter with simultaneous model selection for electrocardiogram signal modeling. In *34th Annual International Conference of the IEEE EMBS*, San Diego, California USA, 2012. doi:[10.1109/EMBC.2012.6346915](https://doi.org/10.1109/EMBC.2012.6346915). C35
- [2] A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis: Second Edition*. Chapman & Hall/CRC, 2003. C38
- [3] J. Hampton. *The ECG Made Easy*. Elsevier/Churchill Livingstone, seventh edition, 2008. C33, C35, C40

- [4] C. Lin, C. Mailhes, and J.-Y. Tourneret. P- and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler. *IEEE Transactions on Biomedical Engineering*, 57(12):2840–2849, 2010. doi:[10.1109/TBME.2010.2076809](https://doi.org/10.1109/TBME.2010.2076809). C35
- [5] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3):289–294, 2003. doi:[10.1109/TBME.2003.808805](https://doi.org/10.1109/TBME.2003.808805). C36
- [6] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi:[10.1063/1.1699114](https://doi.org/10.1063/1.1699114). C38
- [7] G. B. Moody, R. G. Mark, and A. L. Goldberger. PhysioNet: a web-based resource for the study of physiologic signals. *IEEE Engineering in Medicine and Biology*, 20(3):70–75, 2001. doi:[10.1109/51.932728](https://doi.org/10.1109/51.932728). C35, C40
- [8] M. Niknazar, B. V. Vahdat, and S. R. Mousavi. Detection of characteristic points of ECG using quadratic spline wavelet transform. In *International Conference on Signals, Circuits and Systems*. IEEE, 2009. doi:[10.1109/ICSCS.2009.5412588](https://doi.org/10.1109/ICSCS.2009.5412588). C35
- [9] J. Pan and W. J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, 1985. doi:[10.1109/TBME.1985.325532](https://doi.org/10.1109/TBME.1985.325532). C34, C35
- [10] K. Thygesen, J. S. Alpert, and H. D. White. Universal definition of myocardial infarction. *Journal of the American College of Cardiology*, 50(22):2173–2195, 2007. doi:[10.1016/j.jacc.2007.09.011](https://doi.org/10.1016/j.jacc.2007.09.011). C43
- [11] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). *Guidance for*

- Industry E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs*, October 2005. <http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm129357.pdf>. C40
- [12] W. Zong, G. B. Moody, and D. Jiang. A robust open-source algorithm to detect onset and duration of QRS complexes. *Computers in Cardiology*, 30:737–740, 2003. doi:10.1109/CIC.2003.1291261. C35
- [13] W. Zong, M. Saeed, and T. Heldt. A QT interval detection algorithm based on ECG curve length transform. *Computers in Cardiology*, 33:377–380, 2006. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4511867](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4511867). C35

## Author addresses

1. **T. Bodisco**, Science & Engineering Faculty, Queensland University Of Technology, Australia  
<mailto:timothy.bodisco@qut.edu.au>
2. **J. D'Netto**, High Performace Computing & Research Support, Queensland University Of Technology, Australia  
<mailto:j.dnetto@qut.edu.au>
3. **N. Kelson**, High Performace Computing & Research Support, Queensland University Of Technology, Australia  
<mailto:n.kelson@qut.edu.au>
4. **J. Banks**, Science & Engineering Faculty, Queensland University Of Technology, Australia  
<mailto:j.banks@qut.edu.au>
5. **R. Hayward**, Science & Engineering Faculty, Queensland University Of Technology, Australia

<mailto:r.hayward@qut.edu.au>

6. **T. Parker**, Faculty of Health, Queensland University Of Technology, Australia

<mailto:t.parker@qut.edu.au>