RESEARCH

Real-time changepoint detection for noisy biological signals

Nathan Gold[1*†], Martin G Frasch[2†], Christophe Herry[3], Bryan S Richardson[4], Andrew J E Seely[5] and Xiaogang Wang[1]

[*]Correspondence: ngold5@my.yorku.ca
[1] Department of Mathematics and Statistics, York University, 4700 Keele St., M3J 1P3, Toronto, Canada
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

Background: Experimentally and clinically collected time series data are often contaminated with significant confounding noise, creating short, non-stationary time series. This noise, due to natural variability and measurement error, poses a challenge to conventional changepoint detection methods.

Results: We proposed a novel, real-time changepoint detection method for effectively extracting important time points in non-stationary, noisy time series. We validated our method with three simulated time series, a widely benchmark data set, two geological time series, and two physiological data sets, and compared it to three existing methods. Our method demonstrated significantly improved performance over existing methods in several cases.

Conclusion: Our method is able to effectively discern meaningful changes from system noise and natural fluctuations, accurately detecting change points. The ability of the method to extract meaningful change points with minimal user interaction opens new possibilities in clinical monitoring dealing with Big Data.

Keywords: Machine learning; Changepoint detection; Non-stationary noisy time series

Background

Various biological and medical settings require constant monitoring, collecting massive volumes of data in time series typically containing confounding noise [1, 2, 3]. This noise, as well as natural fluctuations in the biological system, create non-stationary time series, known as *piecewise locally stationary time series*, which are difficult to analyze in real time [4, 5, 6, 7]. Immensely important in clinical and

experimental decision making is the accurate and timely detection of pathological changes in the observed time series as they occur. Statistically, this may be interpreted as a changepoint detection problem for piecewise locally stationary time series [8, 9, 10].

The heavy contamination of noise due to measurement error and naturally varying phenomena, however, make the detection of changepoints challenging, as existing techniques will often observe non-pathological changes, resulting in false-alarms and mistrust of detection techniques [11, 12, 13] - the so-called *cry-wolf* effect. Extracting meaningful changepoints from naturally occurrig fluctuations and noisy corruptions remains a challenging research problem. With this setting in mind, we proposed a novel real-time changepoint detection method to extract a meaningful changepoint from other changepoints, which may occur due to noise and natural signal variability. Our method operates autonomously on each individual data steam, providing clinical information in real-time to assist in decision making.

While existing methods [8, 14, 15, 16, 17], can determine the location of changepoints, they are not able to extract meaningful changepoints in time series with piecewise locally stationary structure that contain minor changes and large noise corruption. We propose a novel method, termed the *Delta point* method, extending the Bayesian online changepoint detection method of Adams and MacKay [17] and later Turner [18], to allow a meaningful changepoint to be extracted. Our method uses fixed time intervals to construct the joint distribution of the number of changepoints per interval, and the average length of time between changepoints in each interval. The Delta point method is computationally efficient, and rapidly scales to large data sets. We demonstrate the effectiveness of the Delta point method on three simulated time series of our own design inspired by existing literature, the widely used Donoho-Johnstone Benchmark curves [19], nuclear magnetic resonance recordings from well-log measurements [20],annual lowest water levels of the Nile River [21], ECG recordings from clinical setting [22],and an experimental data set of recordings of fetal sheep heart rate variability during experiments mimicking human labour [23, 24, 25]. The fetal sheep data set contains short time series with large amounts of measurement noise, while the ECG dataset consists of short time series, with varying features.

Methods

Bayesian online change point detection

We begin with a review of the Bayesian online changepoint detection (BOCPD) algorithm [17]. BOCPD uses a combination of a predictive model of future observations of the time series, an integer quantity, r_t, known as the *run length*, or the time since the last changepoint, and a hazard function $p(r_t|r_{t-1})$, which calculates the probability of a changepoint occurring with respect to the last changepoint, to calculate the probability of a changepoint occurring. Bayes' rule is used to compute the posterior or past distribution of the run length as

$$p(r_t|\mathbf{y}_{1:t}) = \frac{p(r_t, \mathbf{y}_{1:t})}{p(\mathbf{y}_{1:t})} \tag{1}$$

where $\mathbf{y}_{1:t}$ is a vector of past observations of the time series, $p(r_t, \mathbf{y}_{1:t})$ is the joint likelihood of the cumulative run length and observations, calculated at each step, and $p(\mathbf{y}_{1:t})$ is the marginal likelihood of the observations. The joint distribution $p(r_t, \mathbf{y}_{1:t})$ is computed with each new observation using a message passing algorithm,

$$p(r_t, \mathbf{y}_{1:t}) = \sum_{r_{t-1}} p(r_t|r_{t-1}) p(y_t|r_{t-1}, \mathbf{y}_{(r)}) p(r_{t-1}, \mathbf{y}_{1:t-1}) \tag{2}$$

where $p(r_t|r_{t-1})$ is the hazard function, $p(y_t|r_{t-1}, \mathbf{y}_{(r)})$ is the prediction model with observations $\mathbf{y}_{(r)}$ since the last changepoint, and $p(r_{t-1}, \mathbf{y}_{1:t-1})$ is the previous iteration of the algorithm.

Gaussian processes

We implemented a recently developed predictive model using Gaussian processes (GP) [18]. A GP is a Gaussian distribution over functions - that is, the distribution of the possible values of the function follows a multivariate Gaussian distribution [26]. GPs are flexible and expressive priors over functions, allowing patterns and features to be learned from observed data. A GP is completely specified by a mean function, $\mu(\mathbf{x})$, and a positive definite covariance function, or *kernel*, $k(\mathbf{x}, \mathbf{x}')$, where $\mathbf{x}$ and $\mathbf{x}'$ are different observations. The covariance function generates properties of the function drawn from the GP, such as smoothness and shape. In our work, we make use of the rational quadratic covariance function,

$$k_{RQ}(\mathbf{x}, \mathbf{x}') = 1 + \frac{||\mathbf{x} - \mathbf{x}'||_2^2}{2\alpha\ell} \tag{3}$$

where α is a computing parameter, and ℓ is the input scale, or distance between inputs $\mathbf{x}$ and $\mathbf{x}'$, which determines the smoothness of the functions.

The observed inputs are collected into a matrix $\mathbf{X}$, and the covariance function is used to generate a covariance matrix $\mathbf{K}$. Using the regression model, $y = f(\mathbf{x}) + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma_t^2)$, and f is generated by a GP with mean function $\boldsymbol{\mu}$ and covariance function k, $f \sim GP(\mu, k)$, the distribution specified over functions is the multivariate Gaussian,

$$p(\mathbf{f}|\mathbf{X}) = N(\mathbf{f}|\boldsymbol{\mu}, \mathbf{K}) = \frac{1}{\sqrt{2\pi \det \mathbf{K}}} \exp\left\{\frac{1}{2}(\mathbf{f} - \boldsymbol{\mu})^\top \mathbf{K}^{-1}(\mathbf{f} - \boldsymbol{\mu})\right\} \tag{4}$$

To predict future observations of the time series, we appeal to Bayes' rule, using the GP as a prior over functions. By the rules of conditioning Gaussian distributions [26], the predictive distribution for our next function value f_* given input $\mathbf{x}_*$ is

$$p(f_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = N(f_*|\mu_*, \sigma_*^2) \tag{5}$$

$$\mu_* = \mathbf{K}_*^\top \mathbf{K}^{-1}\mathbf{y} \tag{6}$$

$$\sigma_*^2 = \mathbf{K}_{**} - \mathbf{K}_*^\top \mathbf{K}^{-1}\mathbf{K}_* \tag{7}$$

where $\mathbf{K}_*$ is the covariance matrix between the new input $\mathbf{x}_*$ and past inputs $\mathbf{X}$, $k(\mathbf{x}_*, \mathbf{X})$, and $\mathbf{K}_{**}$ is the covariance between the new inputs, $k(\mathbf{x}_*, \mathbf{x}_*)$.

Delta point method

Given a vector of suspected changepoints, we view each observed changepoint as a realization of a point process. We construct a user-defined, domain specific interval of fixed length $[t_i, t_j]$, where $i < j$ and introduce a new variable $c = c_{[t_i,t_j]}$, the number of changepoints in the interval $[t_i, t_j]$. The average run length $\bar{r}$ is calculated for each interval $[t_i, t_j]$ by

$$\bar{r} = \frac{1}{c}\sum_{k=1}^{c} r_k \tag{8}$$

where r_k is the run length associated with each changepoint in the interval. We then consider the joint distribution of c and $\bar{r}$, and condition $\bar{r}$ by c, yielding, $p(\bar{r}, c) = p(\bar{r}|c)p(c)$. The interval with the longest average run length and fewest

number of changepoints is selected as the interval with the highest probability of containing the changepoint of interest. The selected interval is then searched, and the changepoint with the longest associated run length r_k is chosen. This chosen changepoint is declared the *Delta point*, or the changepoint of interest.

Statistical analysis

To compare the Delta point method to competing methods, we performed several statistical tests on declared changepoints from each method. For the simulated data, we computed the mean square error (MSE) for each method, taking the absolute temporal difference between the user labelled changepoint and the declared changepoint. We then performed two-sided t-tests to compare the mean absolute detection times of each method to the Delta point method, with the null hypothesis that the mean detection times of other methods will not differ from the Delta point method's time. For the clinical ECG recording data set, we computed the absolute differences in detection times for each method between the user-labelled changepoints, and the MSE for each method. We also performed two-sided t-tests to compare the mean absolute detection times between methods with the null hypothesis that the mean detection times of other methods will not differ from the Delta point method's time. For the experimental data sets, we performed a Fisher's exact test to compare successful changepoint detection for each method. In addition, we generated Bland-Altman plots to compare the accuracy of each method to the user defined changepoint of interest and the declared changepoint.

Results

We tested the Delta point method on several simulated and real world time series data sets. The simulated time series consist of three synthetic time series of our own design, and two widely used benchmark curves. The real world data sets are made up of well-log recordings from geophysical drilling measurements, annual water levels of the Nile river, 100 clinical ECG recordings, and an experimental data set of 14 fetal sheep recordings mimicking human labour. We compared the Delta point method to three competing non-stationary change point detection algorithms, namely Takeuchi and Yamanishi (TY) [15], Last and Shumway (LS) [16], and Liu et al (L) [27].

Simulation

To test the efficacy of the Delta point method, we produced 1000 simulations of three different time series, each 500 data points in length. Each time series was designed to simulate changepoints that may be seen in real world settings, and to have a specific changepoint that is of more interest than others in the time series.

Series 1 has two changepoints, with the changepoint of interest occurring at $t = 150$. This time series simulated the change from a scaled random walk to an autoregressive model, and then back to a scaled random walk. This is a relatively subtle changepoint to detect, and was inspired from [16]. Series 1 is given as,

$$X_t = \begin{cases} \epsilon_t/\sqrt{3}, & \text{if } 1 \leq t < 150 \text{ and } 350 < t \leq 500 \\ 0.7X_{t-1} + \epsilon_t, & \text{if } 150 \leq t \leq 350, \end{cases} \tag{9}$$

where $\epsilon_t \sim N(0, 1)$, is Gaussian white noise.

Series 2 is a simulated autoregressive model with a large jump, and then a return back to the original process. The changepoint of interest was chosen as the onset of the jump ($t = 175$). This time series was used to simulate sensor shocks or faults, or a change in the generative parameters of the time series distribution. Series 2 is given as,

$$X_t = \begin{cases} 0.4X_{t-1} + \epsilon_t, & \text{if } 1 \leq t < 175 \text{ and } 325 < t \leq 500 \\ 0.4X_{t-1} + 3 + \epsilon_t, & \text{if } 175 \leq t \leq 325, \end{cases} \tag{10}$$

where, $\epsilon_t \sim N(0, 1)$, is Gaussian white noise.

Series 3 is a simulated autoregressive moving average model with an introduced linear trend and subsequent return to the autoregressive moving average model. The changepoint of interest was chosen as the beginning of the linear trend ($t = 225$). This is a difficult changepoint to detect, as the noise added to the data obscured the introduction of the trend. This time series simulated the accumulation of some product in a system. Series 3 is given as,

$$X_t = \begin{cases} 0.5X_{t-1} + \epsilon_t + 0.5\epsilon_{t-1}, & \text{if } 1 \leq t < 225 \\ 0.5X_{t-1} + \epsilon_t + 0.5\epsilon_{t-1} + 0.02t - 4, & \text{if } 225 \leq t \leq 375 \\ 0.5X_{t-1} + \epsilon_t + 0.5\epsilon_{t-1} + 4, & \text{if } 375 \leq t \leq 500, \end{cases} \tag{11}$$

where, $\epsilon_t \sim N(0,1)$, is Gaussian white noise.

The BOCPD method learned parameter values through training, so we only list the values we used to initiate the method. For Series 1, Series 2, and Series 3, we used a training set of 200 data, taken at the beginning of the time series. The Gaussian process model used a non-biased parameter initialization, with an assumed standard normal distribution prior for Series 1, Series 2, and Series 3. The hazard rate parameter used for the hazard function for initial training for each time Series is $\theta_h = -3.982$. The Delta point interval length for each time series was set at 40 for training, as this should protect against the BOCPD possibly declaring too many erroneous changepoints. The techniques TY, LS, and L all require a threshold value above which a changepoint will be declared. We performed cross-validation of several threshold values for each method, choosing the value for each time series that allows the most accurate detection of the changepoint of interest. We select the changepoint declared by each method that is closest to the significant changepoint described above.

The results of each method are displayed in Table 1. For Series 1, the Delta point method significantly ($p < 0.001$) outperformed methods TY and LS in the mean absolute difference (abs. diff.) of detection time, and had a significantly lower MSE. This difference in performance is confirmed by a two sided t-test with a null hypothesis that other methods do not have a significant mean abs. diff. from the Delta point method. In our simulations, method L had a slightly smaller mean abs. diff. (8.953) compared to the Delta point method (9.718), however the distribution of declared changepoints had a larger standard deviation (12.783 compared to 9.881). As well, the Delta point method had a lower MSE. The two sided t-test confirmed that both methods had indistinguishable performance ($p = 0.135$). For Series 2, the Delta point method significantly ($p < 0.001$) outperformed all methods. For Series 3, method TY performed the best, with the lowest mean absolute difference. Figures 1-3 display boxplots of the results of each method for Series 1-3, respectively.

Donoho-Johnstone Benchmark

To further analyze the performance of the Delta point method, we tested it and the existing methods on the Donoho-Johnstone Benchmark non-stationary time series [19]. The Donoho-Johnstone Benchmark is a classic collection of four non-

stationary time series designed as a test for neural network curve fitting. The curves are known as the Block, Bump, Doppler, and Heavisine, and are 2048 data points in length each. We adapted the curves with the introduction of noise to test for changepoint detection. As the Delta point method is not designed to function with time varying periodic data, rather piecewise locally stationary time series, we did not test with the Doppler and Heavisine curves.

For training for the Delta point method, we used a standard normal distribution prior for the Gaussian process, and hazard rate parameter $\theta_h = -3.982$. We set the Delta point interval to 10 for the Bump curve due to its high variability, and 50 for the Block curve, as this curve had more well-defined changes. The training set consisted of the first 800 data of each curve, to correspond to the rule of thumb of using the first 35% to 40% of time series data for training [18].

The results of the methods for the Bump and Block curves are displayed in Table 2. The Delta point method performed very well in these cases, declaring the changepoint of interest very close to the user-labelled point. The Bump curve was a more difficult curve to detect changepoints in, due to noise. The Delta point method and LS curve have about the same performance for the bump curve (abs. diff of 5 and 4, respectively). The Delta point method outperformed methods TY, LS, and L for accurate detection in the Block curve (abs. diff. 1 compared to 17, 22, and 24, respectively).

Well-log and Nile recordings

The well-log data set consists of 4050 nuclear magnetic resonance measurements obtained during the drilling of a well [20]. These data are used to deduce the geophysical structures of the rock surrounding the well. Variations in the mean reflect differences in stratification of the Earth's crust [17]. The well-log data set is a well studied time series for changepoint detection [20, 17, 18]. We selected the largest jump in the mean of the time series as the most significant change point ($i = 1070$).

The Nile river time series consists of a record of the lowest annual water levels between 622-1284 CE, recorded on the Island of Roda, near Cairo, Egypt [21]. The Nile river data set has been used extensively in changepoint detection [18], making it an effective benchmark for the Delta point method. Geophysical records suggest the

installation of the *Nilometer* in 715 CE, a primitive device for more accurate water level measurements. As such, we selected this as the changepoint of significance.

For training for the Delta point method for both time series, we used a standard normal distribution prior for the Gaussian process, and hazard rate parameter $\theta_h = -3.982$. For the well-log data, the Delta point interval was chosen as 30 due to the length of the time series and sensor noise, and for the Nile river time series, the Delta point interval was chosen to be 50, as the curve is smoother. The training set consisted of the first 1000 data for the well-log series, and first 250 data for the Nile river set.

The results of each method are displayed in Table 3. The Delta point method performed better than the other methods TY, LS, and L for the well-log data set (abs. diff. 2 compared to 13, 15, and 33, respectively). For the Nile river data set, all methods performed well, with declared changepoints of interest very close to the labelled installation of the Nilometer in 715 CE. The Delta point method for the well-log set is displayed in Figure 4 (Top Panel), and the Delta point method for the Nile river data set is displayed in Figure 4 (Bottom Panel).

ECG recordings

To determine the effectiveness of the Delta point method in detecting a significant changepoint in short time series, we tested each method's accuracy in detecting the QRS complex, and the difference between the labelled beginning and detected time. As each time series is short, and the QRS complex rapidly begins and ends in the recording, accurate detection of the changepoint was considered very important.

For training for the Delta point method, we used a standard normal distribution prior for the Gaussian process, and hazard rate parameter $\theta_h = -3.982$. Due to the short nature of these time series, the training set length was selected to be the first 30 data points; the training set never included the QRS complex for any of the 100 instances. The Delta point interval was set to 5, as the QRS complex is very short, and occurs rapidly in the series. The time series rapidly changes here, so a shorter interval performed best.

The performance of each method is displayed in Table 4. The Delta point method significantly outperformed the TY and LS method in mean abs. diff. from the labelled detection times of the complex ($p < 0.001$ CI $= [-2.138, -1.002]$ and

$[-.1339, -0.0611]$, respectively). Method L had indistinguishable performance from the Delta point method ($p = 0.954$ CI $= [-0.709, 0.609]$). The Delta point method had the lowest MSE of all methods (14.13 compared to 32.2, 26.31, and 22.73, respectively). A boxplot of the abs. diff. of all of the methods is displayed in Figure 5.

Fetal sheep model of human labour

We applied the Delta point method to a data set consisting of 14 experimental time series of a measure of fetal heart rate variability (HRV) known as the root mean square of successive differences (RMSSD) of R-R intervals of ECG recorded during umbilical cord occlusions (UCO) [28, 29]. The RMSSD may be used a measure to study the relationship between fetal systemic arterial blood pressure (ABP) and fetal heart rate in a fetal sheep model of human labour [23, 24, 25]. RMSSD is a sensitive measure of vagal modulation of HRV, and is known to increase with worsening acidemia, a dangerous condition that may occur during labour [28, 29, 30, 31].

During UCO mimicking human labour, a hypotensive blood pressure response to the occlusions manifests as the introduction of a new trend in the recorded time series. This response is induced by the vagal nerve activation triggered by worsening acidemia during UCO as discussed in [32]. These points are detected by *expert* visual detection and are known as ABP *sentinel points*. These sentinel points are defined as the time between the onset of blood pressure responses and the time when pH nadir (ph < 7.00) is reached. A change point detection algorithm should be able to detect these sentinel points from the non-invasively obtainable fetal heart rate derived RMSSD signal in an online manner to assist in clinical decision making.

The experimental time series are short - less than 200 observations - and confounded with a large amount of noise due to experimental conditions and measurement error. The time series are piecewise locally stationary, and contain naturally occurring biological fluctuations due, for example, to non-linear brain-body interactions [33]. These factors make the detection of the expert sentinel point difficult for existing change point techniques [1, 2].

To avoid false alarms, we defined a clinical *region of interest* (ROI) of 20 minutes before the sentinel point where a declared change point of interest is determined

to be a success. We also took into account detections that are at most 3 minutes posterior to the sentinel point, as this is one experimental cycle late. The defined region of interest is to assist clinicians in decision making, as it provides a feasible window of time to provide clinical evaluation, as well as reject false alarm. For training for the Delta point method, we used a standard normal distribution prior for the Gaussian process, and hazard rate parameter $\theta_h = -3.982$. We trained the Delta point method with 48 data points per time series, corresponding to two hours of recording. The Delta point interval was set at 10 data, which corresponded to 25 minutes of experiment time. This interval was chosen to coincide with the clinical ROI.

he Delta point method significantly outperformed competing methods, with 11 of 14 declared change points in the ROI, compared to 3 of 14 for TY with Fisher's exact test statistic 0.007028, 5 of 14 for LS with Fisher's exact test statistic 0.054238, and 2 of 14 for L with Fisher's exact test statistic 0.001838. The Delta point method applied to one animal from the data set (ID473378) is displayed in Figure 6, and the results and detection times of each method are shown in Table 5.

We also computed Bland-Altman plots for the experimental time series to compare the Delta point method to each other method. In Figure 7(A), we display the Bland-Altman plot for the Delta point method and TY with mean difference (6.93±89.03). Figure 7(B) displays the Bland-Altman plot for the Delta point method and LS, with mean difference (-1.36±62.6). Figure 7(C) displays the Bland-Altman plot for the Delta point method and L, with mean difference (14.4±59.9). In Figure 7(D), we display a modified Bland-Altman plot of the differences in detection times for each method, along with the upper and lower ROI.

Discussion

The detection of change points in a non-stationary time series is a well studied problem, which has produced many techniques [14, 8, 15, 17, 16]. Some of the first work in change point detection is due to [14], where changes were detected by comparing probability distributions of time series samples over past and present intervals. This work was extended by Takeuchi and Yamanishi (method TY) [15], who proposed a scoring procedure along with outlier detection, to compare past and present probability distributions in real-time. The scoring method is based on an

automatically updating autoregressive model, known as the sequentially discounting autoregressive (SDAR) model. Points with scores above a user-defined threshold are declared as change points. These scoring based methods, however, are unable to automatically extract a single change point of interest, unlike the Delta point method, as they require a user to observe all change points above the threshold.

Non-stationary time series may also be viewed as segments of piecewise locally stationary time series [8]. We follow this spirit in our work for the Delta point method. In [8], the locally stationary segments are broken into small pieces and the distance between power spectra for two adjacent pieces are calculated. A variety of distance measures such as the Kolmogorov-Smirnov distance looking at the distance between cumulative power spectra, and the Cramer-Von Mises distance between power spectra was used in [8]. An extension was proposed in [16] (method LS) where the Kullback-Liebler discrimination information between power spectra is used to identify change points. These power spectra methods are particularly well-suited for time series with periodic structure, as they compare power spectra in the frequency domain.

Recently, advances in real-time change point detection have been made in the machine learning community with promising results, such as relative density ratio estimation method [27] (method L). Relative density ratio estimation uses a divergence measure to estimate the divergence between subsequent time series samples' density ratios. The Delta point method does not require the need to estimate density ratios between time points, as it instead considers the joint distribution of declared change points and the time between each. By only considering the time between change points, the Delta point method is able to infer the location of the change point of interest.

With respect to our own testing and results, we observed that the Delta point method is effective at finding changepoints of interest in piecewise locally stationary time series of different types. For the simulated time series of our own design, the Delta point method performed better or indistinguishably from the best performing methods for Series 1 and Series 2. For Series 1, the Delta point method had the lowest MSE, which suggested it is accurately identifying changepoints of interest. For Series 2, the Delta point method significantly outperformed the competing methods in terms of mean absolute difference in detection time for labelled changepoints of

interest. Although method LS had a lower MSE for this series, its mean detection difference is closer to 0. In Series 3, method L performed the best, with the smallest mean absolute difference in detection time, and MSE. Series 3 consisted of the introduction of the linear trend to the autoregressive moving average model, of which the introduction of the trend was obscured by added noise. Since method L compares density ratios of the time series, its good performance on this time series is likely due to noticing these changing ratios before other methods noticed the trend.

For the Donoho-Johnstone Bump curve, the Delta point method performed nearly as well as the best performing method - method Ls - with a much smaller abs. diff. in detection time compared to the other methods, TY and L. The Delta point method performance for the Donaho-Johnstone Block curve was better than the other methods, exemplifying the strength of the Delta point method for piecewise locally stationary time series. Our test results for the well-log data set also provides evidence of the performance of the Delta point method for piecewise locally stationary time series. For the Nile river data set, as the installation of the Nilometer is the most significant changepoint in the time series, and can even be noticed visually, we expected that all methods should accurately detect this changepoint with little variation. Indeed, our results confirm this hypothesis.

For the clinical ECG data set, ECGFiveDays, the Delta point method performs significantly better than methods TY and LS, however has an indistinguishable performance difference with method L, although the Delta point method has the lowest MSE. Due to the rapidly varying nature of the time series when the QRS complex begins, the ability of method L to compare density ratios between components of the time series is beneficial and improves its performance compared to other methods.

With regards to the fetal sheep experimental data set, the early detection of acidemia is better than late detection from a clinical perspective. Hence, we defined the clinical ROI according to expert physician input. The 20 minute window before the expert-labelled sentinel point provides adequate warning to clinicians to increase monitoring, or expedite delivery, while the 3 minute window posterior to the expert-labelled sentinel point is sufficiently close to be included in the experimental procedure. In clinical settings, we believe that earlier detection is better, as it provides longer decision making time, and justification for increased monitoring.

The novelty of the current work is that our method permits statistical-level predictions about concomitant changes in individual bivariate time series, simulated or physiological such as HRV measure RMSSD and ABP in an animal model of human labour. Our method is able to predict cardiovascular decompensation by identifying ABP responses to UCO, a sensitive measure of acidosis. These predictions are reliable even in the instances when the signals are noisy. This is based on our observation that here, to mimic the online recording situation, no artefact correction for RMSSD was undertaken as is usually done for HRV offline processing [34]. The two hour training time used for the Delta point method is also acceptable for delivery room settings, due to the typical time length of human labour between 6 to 8 hours on average ([35]). To our knowledge, no comparable statistical methods exist. Another benefit of the Delta point method is the ability to automatically extract the change point of interest with minimal user interaction, as opposed to other methods which require user specific thresholds and criteria.

Conclusion

We have developed a novel, real-time change point detection method for effectively isolating a change point of interest in short, noisy, piecewise locally stationary time series. Our method is able to effectively extract clinically relevant changes in real-time, allowing informed decision making [34, 36]. By considering the joint distribution of the between change points and the number of change points in disjoint intervals, the Delta point method remains robust to signal artifacts and confounding noise. We demonstrated our method on three simulated time series of our own design inspired by existing literature, curves from the Donoho-Johnstone benchmark curve data set, nuclear magnetic resonance reading from well-log measurements of geophysical drilling, annual water levels of the Nile river, a clinical ECG recording data set, and an experimental data set of fetal sheep simulating human labour. We compared the performance of the Delta point method to three existing changepoint detection methods. The Delta point method displays useful performance benefits in accurately extracting a meaningful change point automatically, with the need for minimal user-interaction.

Competing interests

BSR and MGF are inventors of related patent applications entitled "EEG Monitor of Fetal Health" including U.S. Patent Application Serial No. 12/532.874 and CA 2681926 National Stage Entires of PCT/CA08/0058 filed March 28, 2008, with priority to US provisional patent application 60/908,587, filed March 28, 2007 (US 9,215,999). No other disclosures have been made.

Author's contributions

XW, BSR, and MGF conceived of and designed the study. MGF, BSR, and NG acquired the datasets. MGF, NG, CH, and XW interpreted the data. NG implemented the algorithm. MGF, NG, CH, and XW drafted the manuscript. MGF, NG, CH, BSR, AJES, and XW revised the manuscript for intellectual content and approved the final transcript.

Acknowledgements

This work was funded by CIHR (to MGF and AJES), FRQS (to MGF), Canada Research Chair Tier 1 in Fetal and Neonatal Health and Development (BSR); Women's Development Council, London Health Sciences Centre (London, ON, Canada) (to BSR and MGF).

The authors gratefully acknowledge technical support by Brad J. Matushewski, Dr. L. Daniel Durosier, Dr. Qiming Wang, and Dr. Michael Last. We would also like to thank Ms. Patrycja Jankowski for excellent artwork assistance.

Author details

[1] Department of Mathematics and Statistics, York University, 4700 Keele St., M3J 1P3, Toronto, Canada. [2]Department of Obstetrics and Gynecology, University of Washington, Seattle, WA, USA, 1959 NE Pacific St, Box 356460, 98195 Seattle, USA. [3] Dynamical Analysis Laboratory, Ottawa Hospital Reserach Insitute, Clinical Epidemiology Program, Campus Room 6371, 501 Smyth Road - Box 708, K1H 8L6, Ottawa, Canada. [4] Department of Obstetrics and Gynecology, London Health Sciences Centre, Victoria Hospital, 800 Commissioners Road E., N6A 5W9, London, Canada. [5] Dynamical Analysis Laboratory, Divisions of Thoracic Surgery and Critical Care Medicine, Ottawa Hospital Research Institute, University of Ottawa, 501 Smyth Road - Box 708, K1H 8L6, Ottawa, Canada.

References

1. Barahona, M., Poon, C.-S.: Detection of nonlinear dynamics in short, noisy time series. Nature **381**, 215–217 (1996)
2. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Physica D **9**(1-2), 189–208 (1983)
3. Sugihara, G., May, R.M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature **344**, 734–741 (1990)
4. Bodenstein, G., Praetorius, H.M.: Feature extraction from electroencephalogram by adaptive segmentation. Proceedings of the IEEE **65**(5), 642–652 (1977)
5. Corge, J., Puech, F.: Analyse du rythme cardiaque foetal par des methodes de detection de ruptures. In: 7th INRIA Int. Conf. Analysis and Optimization of System, Antibes, FR (1986)
6. Dowse, H.B.: Statistical analysis of biological rhythm data. In: Rosato, E. (ed.) Circadian Rhythms: Methods and Protocols vol. 362, 1st edn., pp. 29–45. Humana Press, Totowa, NJ (2007)
7. Ndukum, J., Fonseca, L.L., Voit, E.O., Datta, S.: Statistical inference methods for sparse biological time series data. BMC Systems Biology **5**(57) (2011)
8. Adak, S.: Time-dependent spectral analysis of nonstationary time series. Journal of American Statistical Association **93**, 1488–1501 (1998)
9. Ombao, H.C., Raz, J.A., Von Sachs, R., Malow, B.A.: Automatic statistical analysis of bivariate nonstationary time series, in memory of jonathan a. raz. Journal of Americal Statistical Association **96**, 543–560 (2001)
10. Davis, R.A., Lee, T.C.M., Rodriguez-Yam, G.A.: Structural break estimation for nonstationary time series models. Journal of American Statistical Association **11** (2006)
11. Beneken, J.E., Van der Aa, J.J.: Alarms and their limits in monitoring. J Clin Monit **5**(3), 205–210 (1989)
12. Lawless, S.T.: Crying wolf: false alarms in a pediatric intensive care unit. Crit. Care Med. **22**(6), 981–986 (1994)
13. O'Carrol, T.M.: Survery of alarms in an intensive therapy unit. Anaesthesia **41**(7), 742–745 (1986)

14. Basseville, M., Nikiforov, I.V.: Detection of Abrupt Changes: Theory and Application. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1993)
15. Takeuchi, J., Yamanishi, K.: A unifying framework for detecting outliers and change points from time series. IEEE Transactions on Knowledge and Data Engineering **18**(4), 482–492 (2006)
16. Last, M., Shumway, R.: Detecting abrupt changes in a piecewise locally stationary time series. J Multivar Anal. **99**(2), 191–214 (2008)
17. Adams, R.P., MacKay, D.J.C.: Bayesian online changepoint detection. arXiv:0710.3742v1 (2007)
18. Turner, R.D.: Gaussian processes for state space models and change point detection. PhD thesis, University of Cambridge, Engineering Department (2011)
19. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation via wavelet shrinkage. Biometrika **81**, 425–455 (1994)
20. Ó Ruanaidh, J.J., Fitzgerald, W.J., Pope, K.J.: Recursive bayesain location of a discontinuity in time series. In: 4th IEEE International Conference on Acoustics, Speech, and Signal Processing, Adelaide, SA, Australia (1994)
21. Beran, J.: Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability, vol. 61. Chapman & Hall
22. Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G.: The UCR Time Series Classification Archive (2015)
23. Wang, X., Durosier, L.D., Ross, M.G., Richardson, B.S., Frasch, M.G.: Online detection of fetal acidemia during labour by testing synchronization of eeg and heart rate: A prospective study in fetal sheep. PLoS ONE **9**(9) (2014)
24. Ross, M.G., Jessie, M., Amaya, K., Matushewski, B., Durosier, L.D., Frasch, M.G., Richardson, B.S.: Correlation of arterial fetal base deficit and lactate changes with severity of variable heart rate decelerations in the near-term ovine fetus. Am J Obstet Gynecol (205), 281–286 (2013)
25. Frasch, M.G., Mansano, R.Z., McPhaul, L., Richardson, B.S., Ross, M.G.: Measures of acidosis with repetitive umbilical cord occlusions leading to fetal asphxyia in the near-term ovine fetus. Am J Obstet Gynecol (200), 201–207 (2009)
26. Rasmussen, C.E., William, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press, Cambridge, MA (2006)
27. Liu, S., Yamada, M., Collier, N., Sugiyama, M.: Change-point detection in time-series data by relative density-ratio estimation. Neural Networks **43**, 72–83 (2013)
28. Frasch, M.G., Muller, T., Weiss, C., Lohle, M., Schwab, K., Schubert, H., Nathanielsz, P.W., Witte, O.W., Schwab, M.: Fetal body weight and the development of the control of the cardiovascular system in fetal sheep. J Physiol (579), 893–907 (2007)
29. Frasch, M.G., Müller, T., Weiss, C., Schwab, K., Schubert, H., Schwab, M.: Heart rate variability analysis allows early asphyxia detection in ovine fetus. Reproductive Science **16**(5), 509–517 (2009)
30. Xu, A., Durosier, L.D., Ross, M.G., Hammond, R., Richardson, B.S., Frasch, M.G.: Adaptive brain shut-down counteracts neuroinflammation in the near-term ovine fetus. Front. Neurol. **5**(110) (2014)
31. Durosier, L.D., Green, G., Batkin, I., Seely, A.J., Ross, M.G., Richardson, B.S., Frasch, M.G.: Sampling rate of heart rate variability impacts the ability to detect acidemia in ovine fetuses near-term. Frontiers in pediatrics **2**, 38 (2014)
32. Frasch, M.G., Durosier, L.D., Gold, N., Cao, M., Matushewski, B., Keenliside, L., Louzoun, Y., Ross, M.G., Richardson, B.S.: Adaptive shut-down of eeg activity predicts critical acidemia in the near-term ovine fetus. Physiological Reports **3**(7), 12435 (2015)
33. Berntson, G., Nagaraja, H.K., Porges, S.W., Saul, P.: Heart rate variability: Origins, methods, and interpretive caveats. Psychophysiology **34**, 623–648 (1997)
34. Seely, A.J., Green, G.C., Bravi, A.: Continuous multiorgan varaibility monitoring in critically ill patients – complexity science at the bedside. In: IEEE Eng Med Biol Soc, Boston, MA (2011)
35. Albers, L.L.: The duration of labor in healthy women. Journal of Perinatology **19**(2), 114–119 (1999)
36. Seely, A.J.E.: Data intelligence is the future of monitoring. Journal of Clinical Monitoring and Computing **28**(4), 325–327 (2014)

Figures

Figure 1 Simulation Series 1 boxplot. Legend: TY = [15], LS = [16], L = [27]. Boxplot of absolute differences of detected changepoints for 1000 simulations of simulation data set Series 1. The true changepoint location is located at 0.

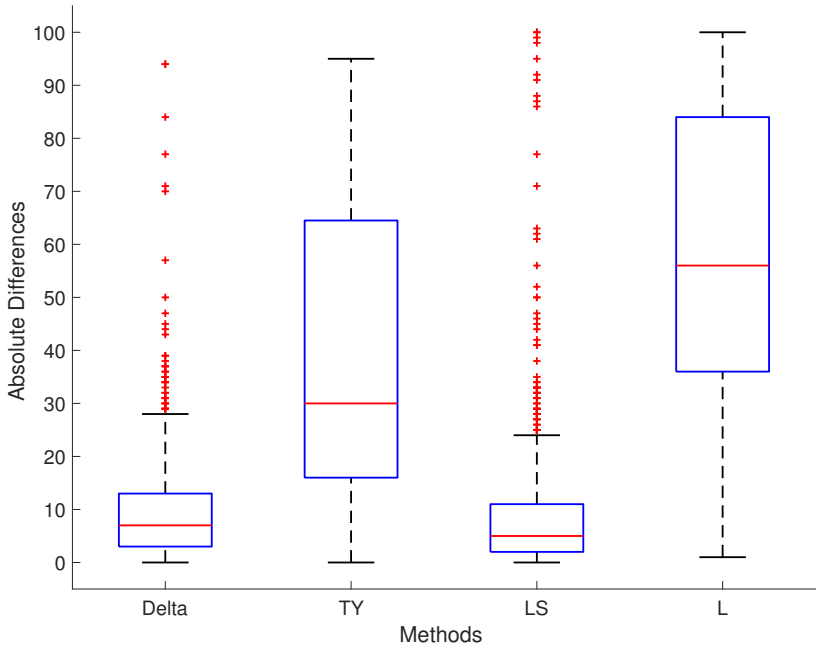


Figure 2 Simulation Series 2 boxplot. Legend: TY = [15], LS = [16], L = [27]. Boxplot of absolute differences of detected changepoints for 1000 simulations of simulation data set Series 2. The true changepoint location is located at 0.

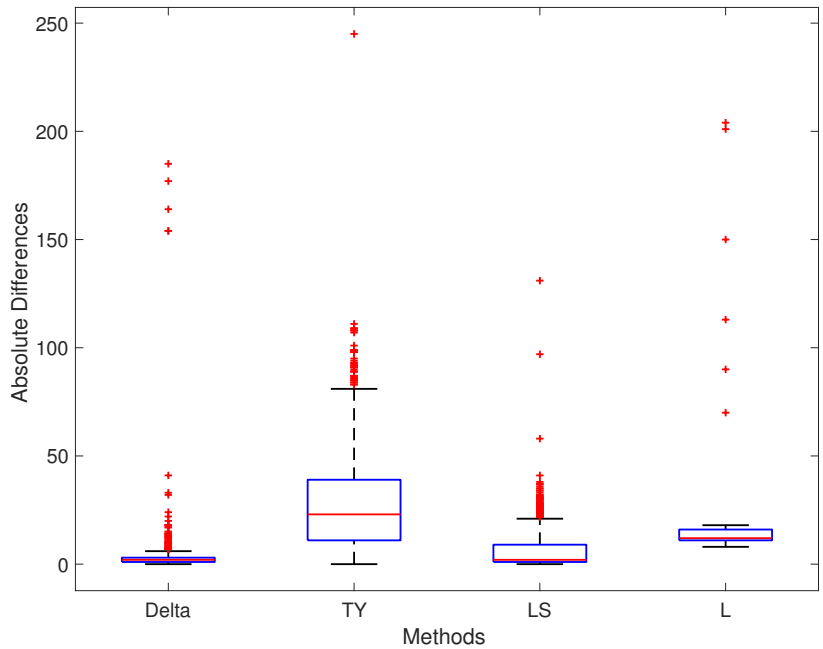


Figure 3 Simulation Series 3 boxplot. Legend: TY = [15], LS = [16], L = [27]. Boxplot of absolute differences of detected changepoints for 1000 simulations of simulation data set Series 3. The true changepoint location is located at 0.

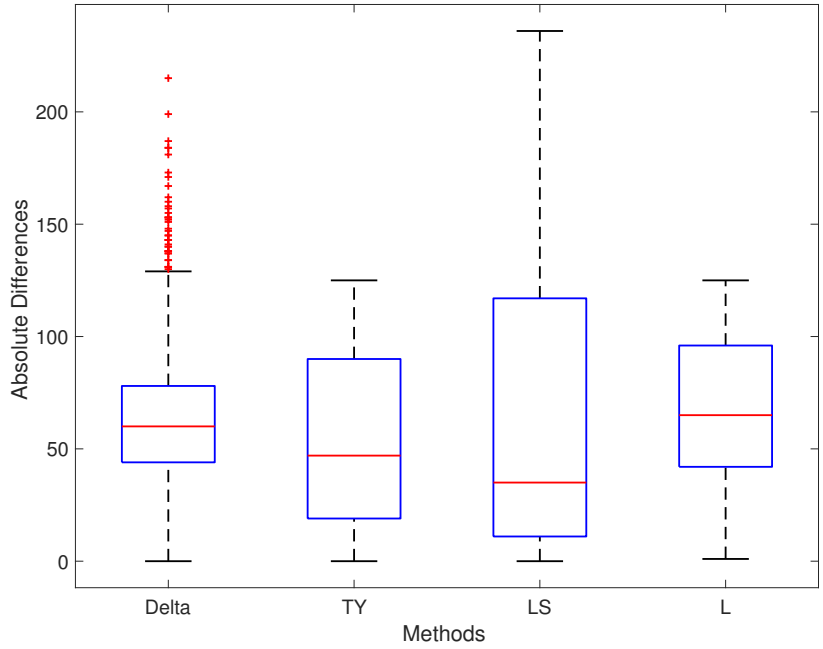


Figure 4 Well-log and Nile River level Delta point method. Top Panel: Delta point method for Well-log data set. The y-axis denotes NMR reading during well digging, and the x-axis denotes the measurement instance. Suspected changepoints are denoted with red crosses, the user-labelled changepoint with a black cross(1070), and the detected Delta point with an orange box(1072). Bottom Panel: Delta point method for annual lowest levels of Nile River. The y-axis denotes the water level (mm), and the x-axis denotes the years. Suspected changepoints are denoted with red crosses, the installation of the Nilometer(715) with a black cross, and the detected Delta point with an orange box (720).

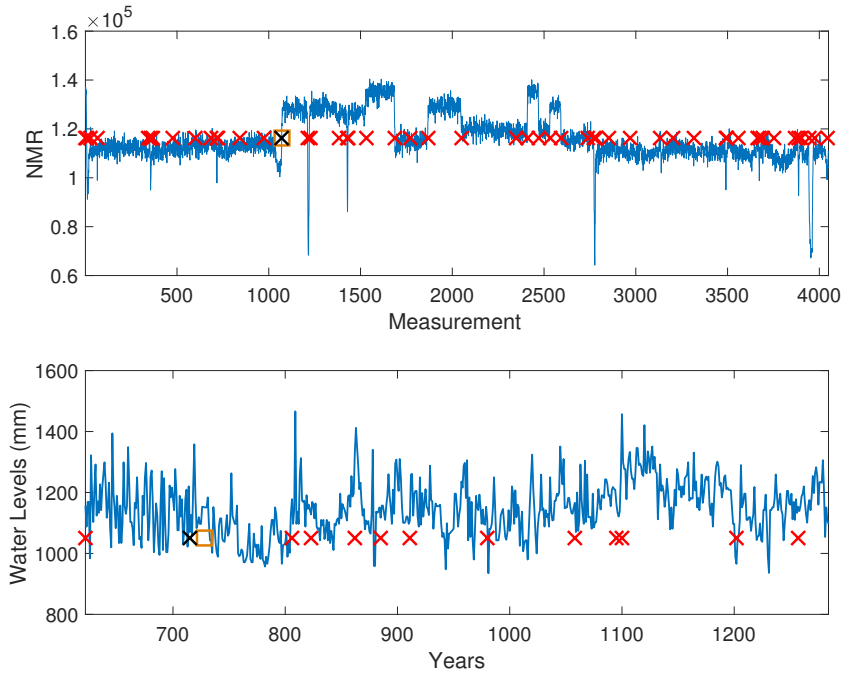


Figure 5 ECG (ECGFiveDays) boxplot. Legend: TY = [15], LS = [16], L = [27]. Boxplot of absolute differences of detected QRS complexes for 100 short time series of ECG recordings. The true changepoint is located at 0.

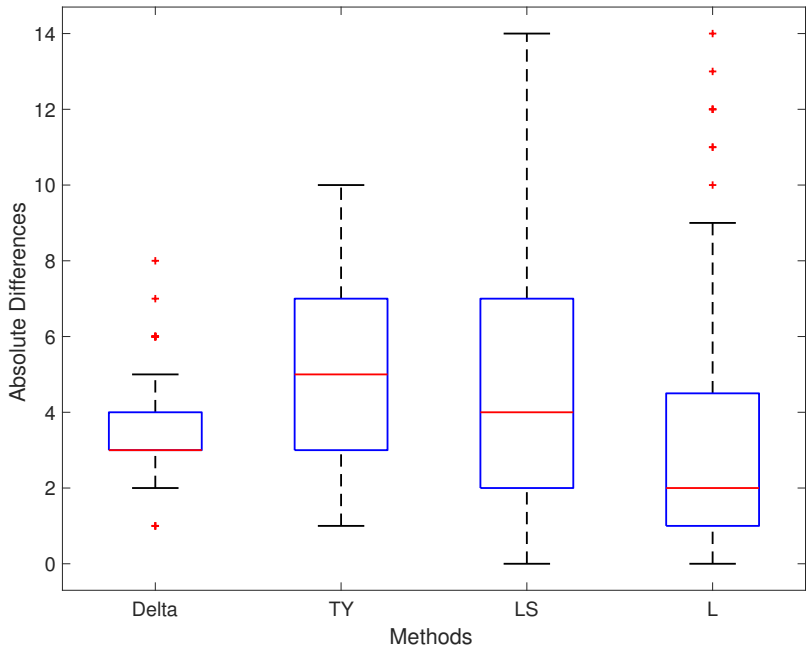


Figure 6 Fetal Sheep ID473378 Delta point method. Delta point method for Fetal Sheep ID473378 RMSSD time series. The y-axis denotes the RMSSD of the animal over the experimental course, and the x-axis denotes experimental time. Suspected changepoints are denoted with red crosses, the expert sentinel value with a black cross (6:13), and the detected Delta point with an orange box (6:24).

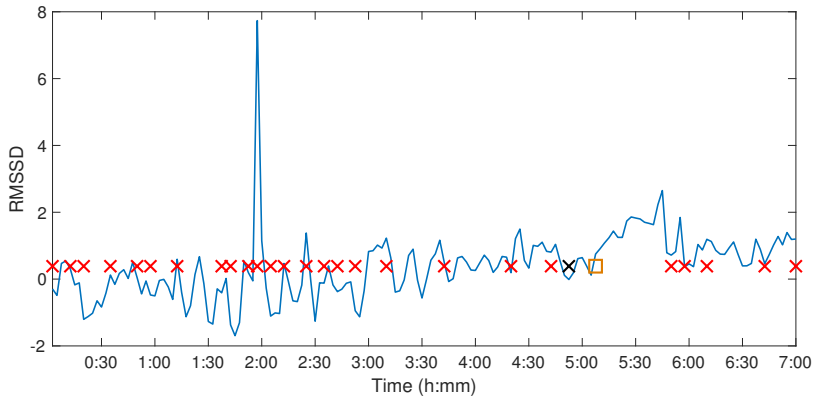
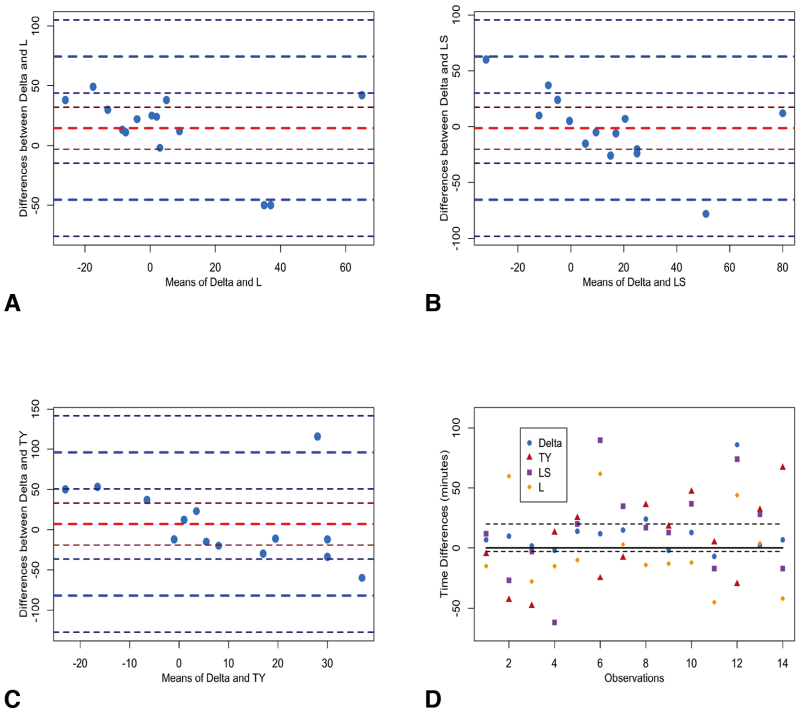


Figure 7 Bland-Altman plots of methods for Fetal Sheep dataset. Legend: TY = [15], LS = [16], L = [27]. Panel A-C: Bland-Altman plot comparing Delta point method to TY, LS, L, respectively. The y-axis displays the differences in the detection times, and x-axis displays the means of detection times for the two methods for each observation. The means (red) and two standard deviations (blue) are displayed with associated confidence intervals (maroon;navy). Panel D: Modified Bland-Altman plot of difference of each method and expert labelled sentinel time. Dashed black lines 20 minutes above and 3 minutes below sentinel time denote the clinical region of interest. Observations within this region are classified as a success.



Tables

Table 1 Simulation results

Method	Abs. Diff. (mean ± st. dev.)	MSE $\times 10^3$	Significance (p-value) [CI]
Series 1			
Delta	9.718 ± 9.881	0.192	N/A
TY	40.110 ± 28.367	2.413	1 (p <0.001) [-32.255, -28.539]
LS	8.953 ± 12.783	0.243	0 (p =0.135) [-0.237, 1.767]
L	59.459 ± 25.713	4.196	1 (p <0.001) [-51.449, -48.033]
Series 2			
Delta	3.399 ± 12.099	0.158	N/A
TY	13.802 ± 10.7298	0.306	1 (p <0.001) [-11.906, -9.406]
LS	6.205 ± 9.6699	0.132	1 (p <0.001) [-3.7662, -1.8458]
L	27.728 ± 22.961	1.296	1 (p <0.001) [-25.938, -22.719]
Series 3			
Delta	63.603 ± 30.762	4.991	N/A
TY	54.645 ± 39.358	4.534	1 (p <0.001) [5.860, 12.056]
LS	65.263 ± 63.469	8.284	0 (p =0.457) [-6.304, 2.714]
L	69.911 ± 30.418	5.812	1 (p <0.001) [-8.991, -3.625]

Legend: TY = [15], LS = [16], L = [27]. Comparison of methods for detecting significant changepoint for 1000 simulations of simulated time series: Series 1, 2, 3, respectively; 500 observations in length each. Mean ± standard deviation of absolute difference (Abs. Diff.) between labelled times and detected time of each method are given. Mean Square Error (MSE) of each method is displayed where the lowest value displayed had the least detection error. Results of null hypothesis two-sided t-test comparing absolute differences to Delta point method displayed with p-values and confidence intervals [CI].

Table 2 Donaho-Johnstone Benchmark curves results

Method	Detected time	Abs. Diff.
Bump	Labelled: 440	
Delta	445	5
TY	475	34
LS	444	4
L	468	28
Block	Labelled: 1331	
Delta	1332	1
TY	1348	17
LS	1353	22
L	1335	24

Legend: TY = [15], LS = [16], L = [27]. Comparison of changepoint detection results for two Donaho-Johnstone Benchmark curves (Bump and Block) with user-labelled changepoints. User-labelled changepoints are selected to represent a drastic change in the time series (Bump), or a significant shift in the mean (Block).

Table 3 Well-log and Nile River results

Method	Detected Time	Abs. Diff.
Well-log	Labelled: 1070	
Delta	1072	2
TY	1083	13
LS	1085	15
L	1103	33
Nile river	Labelled: 715	
Delta	720	5
TY	723	8
LS	722	7
L	725	10

Legend: TY = [15], LS = [16], L = [27]. Comparison of detected changepoint of importance in nuclear magnetic resonance measurements from a rock drill used to detect changes in rock stratification, and lowest annual water levels of the Nile River from 622-1284. The changepoint of importance is selected as the first significant jump in the mean, indicating the presence of a change in the ground rock, and the instillation of the Nilometer, respectively.

Table 4 ECG (ECGFiveDays) QRS complex results

Method	Abs. Diff. (mean $\pm$ st. dev.)	MSE	Significance (p-value) [CI]
Delta	3.51 $\pm$ 1.352	14.13	N/A
TY	5.08 $\pm$ 2.54	32.2	1 (p <0.001) [-2.138, -1.002]
LS	4.21 $\pm$ 2.944	26.31	1 (p <0.001) [-1.339, -0.061]
L	3.53 $\pm$ 3.221	22.73	0 (p =0.954) [-0.709, 0.609]

Legend: TY = [15], LS = [16], L = [27]. Comparison of methods for detecting onset of QRS complex in 100 short time ECG recordings, 136 observations in length. Mean $\pm$ standard deviation of absolute difference (Abs. Diff.) between labelled times and detected time of each method are given. Mean Square Error (MSE) of each method is displayed where the lowest value displayed had the least detection error. Results of null hypothesis two-sided t-test comparing absolute differences to Delta point method displayed with p-values and confidence intervals [CI].

Table 5 Fetal sheep experiment results

Animal	Sentinel(HH:MM)	Delta(HH:MM)	TY(HH:MM)	LS(HH:MM)	L(HH:MM)
8003	15:56	**00:07**	-00:05	**00:12**	-00:15
473351	13:38	**00:10**	-00:43	-00:27	01:00
473362	11:05	**00:02**	-00:48	**-00:03**	-00:28
473376	12:36	**-00:02**	**00:13**	-01:02	-00:15
473726	12:04	**00:14**	00:25	**00:20**	-00:10
461060	12:43	**00:12**	-00:25	01:30	01:02
473361	12:51	**00:15**	-00:08	00:35	**00:03**
473352	13:17	00:24	00:36	**00:17**	-00:14
473377	12:12	**-00:02**	**00:18**	**00:13**	-00:13
473378	13:22	**00:13**	00:47	00:37	-00:12
473727	11:03	-00:07	**00:05**	-00:17	-00:45
5054	12:53	01:26	-00:30	01:14	00:44
5060	11:26	**00:02**	00:32	00:28	**00:04**
473360	13:59	**00:07**	01:07	-00:17	-00:42
Total		11/14	3/14	5/14	2/14

Legend: Sentinel = expert defined change point TY = [15], LS = [16], L = [27]. Comparison of methods in detecting expert defined change point. Method times are displayed relative to expert Sentinel time (HH:MM), with positive values representing change points of interest detected before the Sentinel time, and negative values representing change points of interest detected after the Sentinel time. Bolded results represent a change point of interest detected in the region of interest 20 minutes to and 3 minutes after the Sentinel time.

Additional Files

Data and Codes

Data and codes associated with the manuscript may be found at the following sources:

BOCPD code:

https://sites.google.com/site/wwwturnercomputingcom/software/ThesisCodeAndData.zip?attredirects=0&d=1

Donoho-Johnstone data: ftp://ftp.sas.com/pub/neural/data/dojo_medium.txt

Well-log data: http://mldata.org/repository/data/viewslug/well-log/

Nile river data: http://mldata.org/repository/data/viewslug/nile-water-level/

ECGFiveDays: http://www.cs.ucr.edu/ eamonn/time_series_data/

Simulation and fetal data: https://github.com/ngold5/fetal_data