Cyclists' Cardiac Conundrum

Andrew Nugent[a]**, Yi Ting Loo**[a]**, and Jack Buckingham**[a]

[a]Mathematics of Systems II, University of Warwick

Arrhythmia is an abnormality of the heart's rhythm, caused by problems in the conductive system and resulting in irregular heartbeats. There is increasing evidence that undertaking frequent endurance sports training elevates one's risk of arrhythmia. Arrhythmia is diagnosed using an electrocardiogram (ECG) but this is not typically available to athletes while exercising. Previous research by our external collaborator at Crickles investigates the usefulness of commonly available heart rate data in detecting signs of arrhythmia. It is hypothesised that a feature termed 'gappiness', defined by jumps in the heart rate while the athlete is under exertion, may be a characteristic of arrhythmia. A correlation was found between the proportion of 'gappy' activities and survey responses about heart rhythm problems. We develop on this measure by exploring various methods to detect spikes in heart rate data, allowing us to describe the extent of irregularity in an activity via the rate of spikes. We first compare the performance of these methods on simulated data, where we find that smoothing using a moving average and setting a constant threshold on the residuals is most effective. This method was then implemented on real data provided by Crickles from 168 athletes, where no significant correlation was found between the spike rates and survey responses. However, when considering only those spikes that occur above a heart rate of 160 beats per minute (bpm) a significant correlation was found. This supports the hypothesis that jumps at only high heart rates are informative of arrhythmia and indicates the need for further research into better measures to characterise features of heart rate data.

Arrhythmia | Heart rate data | Spike detection | Poisson process

1. Introduction

Arrhythmia is an abnormality of the heart's rhythm, the most common type of which is atrial fibrillation (AF) (1). A diagnosis is confirmed by observing an episode of AF during an electrocardiogram (ECG) (2), which can be aided by patients wearing a portable ECG recorder (1, 3). While moderate physical exercise is associated with a reduced risk of AF there is some evidence that strenuous endurance exercise increases an individual's risk (4, 5). As such there has been an interest in attempts to use ECGs in smart watches, worn by athletes to record information about their exercise, to detect AF (6, 7). As these devices with ECGs can be expensive, it could be beneficial to instead determine if commonly available heart rate data contains any informative signs of AF.

The Crickles project (8) was established to provide athletes with cardiac stress analysis, calculated using data from their Strava activities, and inform them about their training load relative to other athletes. This data has been used by Ian Green and Mark Dayer to investigate the possibility of using heart rate data to detect the presence of heart rhythm problems (9). They hypothesised that jumps at high heart rates, for example jumping from 180 to 195 beats per minute (bpm) then immediately dropping again, or jumps to unrealistically high values, could be informative. As an athlete's heart rate

monitor continuously samples current in the heart muscle it may plausibly be affected by arrhythmia. Green *et al.* do not claim that such jumps are a direct observation of an episode of AF, nor that they are an accurate reflection of an athlete's heart rate, but rather that these observed jumps are errors in the heart rate monitor. Their key hypothesis is that the presence and/or frequency of such errors may still be informative.

In (9) each activity is classified as 'Regular', 'Unclear', 'Check_Strap' or 'Irregular', based on two key features. The feature of most interest is 'gappiness' shown in Fig.1: an activity is gappy if it contains gaps in the range of observed heart rate readings over a threshold. This threshold is set as
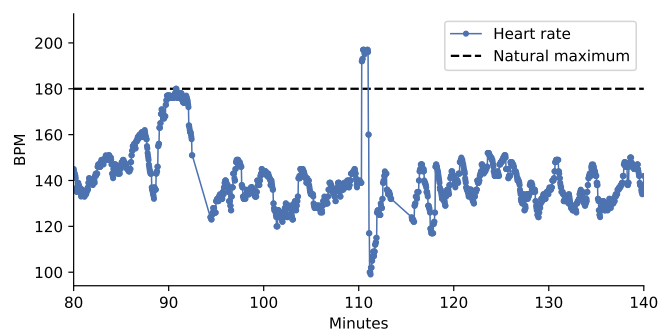

Fig. 1. Time series showing a gappy heart rate, as the heart rate exceeds the natural maximum (dashed line). It is hypothesised that this gap are blips in the heart rate monitor caused by unusual electrical cardiac signals, and so a heightened frequency of these gaps for an athlete may be associated with arrhythmia.

the maximum of the modal heart rate and lactate threshold heart rate (LTHR) of the athlete, which serves as an indication of when the athlete's heart rate is considered to be at

Significance Statement

In recent years there has been increasing evidence that those who frequently participate in high intensity endurance training have an elevated risk of developing Atrial Fibrillation (AF), a type of irregularity in heart rhythm. Although an electrocardiogram (ECG) is available for accurate diagnosis of AF, common sports devices worn by athletes only collect heart rate data in beats per minute (bpm) during an activity, usually sampled at every second which generates a time series. This project analyses characteristics of these time series and their correlation with diagnoses of AF and aims to provide useful indicators of warning signs of irregularity in heart rhythm to an athlete which can prompt earlier diagnosis, and subsequently treatment, of AF.

a high range. Gappiness can be detected by marking heart rates as either 'visited' or 'unvisited' over the course of an activity. If there exists a visited heart rate higher than an unvisited heart rate and the threshold then a gap is observed. As the procedure for detecting gappiness uses the distribution of heart rates across an entire time series it provides a simple Boolean indicator of whether the whole activity is gappy or not. Activities can also be gappy at the bottom end of heart rate readings, using a mirror of the procedure above, but it is assumed that these gaps are not indicative of arrhythmia as the athlete is not under exertion. By design gappiness focuses on jumps that occur at the top end of the heart rate, potentially neglecting any informative behaviour that occurs when the heart rate is normal or relatively low.
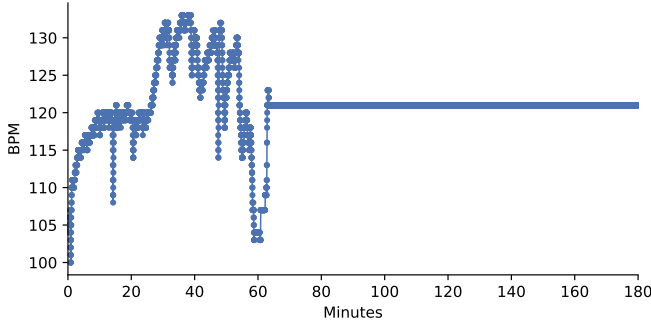

Fig. 2. Time series showing a sticky heart rate, where the heart rate is held constant at 120 bpm for a long period of time which is physiologically implausible especially during endurance training, suggesting an error in the heart rate monitor.

For the second feature, an activity is called 'sticky' if the heart rate is constant for at least 90 seconds, as seen in Fig.2. This indicates that the monitor is not correctly sampling the heart rate. Activities are then classified according to Table 1.

Table 1. Classification of activities

		Non-sticky	Sticky
Gappy	High end only	Irregular	Check_Strap
	Low end only	Unclear	
	At both ends	Unclear	
Non-gappy		Regular	

Activities classified as 'check_strap' or 'unclear' were removed from consideration. This did not pose a problem as athletes typically recorded a large number of activities. For each athlete the *irregular ratio* was defined as the percentage of remaining activities classified as irregular.

The study in (9) aimed to investigate the relationship between an athlete's irregular ratio and their cardiac health. Crickles users were invited to take part in a questionnaire, the key question of which was

> Do you have, or have you had a heart rhythm problem?

The Yes/No response to this question was used to classify athletes into those with and without a diagnosed arrhythmia. The Crickles dataset was restricted to those athletes with

a survey response. The dataset was restricted further by considering only cyclists, who were assumed to use a chest strap to monitor their heart rates. A statistically significant correlation of 0.24 was then found between the irregular ratio and responses to a heart health survey.

Work so far on this problem describes the jumps in time series by labelling the entire activity 'gappy' or 'non-gappy'. We aim to expand on this by providing a more detailed characterisation of the level of irregularity for each activity/athlete, with the goal of strengthening the correlation previously observed. We begin by developing 'spike detection' algorithms to locate jumps throughout the heart rate data, regardless of the range of the heart rate readings, and extract potentially important features such as the height of spikes and the heart rate at which they occur. We then infer the rate of spikes for each athlete and test for a correlation against reported heart rhythm problems.

2. Methods

We begin this section by describing a variety of methods to detect spikes in heart rate data, including smoothing, wavelet transforms and thresholding.

We then introduce a method for simulating heart rate data. Creating simulated data was essential as privacy controls regarding data available to Crickles severely limited our own access. For exploratory analysis we had access to a small number of sample time series, but to utilise a larger dataset we needed the Crickles team to run and return any results. This restriction necessarily shaped our methodology: we first developed various spike detection methods, tuned their parameters, then compared their performance on simulated data to select a single method to apply to the real dataset. We outline our methods and comparison below, then close this section by describing the inference we perform on the detected spikes.

A. Spike detection methods. One method that could be used to detect spikes on time series data is to first detrend the time series (10), for example, by subtracting a smoothed version from the original. Then, a threshold can be set on the residuals, with times when the residuals exceed that threshold registered as spikes. The corresponding heart rate where the spikes occur can be found from the smoothed heart rate time series. The height of the spike can also be calculated by subtracting the smoothed heart rate from the original heart rate data where spikes occurred.

A.1. Smoothing methods. In order to smooth the heart rate time series, one method was to utilize a simple moving average. Letting $\{X(t) : t = 1, \ldots, T\}$ be the heart rate time series, define the smoothed time series, $\overline{X}(t)$ by

$$\overline{X}(t) := \frac{1}{w} \sum_{k=-i}^{i} X(t + k) \qquad [1]$$

where $w = 2i + 1$ is called the width.

A more complex smoothing method finds a sparse representation in a discrete wavelet basis - a widely used technique for signal denoising (11, 12). The mathematical theory of this method is explained in section S1.C of the supplementary material.

A.2. Thresholding methods. A simple thresholding method is to set a constant threshold by evaluating the 95th percentile of the residuals, multiplied by a constant.

Another method is to have an adaptive threshold (13) on the residuals based on a rolling window, instead of evaluating a constant threshold from the entire set of residuals. For each time t in the residual series, a window of width w centred on time t is considered and the threshold, Thresh(t) is calculated by:

$$\text{Thresh}(t) = \text{offset} + \mu(t) + \text{constant} * \sigma(t) \qquad [2]$$

where $\mu(t)$ and $\sigma(t)$ is the mean and standard deviation of residuals within the window respectively. The offset was introduced and set to a low value of 0.01 to prevent any detection of spikes when there are regions of missing data in the heart rate time series. Indeed, the standard deviation $\sigma(t)$ should be zero in these regions, but fluctuations due to finite numerical precision meant that the residuals here are not exactly zero.

A.3. Continuous wavelet transform. The methods discussed so far detect spikes by thresholding the residuals after detrending the time series. An alternative approach instead uses the continuous wavelet transform (CWT). Loosely speaking, the CWT measures the similarity of the signal to the shape of a given 'mother wavelet' at different timescales and locations. By thresholding these coefficients for a suitable, small timescale, we can identify locations in the time series which resemble spikes. Again, this threshold can be either constant or adaptive. The CWT is closely related to the discrete wavelet transform mentioned earlier and both transforms, along with the spike detection methods derived from them, are described in more detail in section S1 of the supplementary material.

B. Simulation algorithm. Our simulation model is inspired by existing models for electricity prices, which are known to spike in a similar fashion (14, 15). These models separate the process into multiple components, each designed to capture a different feature of the data, one of which is sudden short spikes. Taking a similar approach we consider three components which are simulated in order:

1. The base heart rate (X_t)

2. An additional noise process (Y_t)

3. A spike process (Z_t)

All variables are one-dimensional and in units of bpm.

The base heart rate provides the general shape of the time series and is where different activities, such as interval training, can be included. X_t represents the 'true' heart rate of the athlete. We model X_t using an SDE of the form

$$dX_t = -V'(X_t, t)\, dt + \sigma\, dW_t^{(1)} \qquad [3]$$

where $V(x, t)$ is an asymmetric potential surface, σ represents the amount of noise and $dW_t^{(1)}$ is a standard Brownian motion. V may be time dependent, allowing us to simulate complex activities such as modelling interval training by moving the minimum of V, whereas σ is constant. V is chosen to be asymmetric as athletes' heart rates often remain close to, but do not regularly exceed, a natural ceiling rate during exercise.

However, excursions into lower heart rates, in the region 100-160bpm) were not uncommon in our example data and so are penalised less by the asymmetric potential surface. The precise shape of V and values of σ are fit by eye to available data as their purpose is to provide a reasonable background process against which we can try to detect spikes. A plot of V can be found in Fig.S2.1.

The additional noise process allows us to incorporate heteroscedasticity, as step changes in the level of noise are sometimes observed during an activity. An example of this in real data is shown in Fig.3. In some cases, it is the first ten minutes or so of the activity which is noisier than the rest. A possible explanation (9) of this phenomenon is that at the start of an activity a lack of sweat build-up leads to contact failures in heart rate monitors and hence noisy data. As this noise is considered to be an effect of the monitor, not increased volatility in the athlete's heart rate, we include this as a separate independent process.
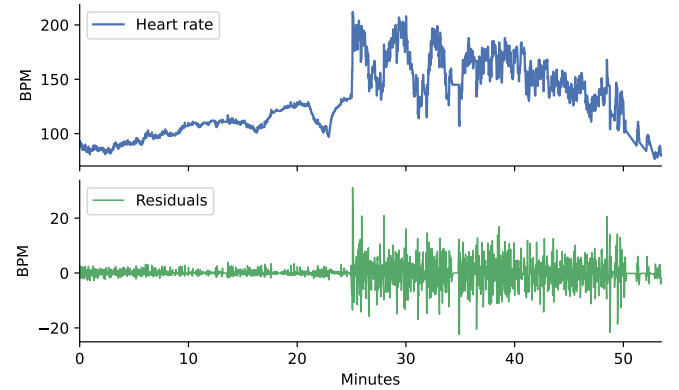

Fig. 3. Time series showing heteroscedasticity in a real heart rate sample. Both the recorded heart rate (top) and residuals after removing a moving average with a ten second window (bottom) are shown. It can clearly be seen that there is a significant increase in point variance after about 25 minutes.

The additional noise process is a simple mean-reverting Ornstein–Uhlenbeck process (16), modelled using the SDE

$$dY_t = -\alpha(t)\, Y_t + \beta(t)\, dW_t^{(2)} \qquad [4]$$

where $\alpha > 0$ describes the rate at which the process reverts to its mean at 0, β is a diffusion coefficient and $dW_t^{(2)}$ is a standard Brownian motion, independent of $dW_t^{(1)}$. α and β are time-varying but do not depend on the state of the process, and are chosen to create additional noise at the start of the activity.

Finally, the spike process, Z_t, is simulated last. As the high heart rates observed during spikes are not considered to be a true reflection of the individual's heart rate (9) it is reasonable to consider them as a separate process. Since arrhythmia-related spikes are also considered to be more likely when the heart rate is high and the heart is under strain (9) Z_t will depend upon X_t.

The spike process is given by the increments of a non-homogeneous jump process J_t. Let $r : \mathbb{R} \to \mathbb{R}$ be a function that describes the rate of spikes as a function of the heart rate.

This function is then used to calculate the Poisson intensity

$$\lambda(t) = r(X_t) \quad\quad [5]$$

which is used to simulate a non-homogeneous Poisson process P_t that determines the timing of spikes. The sizes of spikes are then drawn from a log-normal distribution to give the jump process J_t. The increments of this jump process give the spikes Y_t, hence Y_t is zero except when it is positive at the moment of a spike.

The final output of the simulations is the synthetic heart rate (H_t) which is found by smoothing $X_t + Y_t + Z_t$ using a moving average, then rounding the result to the nearest integer. The simulation method was examined by Crickles and found to produce realistic heart rate time series.

We test the spike detection methods described on simulated data with various features including: interval training, step changes in the level of activity and heteroscedasticity, as well as simple situations with a constant level of activity and noise. Example simulated time series can be found in S2. For each time series the simulation algorithm also outputs a list of spikes containing the time and height of the spikes, as well as the underlying heart rate at the time the spike occurred. We compare this against the spikes detected by each method.

C. Comparison metrics. To compare the performance on simulated data of the various spike detection methods described we consider two error metrics. Both use a list of true spikes taken from the simulated data and a list of detected spikes outputted by the chosen spike detection method.

For the first error metric we set a strict time-window of 5 seconds and scan through the detected spikes. For each detected spike we count it as a 'correct spike' if it can be paired with a true spike within the time-window. If so, both spikes are removed from their respective lists. After scanning through the detected spikes we calculate the precision and recall

$$\text{precision} = \frac{\#\text{ correct spikes}}{\#\text{ detected spikes}}$$
$$\text{recall} = \frac{\#\text{ correct spikes}}{\#\text{ true spikes}}$$

and from this the F_1 score for each activity

$$F_1 \text{ score} = \frac{2}{\text{precision}^{-1} + \text{recall}^{-1}}$$

which is then averaged over all simulated activities. The F_1 score is the harmonic mean of the recall and precision and is a well-established performance metric ([17–19]). The score takes values between 0 (worst score) and 1 (best score).

The second error metric, which we call 'spike density error' involves transforming the locations of spikes from individual points to a distribution of possible points. The motivation for this is that, although spikes peak at a single point, they do have some width, partially due to the smoothing done in real and simulated heart rate time series. For this reason there is not a single time at which the whole spike event occurs, so it seems reasonable to allow the true and detected spike times to be represented by a distribution rather than a point. This metric also allows us to incorporate the height of a spike.

We define a kernel density $f : \mathbb{R} \to \mathbb{R}$, chosen here to be a Gaussian distribution centered at 0 with standard deviation 5 (to correspond to the earlier strict threshold). This kernel density will be translated and scaled to replace the single point value for each spike. Any alternative kernel is acceptable provided it satisfies the following conditions:

1. f is positive and symmetric.

2. f achieves its global maximum at 0.

3. f is integrable and integrates to 1.

Denote by $f_s(t)$ the translated and re-normalised function

$$f_s(t) = \frac{f(t-s)}{\int_0^T f(\tau - s)\, d\tau} \quad\quad [6]$$

where T is the length of the time series.

Given n spikes occurring at times $0 \le s_1 < \cdots < s_n \le T$ with corresponding heights $h_1, \ldots, h_n$, define the spike density $\rho : [0, T] \to \mathbb{R}^+$ by

$$\rho(t) = \sum_{i=1}^{n} h_i\, f_{s_i}(t) \qu\quad [7]$$

Given lists of real and detected spike times, denote the respective densities by ρ_r and ρ_d. The spike density error is given by

$$\epsilon = \frac{1}{T} \int_0^T \left| \rho_r(t) - \rho_d(t) \right| dt \qu\quad [8]$$

which provides an error in units of bmp/hour. Properties of the spike density error are discussed in S4. We also compare this error against that of the null spike detection model, in which no spikes are ever detected. The error for the null model is simply the integral of the true spike density, so we define the error impact

$$\epsilon^* = \frac{1}{T} \int_0^T \rho_r(t) - \left| \rho_r(t) - \rho_d(t) \right| dt \qu\quad [9]$$

which again has units bpm/hour. If ϵ^* is negative this means that overall spikes have been detected erroneously and it would have been more accurate to detect no spikes. If ϵ^* is positive this means the spike detection method has reduced the error (i.e. had a positive impact) by correctly identifying spikes, hence a large ϵ^* is indicative of successful spike detection.

D. Results of comparison. To compare the spike detection methods that were described in subsection 2.A, 45 training sets of heart rate time series of different activity types (as listed in the final paragraph of subsection 2.B) were simulated. This training data was used to tune all the relevant parameters in the spike detection methods, for example the width for moving average and the constant multipliers in the thresholding algorithms. A grid search of parameters was done such that each spike detection method maximizes the F1 score and the error impact with the chosen parameter.

Once all the parameters were tuned such that all spike detection methods were achieving their best scores on the set of training data, they were then applied to the test data which contains 900 simulated heart rate time series, similarly of various activity types. From the results of the spike detection, the

average F1 scores and average error impact across all simulated activities were evaluated, as shown in Fig.4. Furthermore, the average F1 score and average error impact for each subset of simulated activity types were also evaluated for a more detailed analysis.
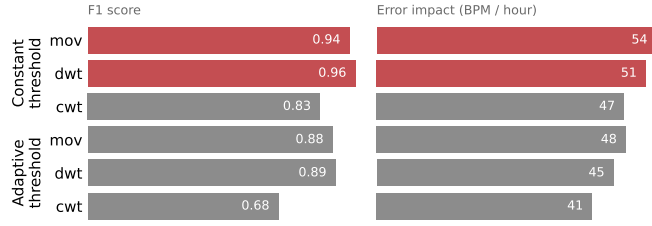

Fig. 4. Bar charts comparing the F1 score (left) and error impact (right) when averaged across all simulated activities. Both metrics show the methods based on moving average (`mov`) and DWT smoothing with a constant threshold to be the best performing (both highlighted in red). Of the two, the moving average method was chosen for spike detection on empirical data because of its simplicity.

From Fig.4, both moving average and DWT smoothing methods with constant threshold were able to detect the true simulated spikes accurately regardless of activity type, showing similar results in both performance metrics. As the moving average method was able to detect a more accurate spike height, this resulted in a greater error impact. This is illustrated in Fig.5, which shows a breakdown of this spike detection method on an example simulated heart rate time series. From these results, and because of its relative simplicity, the method with moving average and constant threshold was chosen as the spike detection method to be used going forward.

This spike detection method, with the same set of parameters as used for the comparison, i.e., width $w = 200\,\mathrm{s}$ for the moving average smoothing and constant multiplier 2.5 for thresholding, were then applied to the empirical data of 168 athletes provided by Crickles.

E. Inference. Once this method has been applied to the empirical data it remains to characterise the frequency of spikes for each athlete. We consider two key lines of enquiry:

1. Is there a correlation between the overall frequency of spikes and reports of heart rhythm problems?

2. Is there a correlation between the frequency of spikes and the heart rate at which they occur? Furthermore, is this relationship the same for individuals with and without heart rhythm problems?

To address the first question we calculate the rate of spike events. We denote the total number of spikes across all an individual's time series by N, and the total length of time series by T. Define the spike frequency

$$\lambda = \frac{N}{T} \qquad [10]$$

which is in fact the maximum likelihood estimate (MLE) if the cumulative number of spike events were a Poisson process with constant rate λ (see S3). This value is calculated for all individuals. As in (9) we then perform a point-biserial correlation check and logistic regression to explore the association between spike rate and heart rhythm problems.

To address our second line of enquiry, the relationship between spike frequency and heart rate, we consider two approaches. Firstly we consider a more detailed model of the spike frequency. We again assume spikes occur at times determined by a Poisson process, but now allow this process to be inhomogeneous with time-dependent rate $\lambda(t)$. Letting $\overline{X}(t)$ denote a smoothed heart rate time series with spikes removed, we assume $\lambda(t)$ is a piecewise constant function of $\overline{X}(t)$. Given these two models we consider the hypotheses:

- H_0: spikes occur according to a Poisson process with constant intensity.

- H_1: spikes occur according to a Poisson process whose intensity is a piecewise constant function of the heart rate.

We perform a likelihood ratio test to determine if the more complex, heart rate dependent model is a significantly better fit to the observed spike locations. The full details of this hypothesis test are described in S3. This hypothesis test is performed for each individual and for the athletes collectively. If the null hypothesis were rejected this would indicate there exists a relationship between heart rate and spike rate that requires further investigation. It is also possible that a relationship between inhomogeneity and arrhythmia may be observed.

We also considered fitting the rate of an inhomogeneous Poisson process whose rate is a continuous function of time or heart rate, but the number of spikes observed in the sample data was too low for such a method to produce meaningful results.

3. Results and Discussion

Spike detection and inference were performed on data from 168 athletes, all of whom had at least 100 activities of different lengths recorded, while 45 of them reported having heart rhythm problems in the Crickles survey. Activities shorter than 5 minutes and those marked as 'sticky' were removed from the dataset, leaving a total of approximately 125,000 activities. On average there were 814 hours of activity data for each athlete.

A. Initial results. Fig.6 shows on the left a box plot of the Poisson rate of spikes for each athlete, split into those athletes with and without arrhythmia. On the right is a corresponding box plot for the percentage of activates marked as irregular, called the irregular ratio, which is the original measure of gappiness in (9).

The majority of athletes have a very low rate of spikes, typically fewer than 0.5 spikes per hour. A point-biserial correlation check is performed to asses the correlation between Poisson spike rate and arrhythmia. We also show results of the same test performed on the irregular ratio.

Table 2. Results of point-biserial correlation check for Poisson spike rate and irregular ratio against cardiac health survey.

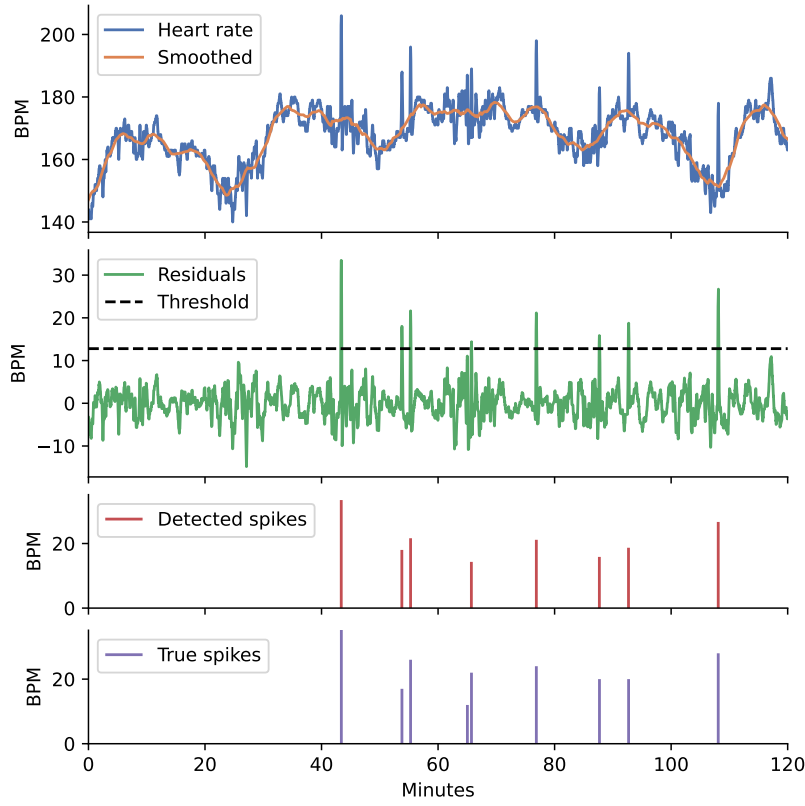|	Correlation	p-value
Poisson spike rate	0.06	0.44
Irregular ratio	0.28	0.00027

Fig. 5. Example of spike detection on simulated heart rate data with method mov_constant from Fig.4. The top plot shows the simulated time series being smoothed by moving average (orange). The second plot shows the resulting residuals (green) and the evaluated constant threshold (dashed). The third and fourth plots are the resulting detected spikes and true simulated spikes respectively, which shows the accuracy of this method in recovering the true spikes.
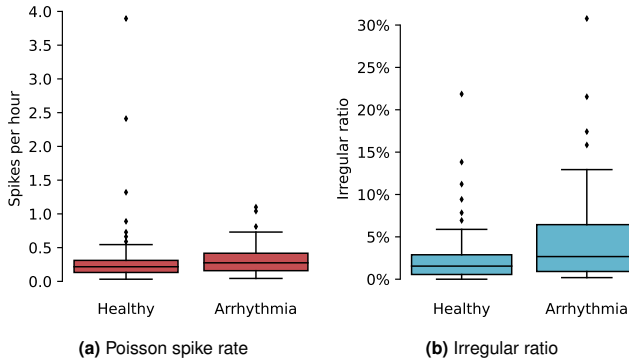

(a) Poisson spike rate   **(b)** Irregular ratio

Fig. 6. These box plots show the distribution of Poisson spike rates (a) and irregular ratios (b) for athletes with and without arrhythmia. Some outliers for both Poisson spike rates and irregular ratios can clearly be observed and will be analysed in subsection 3.B. A correlation with arrhythmia was found from the irregularity ratio, but not for the Poisson spike rates.

From these values, we cannot conclude a statistically significant correlation between the Poisson spike rate and the reported heart rhythm problems. Conversely, the irregular ratio does shows a significant correlation. We now address several possible causes of this difference.

B. Analysis of outliers. In Fig.6a we observe a small number of individuals without arrhythmia for whom the Poisson spike rate is substantially higher than expected. For the purpose of investigating the effect they have on the correlation, outlying

athletes can be removed from the dataset by removing those above a threshold percentile of the data. To ensure we compare the Poisson spike rates and irregularity ratios of the same set of athletes we require that athletes are not outliers in either measure. The effect on correlation and p-values of removing outliers is shown in Fig.7.
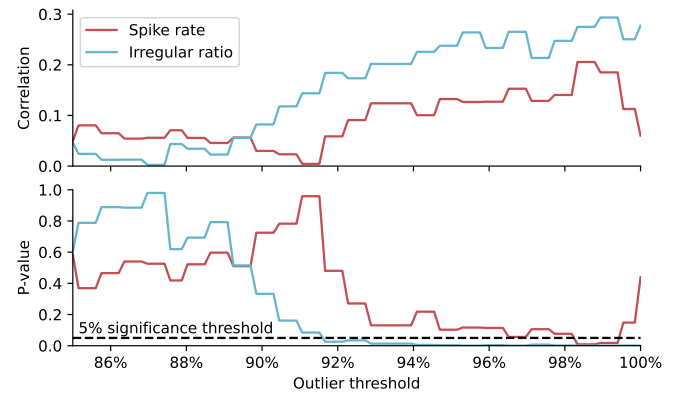

Fig. 7. The plots show the correlation coefficient (top) and corresponding p-values (bottom) of Poisson spike rate (red) and irregular ratio (blue) with arrhythmia, against the outlier threshold, where we remove datasets with either of Poisson spike rate or irregular ratio above the threshold percentile of the data. When those above the 98th percentile are removed, the Poisson spike rate shows a significant correlation, indicating that it is sensitive to some outliers present in the data.

The correlation between the irregularity ratio and reported heart rhythm problems gradually weakens, with an increasing

p-value, as more athletes are removed from the dataset. When considering the Poisson spike rate we observe that removing a small number of outlying athletes (those above the 98th percentile) strengthens the correlation and decreases the p-value, indicating a significant result. However, removing additional athletes then reduces the correlation and significance. This indicates that the correlation between Poisson spike rate and reported heart rhythm problems is highly sensitive to the presence of these outliers.

Unfortunately, due to our limited access to the data, we were not able to identify the cause of these outliers. It is possible that these athletes' activity time series contain some feature that was not represented in our samples that caused our spike detection method to behave unexpectedly. As the Poisson rate is calculated using spikes across all an athletes activities, errors arising in one activity have the capacity to corrupt the overall rate. This sensitivity to individual activities may contribute to these outlying results. By comparison the Boolean classification of 'irregular' activities for the gappiness metric in (9) limits the influence of any one activity on the athlete's irregular ratio, leading to greater robustness at the cost of reduced information.

Alternatively it may be that the spike detection method has performed as expected, but that outliers have a higher rate of spikes due to a fault in their heart rate monitors or other non-arrhythmia related effect. The precise mechanism by which irregular heart rhythms translates to gappy time series is unclear, and the spikes we observe have other possible causes. Hence, even if the spike detection methods operate correctly, the inferred rate of spikes still may not correlate clearly with arrhythmia.

Finally it may possibly be the case that outlying individuals have an un-diagnosed heart condition, meaning that the survey results do not accurately correspond to true heart rhythm problems. Additionally there may be other individuals who responded as having had heart rhythm problems but who are now receiving treatments. As the methods proposed here are not intended as a diagnostic tool we do not speculate as to the heart health of any of the athletes whose data we analyse.

As the cause of these outliers is unclear we would not feel justified in excluding them from our analysis, and so recognise that the current method of spike detection and inference does not improve upon the original gappiness metric proposed in (9).

C. Analysis of heart rate dependence. For the in-homogeneity hypothesis test of spike rates for all athletes, as explained in section 2.E, the results show that for 123 athletes out of the 168, there was significant evidence to reject the null hypothesis, suggesting that spikes occur according to a Poisson process with a rate which is heart rate dependent. A histogram of the frequency of spikes against the heart rate where the spikes occur is shown in Fig.S5.1, which shows that a large number of spikes were detected in the range $105 - 140$bpm.

From the paper (9) by Crickles, it is also hypothesised that only features of heart rate data at high ranges above the lactate threshold heart rate (LTHR) of the athlete are informative of arrhythmia, since the athlete is under exertion at that range of heart rate. Hence, to test our results against this

hypothesis, spikes below a certain heart rate threshold h_{thresh} are removed for all athletes. The corresponding Poisson spike rate for each athlete is then recalculated, and similarly, the point-biserial correlation between the new Poisson spike rate and arrhythmia is evaluated. The plot in Fig.8 shows the correlation coefficients and p-values against increasing heart rate threshold, h_{thresh}.

In the initial section of the plot in the range $h_{\text{thresh}} \leq 75$, both the correlation coefficient and p-value stay relatively constant at the value prior to filtering out any spikes. This is because there are only a very small number of spikes present in the low range of the heart rate, so the correlation tests produce a similar result to that found in subsection 3.A.
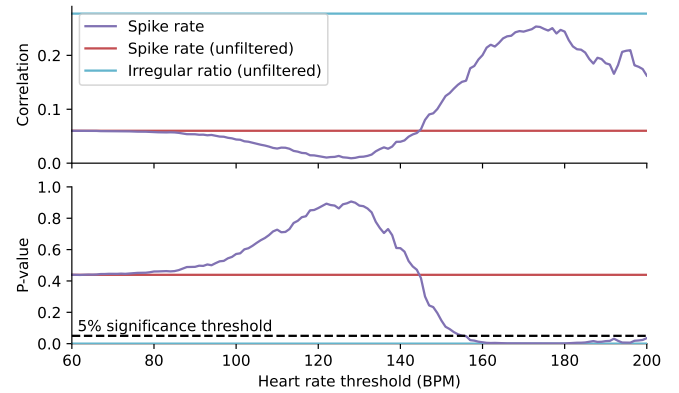

Fig. 8. Correlation with arrhythmia against heart rate threshold, where the top figure shows the correlation coefficient and bottom figure are the corresponding p-values. The horizontal blue and red lines display the results of irregularity ratio from Crickles and Poisson spike rate without filtering out any spikes respectively. The purple curve shows the Poisson spike rate with spikes below the heart rate threshold being removed.

Within the heart rate threshold range $75 \leq h_{\text{thresh}} \leq 130$, there is a decrease in correlation between Poisson spike rate and arrhythmia, which is caused by it being a gray area where this heart rate range is considered as high for some athletes and low for others. From the histogram in Fig.S5.1, the modal value of heart rate where spikes occur also falls in this range. Therefore, past this range, a large number of spikes would be removed relatively randomly from all athletes, so affecting the significance of the spikes detected and thus reducing the correlation with arrhythmia.

Within the range $130 \leq h_{\text{thresh}} \leq 175$, however, a steady increase in correlation coefficient and decrease in p-value is observed. The correlation of spike rate with arrhythmia peaks at $h_{\text{thresh}} = 175$. In this range, the heart rate is likely considered to be high for all athletes. Since now only spikes at high heart rates remain, the increase in correlation justifies that only spikes at a higher heart rate range are significantly indicative of reports of heart rhythm problems. This not only agrees with the hypothesis as mentioned above, but also fits with the medical understanding of arrhythmia that athletes under exertion may induce irregularity patterns in heart rate (20). The correlation value with arrhythmia using Poisson spike rate never achieves a value as high as using Crickles' irregularity ratio, since only a global heart rate spike threshold was used here, instead of an individualised LTHR for each athlete, as was done in Crickles' research. Hence, further re-

search to obtain LTHR data and implement this into the spike detection methods would likely further improve the correlation of Poisson spike rates with arrhythmia.

Finally when $h_{\text{thresh}} \geq 175$, both correlation and p-values decrease as at this point most spikes have been removed. More heart rate data with spikes at such high heart rates would be needed to obtain a better correlation and draw a meaningful conclusion.

4. Further Work

The first step of further work on this project would be as mentioned in the second last paragraph of section C, where the LTHR for each athlete should be incorporated in the spike detection methods as a lower bound so that only spikes at heart rate ranges considered high for each individual athlete are considered. This would require access to further historical activity data from the athlete for an accurate estimate of the athlete's LTHR, which will likely be a challenge due to data privacy.

With the current spike detection method, it is also observed that a large proportion of spikes were detected from the first 10 minutes of the activity data, shown in the histogram in Fig.S6.1. This may be largely caused by the defects of heart rate monitors that struggle to obtain accurate readings when athletes start their activities. For example, chest straps are prone to contact failures with the skin before the build up of sweat, resulting in inaccurate and noisy heart rate readings. More analysis will need to be done on coping with noisy data from the heart rate monitor such as this, so that the spike detection methods can properly distinguish noise caused by external effects from spikes caused by heart rhythm irregularities. To cope with initial noisy readings, one method will be to discard readings from the first 5 to 10 minutes of activity before implementing the spike detection methods.

Instead of explicitly detecting spikes in heart rate data, further work can also be done to characterise other features of heart rate time series, and their correlation with arrhythmia. It is observed from empirical data that the volatility of heart rate varies not only between separate activities but also within heart rate data recorded by an athlete from a single activity. Hence, analysing and measuring the volatility of heart rate data could also potentially be informative of heart rhythm problems.

In addition to heart rate measurements, sport devices often record other data such as the velocity, distance, elevation, power output by the athlete, etc., during the activity. Analysis on this additional data could perhaps be used to filter out any characteristics in the heart rate data such as change in volatility or drastic trends in heart rate that are caused by a sudden exertion from the athlete while exercising.

In conclusion, other than exploring different data driven methods, there is scope for lots of further work to be done in improving the current spike detection method and modelling other features of the of heart rate time series, in order to obtain stronger correlations of these metrics with arrhythmia.

Bibliography

1. NHS (2021) Arrhythmia (https://www.nhs.uk/conditions/arrhythmia/). Accessed: 03/06/2021.
2. with the Special Contribution of the European Heart Rhythm Association (EHRA) D, et al. (2010) Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the european society of cardiology (esc). *European heart journal* 31(19):2369–2429.
3. Camm AJ, Corbucci G, Padeletti L (2012) Usefulness of continuous electrocardiographic monitoring for atrial fibrillation. *The American journal of cardiology* 110(2):270–276.
4. Aizer A, et al. (2009) Relation of vigorous exercise to risk of atrial fibrillation. *The American journal of cardiology* 103(11):1572–1577.
5. Centurión OA, et al. (2019) The association between atrial fibrillation and endurance physical activity: how much is too much? *Journal of Atrial Fibrillation* 12(3).
6. McManus DD, et al. (2013) A novel application for the detection of an irregular pulse using an iphone 4s in patients with atrial fibrillation. *Heart Rhythm* 10(3):315–319.
7. Li KHC, et al. (2019) The current state of mobile phone apps for monitoring heart rate, heart rate variability, and atrial fibrillation: narrative review. *JMIR mHealth and uHealth* 7(2):e11606.
8. (year?) Crickles (https://crickles.casa/). Accessed: 03/06/2021.
9. Green I, Dayer M (2021) Detecting arrhythmia from exercise data. *Unpublished*.
10. Mills TC (2003) 'classical' techniques of modelling trends and cycles in *Modelling Trends and Cycles in Economic Time Series*. (Palgrave Texts in Econometrics. Palgrave Macmillan, London).
11. Graps A (1995) An introduction to wavelets. *IEEE Computational Science and Engineering* 2(2).
12. Mallat S (2009) *A Wavelet Tour of Signal Processing*. (Academic Press), Third edition.
13. Van Brakel JPG (2014) Robust peak detection algorithm using z-scores (https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data/22640362#22640362). Accessed: 26/05/2021.
14. Mayer K, Schmid T, Weber F (2015) Modeling electricity spot prices: combining mean reversion, spikes, and stochastic volatility. *The European Journal of Finance* 21(4):292–315.
15. De Jong C (2006) The nature of power spikes: A regime-switch approach. *Studies in Nonlinear Dynamics & Econometrics* 10(3).
16. Gardiner CW, , et al. (1985) *Handbook of stochastic methods*. (springer Berlin) Vol. 3.
17. Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26(3):297–302.
18. Sørensen TJ (1948) *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*. (I kommission hos E. Munksgaard).
19. Chinchor N, Sundheim BM (1993) Muc-5 evaluation metrics in *Fifth Message Understanding Conference (MUC-5): Proceedings of a Conference Held in Baltimore, Maryland, August 25-27, 1993*.
20. Guasch E, Mont L (2017) Diagnosis, pathophysiology, and management of exercise-induced arrhythmias. *Nat Rev Cardiol* 14:88–101.
21. Daubechies I (1992) *Ten lectures on wavelets*. (SIAM).
22. Kaiser G (2011) *A Friendly Guide to Wavelets*. (Birkhäuser Boston).
23. Lee GR, Gommers R, Waselewski F, Wohlfahrt K, O'Leary A (2019) Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software* 4(36):1237.
24. Gao HY (1998) Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics* 7(4):469–488.

Supplementary Information

S1. Spike detection methods

A. An introduction to wavelets. What follows is a very fast paced overview of the relevant theory of wavelets for this project. For a friendlier introduction the reader is referred to the excellent texts by Daubechies (21) and Kaiser (22). In particular, Daubechies' book formed the basis for this section.

The Fourier transform gives a decomposition of a signal into a frequency spectrum. Low frequency components describe the slower changing aspects of the signal such as the overall trend, while features like spikes require the higher frequency components. However, representation of a spike in Fourier space requires contributions from all frequencies. Indeed, the most extreme case of a spike is the Dirac delta function centered at a point $c \in \mathbb{R}$, $\delta_c(t)$, whose Fourier transform is $\hat{\delta}_c(\omega) = e^{-i\omega c}$. This makes it hard to use the Fourier transform to isolate spikes within a signal and it is therefore desirable to look for a decomposition of the signal which localises 'content' in both frequency and time.

Wavelet theory (12, 21) covers many related transforms to perform this time-frequency decomposition, the most important being the continuous and discrete wavelet transforms. They each begin with a *mother wavelet*, $\psi \in L^2(\mathbb{R})$, such as the one in Fig.S1.1, which defines a short-lived oscillation. By translating and rescaling ψ, we can generate a family of *(child) wavelets*

$$\psi^{a,b}(t) = |a|^{-1/2} \psi \left(\frac{t-b}{a} \right). \qquad [11]$$

Here $a \in \mathbb{R} \setminus \{0\}$ is a scaling parameter and $b \in \mathbb{R}$ is a translation parameter. By rescaling by a factor of $|a|^{-1/2}$ we preserve the L^2 norm of the wavelet, $\|\psi^{a,b}\| = \|\psi\|$.
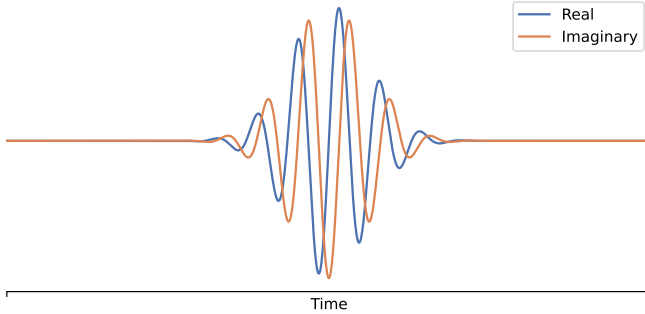

Fig. S1.1. An example of a mother wavelet: the complex Morlet wavelet with bandwidth $2\,\mathrm{s}^2$ and central frequency $1\,\mathrm{Hz}$, $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} e^{2\pi i t}$.

We must place some restrictions on what is an acceptable mother wavelet. It is sufficient to require that $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ with

$$\int_{\mathbb{R}} \psi(t)\, dt = 0, \qquad \|\psi\| = \int_{\mathbb{R}} |\psi(t)|^2\, dt = 1. \qquad [12]$$

The requirement that ψ integrates to zero will allow us to invert the wavelet transforms and such a mother wavelet is called *admissible*. The condition on the L^2-norm is just a normalisation convention. It is possible to weaken these restrictions so that ψ need not be in $L^1(\mathbb{R})$ and need not integrate to 0 but is still admissible, however, this is of little practical use.

The *continuous wavelet transform (CWT)* of a signal $f(t)$ is

$$\begin{aligned} (\mathcal{W}f)(a,b) &= \langle f, \psi^{a,b} \rangle \\ &= |a|^{-1/2} \int_{\mathbb{R}} f(t)\overline{\psi}\left(\frac{t-b}{a}\right) dt. \end{aligned} \qquad [13]$$

Like the Fourier transform, it is defined using an inner product. However, unlike with the Fourier transform, the CWT includes a lot of redundancy. This means that there is no single formula for inverting the CWT. However, a canonical formula known as the *resolution of identity* does exist

$$f(t) = C_\psi^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mathcal{W}f)(a,b)\, \psi^{a,b}(t)\, \frac{1}{a^2}\, da\, db \qquad [14]$$

where

$$C_\psi = 2\pi \int_{-\infty}^{\infty} |\hat{\psi}(\xi)|^2 \frac{1}{|\xi|} d\xi.$$

Here $\hat{\psi}$ denotes the Fourier transform of the mother wavelet. It can be shown that for an admissible wavelet (Eq. (12)), the integral defining the constant C_ψ exists and is finite.

In the *discrete wavelet transform (DWT)*, the scaling and translation parameters only take discrete values

$$a_m = 2^m, \quad b_n = n2^m \qquad [15]$$

giving child wavelets

$$\psi_{m,n}(t) = \frac{1}{\sqrt{2^m}} \psi \left(\frac{t}{2^m} - n \right). \qquad [16]$$

Therefore the DWT can be viewed as a discrete sampling of the CWT, as illustrated in Fig.S1.2. At higher frequencies, the DWT has higher temporal resolution to capture the faster changes but lower frequency resolution since this is more important for lower frequency components. Note that at this point, we are still in the regime where the original signal is sampled continuously.
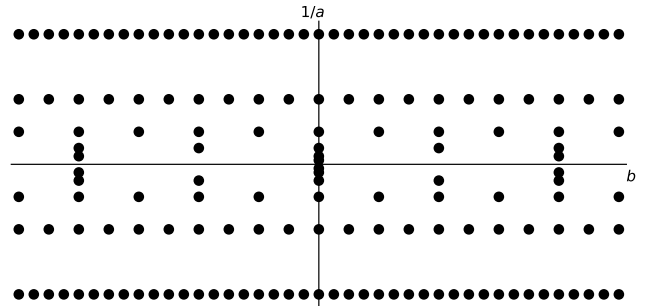

Fig. S1.2. The discrete wavelet transform can be viewed as a discrete sampling of the continuous wavelet transform. The dots in this figure show these discrete locations. Note that the y-axis shows $1/a$ which is proportional to the central frequency of the wavelet. The x-axis plots the translation parameter, b.

For particular choices of the mother wavelet, ψ, the wavelets $\psi_{m,n}$ form a basis for $L^2(\mathbb{R})$. For more specific choices, this basis is orthonormal and the coefficients in the decomposition can be calculated easily by taking the inner product with the respective wavelets

$$f(t) = \sum_{m,n} \langle f, \psi_{m,n} \rangle \psi_{m,n}. \qquad [17]$$

If the basis is not orthonormal then we can still calculate the coefficients using the Riesz representors, $\tilde{\psi}_{m,n}$, of the dual basis. These are functions satisfying $\langle \psi_{m,n}, \tilde{\psi}_{m',n'} \rangle = \delta_{m,m'} \delta_{n,n'}$, where $\delta_{j,k}$ is the Kronecker delta. The coefficients of f in the wavelet basis are given by taking the inner product with this dual basis,

$$f(t) = \sum_{m,n} \langle f, \tilde{\psi}_{m,n} \rangle \psi_{m,n}. \qquad [18]$$

The choice of discretisation made in Eq. (15) is motivated by multiresolution analysis which will be introduced shortly, however other discretisations are possible. The subject resulting from a combination of mother wavelet and discretisation which doesn't yield a basis is that of redundant *frames* and is beyond the scope of this introduction.

For a practical understanding of wavelets, it is necessary to introduce multiresolution analysis. This is by far the most common implementation for the DWT and, for example, is used by `pywavelets` (23), the library used for this project. Aside from computational efficiency, one main reason for this is that it provides a method to construct wavelets which form an orthonormal basis.

Rather than explicitly defining the mother wavelet, ψ, multiresolution analysis instead begins with a nested sequence of (closed) approximation spaces

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots$$

satisfying

$$\overline{\bigcup_{m \in \mathbb{Z}} V_m} = L^2(\mathbb{R}), \qquad \bigcap_{m \in \mathbb{Z}} V_m = \{0\}.$$

We further require that

$$\forall n \in \mathbb{Z}, \quad f \in V_0 \Leftrightarrow t \mapsto f(t - n) \in V_0,$$
$$\forall m \in \mathbb{Z}, \quad f \in V_m \Leftrightarrow t \mapsto f(2^m t) \in V_0.$$

The latter condition here is what gives multiresolution analysis its name. Finally, we require the existence of a function $\phi \in V_0$ such that $\{\phi_{0,n} : n \in \mathbb{Z}\}$ form an orthonormal basis for V_0, where for all $m, n \in \mathbb{Z}$, $\phi_{m,n} = 2^{-m/2} \phi(2^{-m} t - n)$. This ϕ is called the *scaling function* for the multiresolution analysis.

With this construction, for $f \in L^2(\mathbb{R})$ we can build successive approximations $f_m \in V_m$ by projecting orthogonally onto V_m. By construction, we have $f_m \to f$ pointwise as $m \to -\infty$. Furthermore, it is clear that the scaling function, ϕ, is sufficient to reconstruct the full multiresolution analysis set-up.

We are now in a position to construct an orthonormal wavelet basis for $L^2(\mathbb{R})$. For each $m \in \mathbb{Z}$, let W_m be the orthogonal complement of V_m in V_{m-1}, so that

$$V_{m-1} = V_m \oplus W_m, \quad \text{and} \quad W_m \perp W_{m'} \text{ whenever } m \neq m'.$$

Define

$$\psi = \sum_{n \in \mathbb{Z}} g_n \phi_{-1,n} \ \in W_0 \subset V_{-1} \qquad [19]$$

where for each $n \in \mathbb{Z}$, $g_n = (-1)^n \overline{\langle \phi, \phi_{-1,-n+1} \rangle}$. It is proved in (21, Theorem 5.1.1) that ψ is a mother wavelet generating an orthonormal basis of wavelets $\{\psi_{m,n} : m, n \in \mathbb{Z}\}$ for $L^2(\mathbb{R})$ where each $\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m} t - n)$. Further, for each

$m \in \mathbb{Z}$, $\{\psi_{m,n} : n \in \mathbb{Z}\}$ are an orthonormal basis for the subspace W_m. Note however, that this choice of ψ is not unique.

We now describe an efficient algorithm for calculating these coefficients, which is an example of something known as a sub-band filtering scheme. Suppose that we have a fine scale approximation $f \in V_m$ of some function f_* given by orthogonal projection, and that we have calculated the coefficients $(\langle f, \phi_{m,n} \rangle : n \in \mathbb{Z})$ in the basis generated by the scaling function. By rescaling our scaling function, we may assume without loss of generality that $m = 0$. Thus we have $f = \sum_{n \in \mathbb{Z}} \langle f, \phi_{0,n} \rangle \phi_{0,n} \in V_0$. It is our aim to replace this representation of f with one using the natural basis for $V_0 = V_M \oplus W_M \oplus \cdots \oplus W_1$ for a given integer $M \geq 1$ called the *level* of the decomposition,

$$f = \sum_{n \in \mathbb{Z}} a_{M,n} \phi_{M,n} + \sum_{m=1}^{M} \sum_{n \in \mathbb{Z}} d_{m,n} \psi_{m,n}. \qquad [20]$$

The coefficients $a_{M,n}$ on the scaling function part of the basis will be called the *approximation coefficients*, while the coefficients $d_{m,n}$ on the wavelet part of the basis will be called the *detail coefficients*.

We have already defined coefficients g_n and can define coefficients h_n such that

$$g_n = \langle \psi, \phi_{-1,n} \rangle, \quad h_n = \langle \phi, \phi_{-1,n} \rangle. \qquad [21]$$

It is shown in (21, section 5.6) that for $f \in L^2(\mathbb{R})$ and for all $m, n \in \mathbb{Z}$,

$$d_{m,n} = \langle f, \psi_{m,n} \rangle = \sum_{k \in \mathbb{Z}} \overline{g_{k-2n}} \langle f, \phi_{m-1,k} \rangle, \qquad [22a]$$

$$a_{m,n} = \langle f, \phi_{m,n} \rangle = \sum_{k \in \mathbb{Z}} \overline{h_{k-2n}} \langle f, \phi_{m-1,k} \rangle. \qquad [22b]$$

This gives us a recursive algorithm for computing the approximation and detail coefficients of f up to any level $M \geq 0$, beginning with the inner products we have already computed $\{\langle f, \phi_{0,n} \rangle : n \in \mathbb{Z}\}$. Indeed, using Eq. (22) we can generate the coefficients for level $m + 1$ by convolving the approximation coefficients for level m with the respective filters $(\overline{g_{-n}})$ and $(\overline{h_{-n}})$ and discarding the odd terms in the result. This subsampling of the coefficients is referred to as *decimation*. The pair of filters $(\overline{g_{-n}})$ and $(\overline{h_{-n}})$ are known in the signal processing literature as a *filter bank*.

Of course, in practice we do not have a continuously sampled signal. Instead, we assume sampling at regular intervals. This actually works to our advantage. Firstly, we can perform all the convolutions very efficiently using a fast Fourier transform. Secondly, we do not need to further approximate f at the start of the multiresolution analysis. Rather, we can exactly compute the first set of approximation coefficients $a_{0,k} = \langle f, \phi_{0,k} \rangle$ for $k \in \mathbb{Z}$ using a convolution. In this setting, the original sampling of f can be exactly recovered no matter what level M of decomposition is used. However, length of the time series limits how large we can make M. Indeed, at each step the coefficients are decimated by a factor of two and this cannot be continued forever.

More generally, when we have a discretely sampled signal we must be careful not to trust wavelet coefficients corresponding
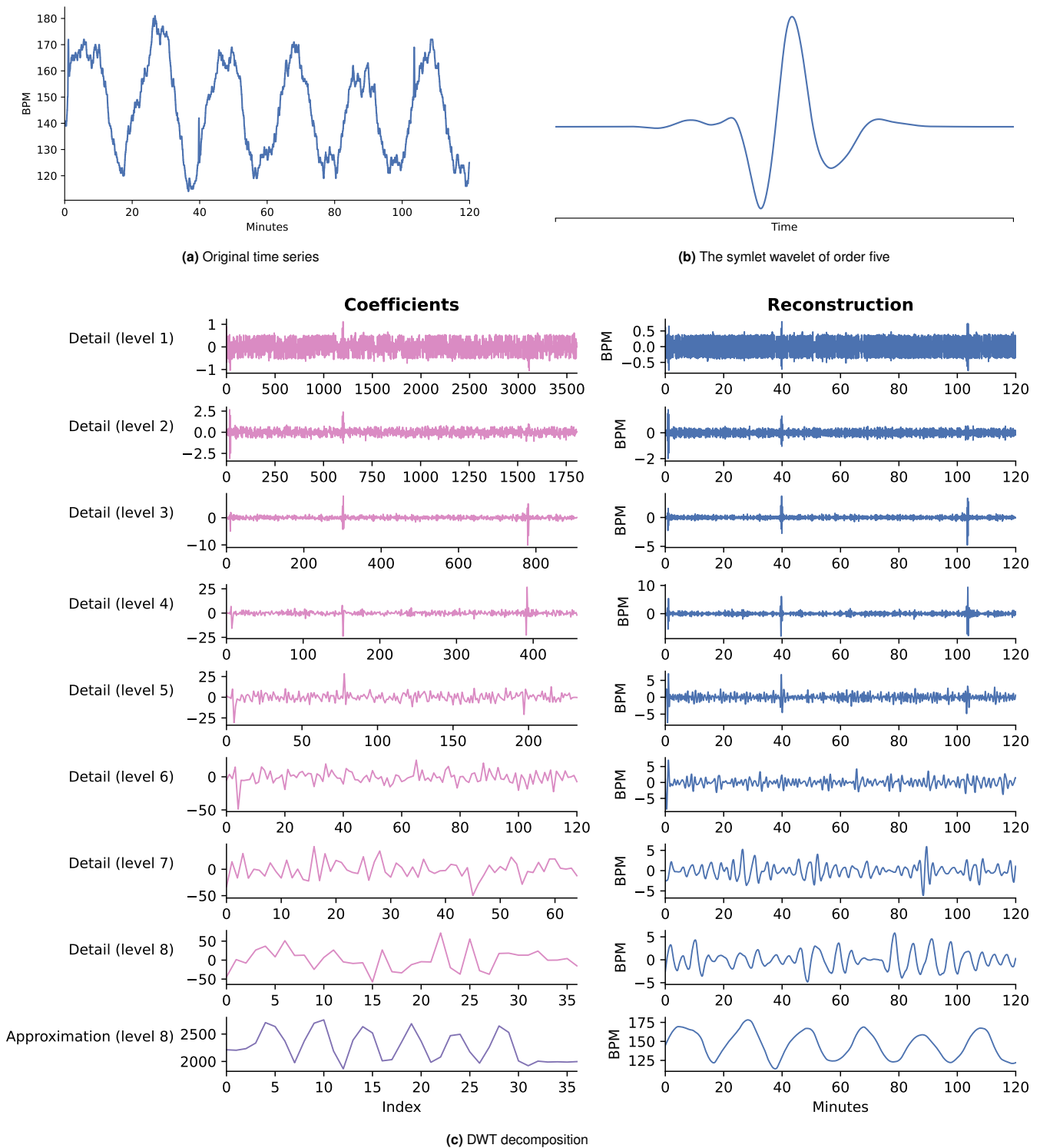
(a) Original time series

(b) The symlet wavelet of order five

(c) DWT decomposition

Fig. S1.3. An example of a multiresolution decomposition using DWT. Figure (a) shows the original, simulated time series; figure(b) shows the fifth order symlet wavelet implicitly defined by the multiresolution analysis and figure (c) shows the detail and approximation coefficients for the first eight decomposition levels. The decomposition was performed after extending the signal symmetrically at both ends to reduce end effects in the coefficients. In figure (c), the coefficients (left) are decimated at each level so that each successive level has half as many coefficients. Their amplitude grows with the level of the decomposition because the wavelet basis is scaled down by a factor of $\sqrt{2}$ at each level. The reconstructions (right) show the contribution of each level to the original time series and summing these will recover the original time series.

to frequencies higher than the Nyquist frequency. This is especially important for the CWT, where the algorithm does not naturally prevent us from going beyond this. Unlike in Fourier analysis however, this is just a rule of thumb, since the concept of a single frequency is stronger for some families of mother wavelet than for others.

It is also common in a practical setting to extend the signal in some way to avoid end effects. By default we could just interpret the signal as zero outside the sampled interval. However, by extending it symmetrically, we can reduce end effects in the coefficients for each level.

Figure S1.3 exhibits the decomposition of a simulated heart rate using an orthogonal wavelet called the 'symlet' of order 5.

Finally, we note that we can relax the condition on the basis $\{\phi_{-1,n} : n \in \mathbb{Z}\}$ being orthonormal (12, section 7.4). If instead they simply form a Riesz basis for V_0 then the wavelet basis we obtain for $L^2(\mathbb{R})$ is not orthonormal, but coefficients in this basis can be calculated using the dual basis, as before. Such wavelet bases are often called *biorthogonal*. The filter banks for the corresponding multiresolution analysis contain four filters: one pair for analysis (deconstruction) and a second pair for synthesis (reconstruction).

B. Spike detection using the continuous wavelet transform. We can use the continuous wavelet transform to perform spike detection on our heart rate time series. As can be seen in figure S1.4, spikes manifest as ridges into the smaller scales (higher frequencies). Choosing an appropriate scale, we obtain a function of the translation parameter alone which we can threshold to classify spikes. As with the methods which threshold residuals of the heart rate time series, we can use either a constant threshold or an adaptive threshold here.

The choice of wavelet is important for performing spike detection using the CWT. A real-valued wavelet will give an oscillating pattern for each spike, making it difficult to know how many spikes there are and where they are located. Therefore, in this report we use a complex wavelet from the family of complex Morlet (or Gabor) wavelets. These have form

$$\psi(t) = \frac{1}{\sqrt{b\pi}} e^{-t^2/b} e^{2\pi i c t} \qquad [23]$$

where the bandwidth, b, and central frequency, c, are parameters. An example is shown in figure S1.1. In this project, we used values of $b = 1\,\mathrm{s}^2$ and $c = 1\,\mathrm{Hz}$. The benefit of using a complex wavelet is that instead of oscillating either side of zero, the wavelet coefficients instead orbit zero in the complex plane. Therefore, taking the absolute value of the coefficients yields a single smooth hump for each spike.

After thresholding the CWT coefficients, we are left with several contiguous regions in which a spike occurs. We assign a spike to each of these regions by taking the midpoint.

C. Spike detection using the discrete wavelet transform. In the main text we described our principal method of spike detection, which proceeded by

1. smoothing the original signal,

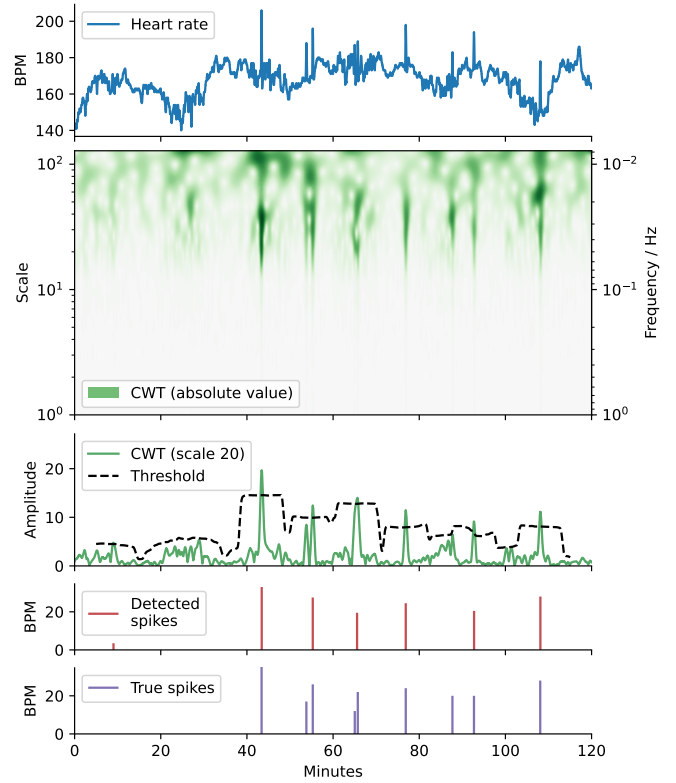2. subtracting the smoothed signal from the original,

Fig. S1.4. Example of spike detection with a continuous wavelet transform. First the simulated heart rate in the top plot is passed through a continuous wavelet transform using a complex Morlet wavelet with bandwidth $2\,\mathrm{s}^2$ and central frequency $1\,\mathrm{Hz}$ (see Fig.S1.1). Its absolute values are shown in the second plot. An adaptive threshold is then used on the CWT at scale $20\,\mathrm{s}$ $(0.05\,\mathrm{Hz})$, shown in the third plot. The bottom two plots show that most of the simulated spikes were detected in this example, although three were missed. Furthermore, an additional spike was detected at the beginning of the time series, although this was assigned a very small height so will not greatly affect the error impact.

3. using a threshold (constant or adaptive) to identify spikes in the residuals.

The discrete wavelet transform provides an alternative smoothing method for this process.

In *signal denoising* (12, chapter 11), we attempt to find a sparse representation for our signal. As we have already seen, using multiresolution analysis we can obtain a representation using a set of approximation and detail coefficients. As is visible in figure S1.3c, not all of these detail coefficients are significant, and many could reasonably be regarded as noise. By shrinking the smaller coefficients to zero, we can obtain a sparse approximation to the original time series.

There are a few choices for the type of thresholding used to obtain the sparse representation. The most popular are the *hard threshold*, $T_\lambda(x)$, and the *soft threshold*, $S_\lambda(x)$, since they are most amenable to statistical risk analysis. However, in this project we used the *non-negative garotte* (24), $G_\lambda(x)$, which has desirable features from both. The three methods are given by

$$T_\lambda(x) = \begin{cases} x & \text{if } |x| \geq \lambda, \\ 0 & \text{if } |x| < \lambda; \end{cases} \qquad [24a]$$

$$S_\lambda(x) = \begin{cases} x - \lambda & \text{if } x \geq \lambda, \\ x + \lambda & \text{if } x \leq -\lambda, \\ 0 & \text{if } -\lambda < x < \lambda; \end{cases} \quad \text{[24b]}$$

$$G_\lambda(x) = \begin{cases} x - \lambda^2/x & \text{if } |x| \geq \lambda, \\ 0 & \text{if } |x| < \lambda. \end{cases} \quad \text{[24c]}$$

D. Handling missing data. One problem encountered when implementing the spike detection method on large sets of heart rate data provided by Crickles was that some recorded time series have portions of missing data, ranging between minutes to hours within a single recorded activity, resulting in unrealistic results, such as detecting an implausibly large amount of spikes.

To handle these issues, dummy heart rate readings were filled into the regions with missing data by linearly interpolating between the two points where the heart rate was last and next recorded, so that there are now heart rate readings every second in the time series. This was done before the heart rate time series was smoothed with a moving average. After this interpolated time series was smoothed, the regions where heart rate readings were added in for interpolation were removed so that the final smoothed version was aligned to the original heart rate data. The residuals were then obtained and a threshold was set on the residuals to detect spikes as explained before. This method was used instead of simply removing the empty regions from the data set as it ensured that sections of recorded heart rate separated by missing regions were unrelated, and not treated as one continuous activity.

S2. Example simulated time series

We show here the potential surface used to simulate the base heart rate in Eqn.3. Following this we give two examples of simulated timeseries.
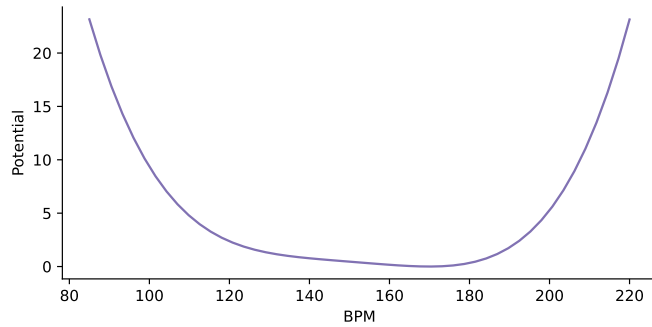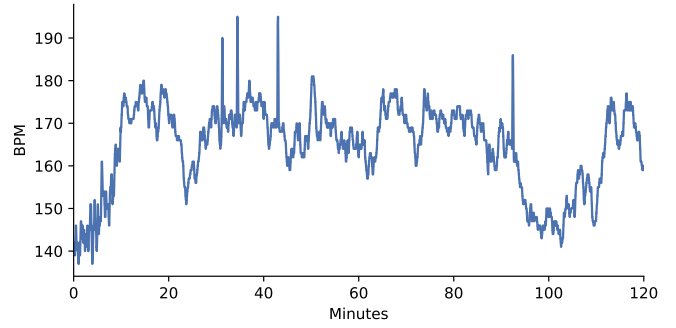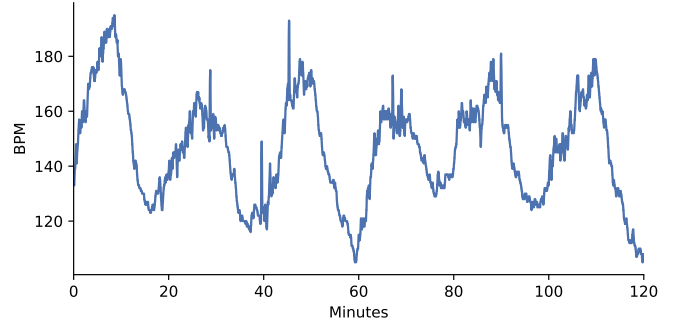

Fig. S2.1. Example of a fixed potential surface as a function of heart rate. The asymmetry in the potential surface ensures that the heart rate does not frequently exceed a natural maximum, while allowing excursions to lower heart rates.

(a) Random activity pattern

(b) Interval training

Fig. S2.2. Example simulated time series with infrequent spikes. In (a) a constant potential surface is used, with additional noise during the first 10 minutes. In (b) the minimum of the potential surface is used to generate an interval training pattern. Additional noise is added in every other 10 minute section, beginning with the first 10 minutes.

S3. Poisson inference

We describe here the likelihood ratio test used to consider the following hypotheses:

- H_0: spikes occur according to a Poisson process with constant intensity.

- H_1 spikes occur according to a Poisson process whose intensity is a piecewise constant function of the heart rate.

We begin by calculating the MLE under H_0. Denote the total number of spikes across all an individual's time series by N, and the total length of the time series by T. The log-likelihood function under H_0 is

$$\ell_0(\lambda_0) = N \log(\lambda_0) - \lambda_0 T \quad \text{[25]}$$

which is maximised by

$$\hat{\lambda_0} = \frac{N}{T} \quad \text{[26]}$$

assuming of course that $T > 0$.

We now consider the model under the alternative hypothesis. Denoted by $X(t)$ a smoothed heart rate time series with spikes removed. Divide a range of possible heart rates into k intervals $[X_0, X_1), [X_1, X_2), \ldots, [X_{k-1}, X_k]$ with $X_0 < X_1, \ldots, X_k$. Define N_i to be the number of spike events that are detected across the time series while $X_t \in [X_{i-1}, X_i)$. Additionally define T_i to be the amount of time for which $X_t \in [X_{i-1}, X_i)$.

Formally

$$N_i = \sum_{j=1}^{N} \mathbb{1}\{X_t \in [X_{i-1}, X_i)\} \qquad [27]$$

$$T_i = \int_0^T \mathbb{1}\{X_t \in [X_{i-1}, X_i)\}\,dt \qquad [28]$$

The log-likelihood function is then

$$\ell(\lambda_1, \ldots, \lambda_k) = \sum_{i=1}^{k} N_i \log(T_i) - T_i \lambda_i \qquad [29]$$

which can be maximised by maximising each component of the sum. This gives the MLE

$$(\hat{\lambda}_1, \ldots, \hat{\lambda}_k) = \left(\frac{N_1}{T_1}, \ldots, \frac{N_k}{T_k}\right) \qquad [30]$$

As the case of a homogeneous Poisson process is a restriction of this more complex model, the more complex model will always provide a 'better' fit to the data. We wish to test if there is a significant relationship between the intensity of the Poisson process and the heart rate by determining if this more complex models provides a statistically significantly better fit.

Remark. In theory an arbitrary number of intervals could be included at physically unrealistic heart rates that would never be attained. To therefore avoid increasing k unnecessarily and encountering issues in the definition of MLEs we assume that all $T_i > 0$ and discount/merge any intervals for which this condition fails.

The test statistic is given by

$$D = 2\big[\ell(\hat{\lambda}) - \ell_0(\hat{\lambda}_0)\big]$$
$$= 2\left[\sum_{i=1}^{k}\left(N_i \log\left(\frac{N_i}{T_i}\right) - 1\right) - N\left(\log\left(\frac{N}{T}\right) - 1\right)\right]$$

This is compared against a chi-squared distribution with $k-1$ degrees of freedom to test for significance.

S4. Spike density error

Some basic properties of the spike density error:

- As it is the integral of a sum of non-negative functions it is always non-negative.

- As there is no theoretical limit on the size or number of spikes the error is unbounded.

- Errors increase if relatively larger spikes are missed or erroneously detected.

We now present a proposition that demonstrates the relationship between accuracy and error behaves as expected.

Proposition 1. Consider a situation in which there is only one true spike and one detected spike of equal size (assumed to be 1 for simplicity) with the detected spike a distance h from the true spike. Let s be the time of the true spike and $s+h$ the time of the detected spike. Assume the density kernel f has finite support and that the location of the spikes are sufficiently far from the beginning or end of the time series that the supports of f_s and f_{s+h} lie entirely within $[0, T]$. In

this situation the spike density error decreases monotonically as h decreases and is given by

$$\epsilon(h) = \frac{2}{T} \int_{-\frac{h}{2}}^{\frac{h}{2}} f(t)\,dt = \frac{4}{T} \int_0^{\frac{h}{2}} f(t)\,dt \qquad [31]$$

Proof. Assume $h > 0$. Recall the definition of the shifted and scaled density kernel (Eqn.6)

$$f_s(t) = \frac{f(t-s)}{\int_0^T f(\tau - s)\,d\tau}$$

Using assumptions about the spikes and density kernel this reduces to

$$f_s(t) = f(t-s)$$

Then define

$$g(t) = f_s(t) - f_{s+h}(t)$$
$$= f_s(t) - f_s(t-h)$$

As the density kernel f is increasing below 0,

$$f_s(t) \geq f_s(t-h)\,\forall\,t < s$$

and so

$$g(t) \geq 0\,\forall\,t < s.$$

Similarly since f is decreasing above 0,

$$g(t) \leq 0\,\forall\,t > s+h.$$

In the region $s < t < s+h$, $f_s(t)$ is decreasing and $f_{s+h}(t)$ is increasing, hence there exists a point $t^* \in (s, s+h)$ such that $g(t^*) = 0$, $g(t) \geq 0\,\forall\,t < t^*$ and $g(t) \leq 0\,\forall\,t > t^*$. As f is not required to be strictly increasing, such a point may not be unique. Since

$$f_s\left(s + \frac{h}{2}\right) = f_s\left(s - \frac{h}{2}\right)$$
$$= f_{s+h}\left(s - \frac{h}{2} - h\right)$$
$$= f_{s+h}\left(s + \frac{h}{2}\right)$$

$t^* = s + \frac{h}{2}$ is a suitable point.

Recall the definition of the spike density error (Eqn.8)

$$\epsilon = \frac{1}{T}\int_0^T \left|\rho_r(t) - \rho_d(t)\right|\,dt$$

We have that

$$\left|\rho_r(t) - \rho_d(t)\right| = \begin{cases} f_s(t) - f_{s+h}(t) & \text{if } t < t^* \\ 0 & \text{if } t = t^* \\ f_{s+h}(t) - f_s(t) & \text{if } t > t^* \end{cases}$$

hence

$$\epsilon = \frac{1}{T} \int_0^T \left| \rho_r(t) - \rho_d(t) \right| dt$$

$$= \frac{1}{T} \int_0^{t^*} f_s(t) - f_{s+h}(t)\, dt + \frac{1}{T} \int_{t^*}^T f_{s+h}(t) - f_s(t)\, dt$$

$$= \frac{1}{T} \int_0^{t^*} f_s(t)\, dt - \frac{1}{T} \int_0^{t^*} f_s(t - h)\, dt$$

$$\quad + \frac{1}{T} \int_{t^*}^T f_s(t - h)\, dt - \frac{1}{T} \int_{t^*}^T f_s(t)\, dt$$

$$= \frac{1}{T} \int_0^{t^*} f_s(t)\, dt - \frac{1}{T} \int_0^{t^* - h} f_s(t)\, dt$$

$$\quad + \frac{1}{T} \int_{t^* - h}^T f_s(t)\, dt - \frac{1}{T} \int_{t^*}^T f_s(t)\, dt$$

$$= \frac{2}{T} \int_{t^* - h}^{t^* + h} f_s(t)\, dt$$

Using $t^* = s + \frac{h}{2}$ then gives Eqn.31. As $f \geq 0$ decreasing h decreases the error. The same argument can be made for $h < 0$ using the symmetry of f.

$\square$

S5. Distribution of spike in heart rate

Fig. S5.1. Histogram showing the distribution of heart rate where spikes were detected in the real heart rate data from Crickles. In order to account for different numbers of activities and spikes for different athletes, a separate histogram was generated for each athlete, and these histograms were then averaged to give the histogram here. The distribution shows a normal distribution-like curve, where the majority of the spikes were detected between ranges $105 - 140$ bpm.
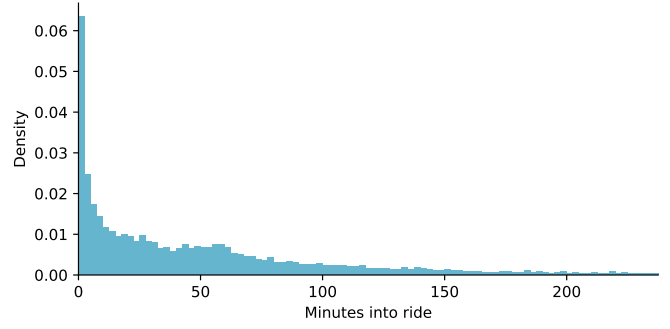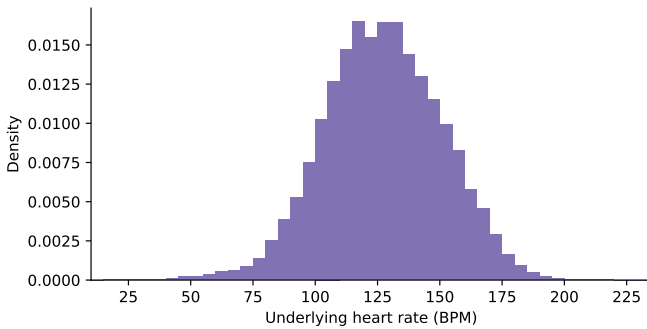
S6. Distribution of spike times

Fig. S6.1. Histogram showing the distribution of times at which spikes were detected in the real heart rate data from Crickles. In order to account for different numbers of activities and spikes for different athletes, a separate histogram was generated for each athlete, and these histograms were then averaged to give the histogram here. This shows that the majority of spikes were detected within the first 10 minutes of the time series, which may be due to do the failure of sports devices to record an accurate heart rate during the start of an activity as discussed in section 4.