Change Point Problems for the Detection of Activity Bouts in Accelerometer Data

Jung Ae Lee, Andrew Bartlett
University of Arkansas, Fayetteville, AR 72701

Abstract

An accelerometer, a wearable motion sensor on hip or wrist, is a promising research tool to perform an objective experiment of human's behavior(s). An important aspect of analyzing the output of this experiment is how to classify the physical activity types regarding intensity and duration of activities, e.g., how long the moderate-to-vigorous physical activity (MVPA) last. Our approach to this is to use the change point analysis in activity series. The accelerometer output of a day is a sequence of activity counts measured at successive points in time (1 minute or 30 sec epoch). Our method justifies a change point in time at which unknown quantities of the distribution abruptly change, such as mean shift. A few well-known change point methods are applied for the detection of the exercise duration, like brisk walking or running, and compared with a threshold-based approach that detects the bouts by consecutive minutes over a threshold with the allowance of two-minutes deviation. We found that the length and the location of bouts are different across methods. This paper provides discussion between a change point approach and a threshold-based approach in terms of mathematical justification and practical implications. The National Health and Nutrition Examination Survey (NHANES) data is used to demonstrate the methods.

Key words: accelerometer, physical activity, MVPA, activity bouts, change point analysis

1 Introduction

A majority of studies with accelerometers, especially obesity research, are interested in energy expenditure (EE) of a subject. There is a large volume of work to translate the raw accelerometer counts to EE metrics, such as kcal or METs [4, 6, 8], because these are physiologically meaningful metabolic unit, allowing us to understand the relationship between physical activities and Oxygen consumption (VO_2). Threshold of these metrics (1,3,6,9 METs) are then used to categorize the level of activity intensity, for instance, METs greater than 3 can be achieved by $>$1951 count per minute with accelerometer [6], and thereby defining the moderate activity by this threshold when using accelerometer experiments in general. However, there exist substantial variations among the current EE prediction techniques from accelerometer data that often produce widely different point estimates of EE and cutoff points for identifying specific intensities of physical activity [2, 11, 13, 15, 23]. Consequently, the bouts detection relying on this intensity cutoff points per minute can present different results for the estimation of time spent on specific activity (exercise or sedentary behaviors) along with its link to health outcome.

The inconsistent predictions from accelerometer counts to EE metrics are not surprising. The existing regression prediction models have been developed in the laboratory using a small-scale of experiments with 20-70 participants along with a narrow range of activities,

which limits the use of the prediction equation in different situations. For example, the energy cost for children's vigorous activity could not be accurately estimated if we use the method developed under a group of adults. In reality, however, producing widely-agreed intensity cutoff points will be an extremely difficult task since it is highly impractical to gather data on all possible body movements and human behaviors that could occur in daily life [3]. Instead, in this paper we develop a bout detection method not depending on these thresholds so that researchers can flexibly apply their own criteria according to their research interest.

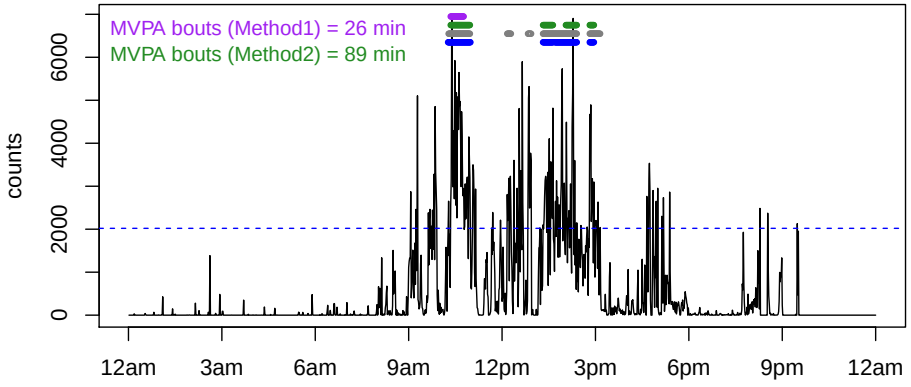
The goal of this study is to classify the physical activity bouts given on the intensity using the accelerometer measurements. Not only the level of activity intensities, but how long these activities last is of great interest among health scientists. CDC has specific guidelines for the exercise's duration, e.g., at least 10 minutes aerobic activity at a time. Prolonged sitting or lying is detrimental to one's health independent of physical activity. The work in this paper addresses how to identify the bouts of moderate-to-vigorous activity (MVPA) in particular from the accelerometer counts. Our method provides a natural way to segment a daily activity series into contiguous regions. We mathematically identify all the change points which will then partition the activity into segments where the activity type is homogeneous.

1.1 Motivating example

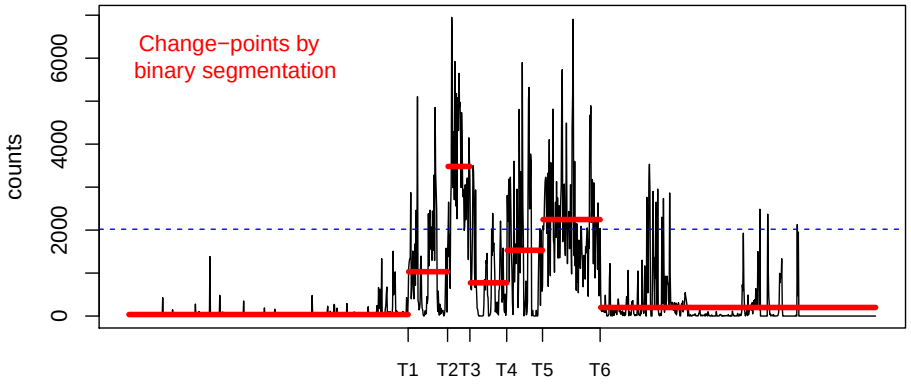
Up until this point, a commonly used method is a heuristic approach in which one can search for the bouts of moderate-to-vigorous physical activity (MVPA) incorporating with the cut-off points per minute; for instance, (100, 760, 2020, 5999) can be a set of count threshold for inactivity, light, moderate, vigorous activities [19] and then the bouts of MVPA is identified by any time interval in which all counts are greater than 2020, or often allowing 2 minutes deviations [14]. The disadvantage of this threshold-based approach is the discrepancy in results by changing the search criteria. In other words, these approaches are sensitive to threshold and tolerance, often resulting in huge differences with no indication which one is correct.

Suppose we search for 10 min bouts of MVPA. In Figure 1(a), Method 1 detects the periods in minute that the counts continuously exceed 2020 for at least 10 minutes, as a result, only 26 minutes are detected for MVPA. Likewise, Method 2 does the same thing but allowing 2 minutes deviations. Compared to Method 1, Method 2 finds more bouts of MVPA resulting in 89 minutes in total. However, it is not clear why 2 minutes allowance is appropriate. Two other methods, colored by gray and blue, use 3 minutes tolerance and 2 minutes tolerance, respectively, with a threshold of 1800, which create a difference as well. Various unexplained exceptions exist with these threshold-based approaches.

On the other hand, a change point analysis is applied in Figure 1(b) to the same dataset, which provides a more unified and mathematically justified procedure for detecting change points within an activity time series. An activity time series is a sequence of activity counts measured at successive points in time (1 min epoch). Intuitively, we can think of a *change point* as a point in time at which unknown quantities (parameters) of the distribution or model abruptly change. For example, in (b), the activity level, displayed by a black line, is flat until a time point T1 and then begin to change for the next hour between T1 and T2 and then all of a sudden the activity level jumps up above the blue dot line with fluctuations between T2 and T3. The moment when we observed a sudden change in the activity data series would be described as a change point in which the active counts distribution changed.



(a) Threshold-based approach



(b) Change-point approach

Figure 1: Illustration of detecting the exercise duration. In (a), MVPA 10-minute bouts are computed by two threshold-based methods; Method 1 searches for continuous minutes exceeding the threshold (2020 count per minute) for 10 min while Method 2 allows 2 minutes deviations. In (b), we use the change point analysis where two durations (T2-T3 and T5-T6) are detected for MVPA activity.

1.2 Our approach

In this study, change point analysis is constructed to effectively categorize the activity types through a daily activity time series. The example in Figure 1(b), in fact, contains all necessary components for the change point analysis, including the number of change points, where this change point occurs, the duration of each local segment, and the average intensity within each local segment as shown by horizontal red lines. Therefore if the horizontal red line exceeds the threshold (blue dots), we find the bouts of MVPA from the detected durations T2-T3 and T5-T6, and classify other durations in a similar fashion.

The ability to detect unknown change points within an activity time series is critical for accurately determining time spent on diverse physical activity types (MVPA, SB and Sleep). Note that in the example of Figure 1(b) we set the maximum number of change points to six previously, but usually the number of change points is unknown. Establishing the existence, and ultimately estimating the number and locations of change points can be an extremely challenging task both theoretically and computationally. Despite the difficulty, the need is essential and therefore has received much attention over the past 40 years linking to many different application areas, including climatology [22], bioinformatics [5, 17], finance [25], oceanography [10], medical imaging [18]. This paper is the first, to our best knowledge, to apply change point analysis to physical activity measurements.

Finding change point(s) can be seen as the subdivision of a series into segments characterized by homogeneous statistical features (e.g., mean and standard deviation) while maximizing the heterogeneity between time segments. Numerous studies have presented methods to detect the change points by testing the mean shift within a time series. When these mean tests are applied to activity data, we expect two main challenges. One is the extreme values from irregular vigorous activity which may inflate the local mean, and the other is the zero-inflations from either non-movement or missing values. These phenomena make it hard to choose a distributional assumption, so our investigation includes what differences can be led by different assumptions and methods. In the following, we will explore parametric vs nonparametric models and the procedure how these methods can be used for identifying MVPA bouts on accelerometer data.

2 Homogeneous Test in Mean

Detection of a shift in random processes was first considered by Page (1955) [20]. Among many others, Hinkley [9] and Sen and Srivastava [24] considered this procedure by a test of change in mean within normally distributed observations. Since these tests can detect a shift after a singe point, multiple change points can be handled by applying these tests sequentially until there are no change points. Variants of the mean shift test are numerous including Pettit's nonparametric test [21].

The notion for testing a mean shift is following. Suppose that $X_1, X_2, \ldots, X_n$ is an activity count series in a day, e.g., $n = 1440$ for one-minute epoch. It is reasonable to model the $\{X_i\}$ by

$$X_i = \mu(i) + \varepsilon_i \,, \quad i = 1, \ldots, n \tag{1}$$

where

$$\begin{aligned}\mu(i) &= \mu_1\ , \quad 0 < i \leq \tau_1\ , \\ &= \mu_2\ , \quad \tau_1 < i \leq \tau_2\ , \\ &= \vdots \\ &= \mu_{r+1}\ , \quad \tau_r < i \leq n\ .\end{aligned}$$

This model implies $r+1$ segments created by r change points with a set of locations, $\mathscr{C} = \{\tau_1, \ldots, \tau_r\}$ where $\mathscr{C} \subset \{1, \ldots, n-1\}$ and $\tau_1 < \tau_2 < \cdots < \tau_r$. The μ_j are the means of the jth segment over the period, $(\tau_{j-1}, \tau_j]$, and thus, $\tau_j + 1$ is the location in time that the mean change occurred where $j = 1, \ldots, r$, $\tau_0 = 0$, and $\tau_{r+1} = n$. It is assumed that the μ_j, local mean, τ_j, location of mean change, and the r, number of change points, are unknown. The ε_i are independent errors, and this assumption is a controversy for one-minute epoch series, where there is empirical evidence of autocorrelation [12]. Nevertheless, we assume the autocorrelations are ignorable since it fades away in 1-3 minutes and no longer present between segments.

For the problem of exercise durations, we are particularly interested in detecting the bout $(\tau_{j-1}, \tau_j]$ such that $\mu_j > c$, where c is a known constant. The known c is a physiologically meaningful cutoff; for instance, the metabolic equivalent (MET) > 3 can be classified as a moderate-to-vigorous activity from the measurement of Oxygen consumption. This value in accelerometer data, equivalent to MET of 3, has been studied under different devices and diverse populations ranged from 1267 to 2743 count per minute in literatures [16]. We recommend to use the judgement of the researcher from these references.

Due to the known c, we can modify the strategy of detecting the change points. One strategy is to select the r, from a reasonable set until we find any bout over c, and stop this search when no more meaningful bout $> c$ are detected. Whether a physical activity has occurred is less important question than how long this activity lasts. Thus we focus on tackling the estimation of the duration in this work.

2.1 Test statistics under Normal

The standard normal mean test has been studied in Hinkely [9], Sen and Srivastava [24], Hawkins [7], Alexandersson [1] and others. In Reeves' review [22], a test statistic for the mean change is presented analogous to two sample t-test. That is, when the data are normally distributed with an unknown variance σ^2, the likelihood ratio statistic for testing the null hypothesis that there is no change against the alternative that there is exactly one change at unknown location i is given by

$$T^2_{\max} = \max_{1 \leq i < n} T_i^2 \quad \text{with} \quad T_i = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{1/i + 1/(n-i)}}, \tag{2}$$

where $\bar{X}_1$ and $\bar{X}_2$ denote the sample means of $\{X_i\}$ before and after i and s_p is the pooled estimate of the standard deviation of $\{X_i\}$. T_i in (2) is the standard two-sample t-test statistics when the variance σ^2 is unknown.

An appropriate modification for physical activity data is to release the equal variance assumption. As shown in Figure 1, activity counts fluctuate in variance, so it is reasonable to replace T_i with Welch's t-statistic in (2) for the detection of the mean shift under unequal variance,

$$T_i^{'} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1/i + s_2/(n-i)}}, \tag{3}$$

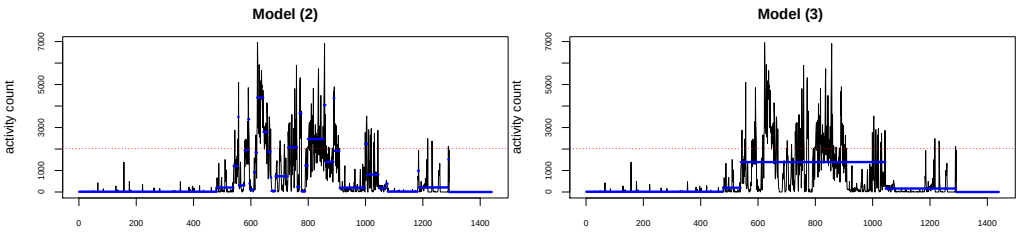
where s_1 and s_2 denote the sample standard deviation of $\{X_i\}$ before and after i.

The distributions of the test statistics are simulated under the null hypothesis of no change points with normally distributed data with 1000 repetitions. Different test statistics are compared: (a) Sen and Srivastava's test assuming a known variance of 1 [24], (b) Pettit's nonparametric test [21], (c) t-test under unknown equal variance assumption in (2), and (d) Welch's t-test under $\sigma_1^2 \neq \sigma_2^2$ in (3). The critical values at 0.95 quantile are summarized in Table 1.

Table 1: The 95% critical values for the null distribution of no change point under Normal

n	$\sigma^2 = 1$ (a)	nonparamatric (b)	unknown σ^2 (c)(2)	$\sigma_1^2 \neq \sigma_2^2$ (d)(3)
30	8.603	6.739	9.522	40.690
60	9.368	8.581	10.090	42.736
120	9.809	9.025	10.150	76.197
700	10.337	9.917	10.256	213.868
1440	11.242	10.512	11.122	295.015

The cutoff points for the test statistic are similar for (a),(b) and (c), but method (d) results in drastically different critical values than method (a),(b), or (c). As a result, the number of change points is over-estimated under (a), (b) and (c), whereas it is underestimated under (d). In Figure 2, for example, the left panel by method (2) identifies too many (false) change points while the right panel by method (3) is too conservative to identify the change points.



36 change points 4 change points

Figure 2: Different change point detection under Model (2) and (3)

2.2 Nonparametric test statistic

The standard homogeneous mean test in the previous section is a likelihood ratio test under independent and identical Gaussian random variables. Normality is a debatable assumption for activity count series. Typical option is a nonparametric test procedure, which prevents false change points from outliers especially near boundaries. The robustness (less sensitive to distributional departures from normality) is due to the rank transformation of the observed values. General reputation of nonparametric tests is a lower power (i.e., missing change points). But if the sample size is large and the parametric assumptions are violated, such nonparametric tests may be only slightly less powerful and will provide better false-detection rates (type I errors).

A nonparametric test statistic in change point context departs from the equivalence to $T_{\max}$ in (2), the maximum of two-sample t statistics over all possible locations [22]. In other words, the nonparametric version is the maximum of Wilcoxon rank-sum statistics (or equivalently, Mann-Whitney statistics) over all possible locations. Thus, a nonparametric standard homogeneous mean test is given by

$$W_{\max} = \max_{1 \le i < n} W_i, \tag{4}$$

where W_i is the square of a normalized Wilcoxon rank-sum statistic for each fixed location i,

$$W_i = 12 \frac{\left(\sum_{t=1}^{i} r_t - i(n+1)/2 \right)^2}{i(n-i)(n+1)}$$

where r_t is the rank of the tth element in the count series (e.g., X_1 is 7th largest value, then $r_1 = 7$). Therefore one detects a change point at location i when $W_{\max}$ is sufficiently large. Critical regions for the $W_{\max}$ under the null hypothesis of no change point were simulated, as with $T_{\max}$, in Table 1.

2.3 Test statistic under Poisson

Standard homogeneous test procedure is a test for the mean shift under the assumption of equal variance. However, the output of accelerometer experiment is a series of nonnegative counts and the variance changes as the mean level changes. The Poisson distribution is a natural choice to fit the model for this data. One attraction of Poisson distribution is the mean-variance relationship, which simplifies the test statistic with one parameter, so the test of mean implies the test of variance.

Generally, suppose we have n observations from $y \sim f(y|\theta)$, and there exists a single change point at τ_1, the maximum log likelihood for a given location τ_1 is

$$ML(\tau_1) = \log f(y_{1:\tau_1}|\hat{\theta}_1) + \log f(y_{(\tau_1+1):n}|\hat{\theta}_2)$$

Since the location of the change points are unknown, the maximum is taken over all possible change point locations under the alternative hypothesis of existence of change point is

$$\max_i ML(i) \quad \text{for} \quad i \in \{1, 2, \ldots, n-1\},$$

implying at which location a change point most likely occurs. Therefore, the test statistics is given by

$$T_{general} = 2 \left[\max_i ML(i) - \log f(y_{1:n}|\hat{\theta}) \right],$$

which value is close to zero under the null hypothesis and becomes large for the alternative. The test involves choosing a critical point, which is still an open question, to determine the existence of a change point along with the location of the change point. Under Poisson, this test statistic is

$$T_{Poi} = \max_i [i\bar{y}_1 \log(\bar{y}_1) + (n-i)\bar{y}_2 \log(\bar{y}_2)] - n\bar{y} \log(\bar{y})$$

We simulated the null distributions under no change point with the Poisson random variables by 1000 repetitions, and the empirical 95% upper quantiles are identified while

the sample size n and the size of λ change, where λ is a Poisson parameter, meaning the average count of activity in accelerometer. The results are shown in Table 2. The 95% critical point slightly moves up as the sample size increases, but the impact of different λ on the critical point is minor with an exception of $\lambda = 10$. Luckily, $\lambda = 10$ indicates an extremely low average count of activity in accelerometer during that time period, which is not a focus in this work.

Table 2: The 95% critical values for the null distribution of no change point under Poisson

n	$\lambda = 10$	$\lambda = 100$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 3000$
30	5.125	4.535	4.585	4.365	4.165
60	5.974	4.871	4.363	4.874	4.579
120	6.191	5.377	5.190	4.904	4.894
700	6.941	5.785	5.397	5.494	5.177
1440	7.137	5.871	5.328	5.409	5.447

2.4 Application to MVPA bout detection

Detection of mean shift by sequential binary segmentation based on level α test is arguable in many aspects. First, there does not exist a simplified probability density function of the null distribution, and the α upper quantile for rejection is not justifiable in theory. In applications, it is rather being achieved by a computationally expensive simulation as shown in Table 1-2. Second, statistical power is different depending on the location of change point (middle or edge). Third, α significance level applied to recursive binary segmentation could be inaccurate depending on how many subsegments exist within, probably needing multiple testing adjustment. If one ignores multiple testing aspects, any detection method will yield too many false change points. Fourth, the assumptions of normality and equal variance would never be correct with real data, which nullify the decision based on these assumptions.

Fortunately, we are in a situation where a statistical decision based on level α test is less critical. Our primary interest lies in detecting a time period (locations) of being meaningfully active rather than knowing the presence of a change. The presence of a change towards MVPA is judged by our own threshold (e.g., 2020 count per minute). In other words, our research question is related to where exactly the location would be if there were a change during this period. This allows us to care less about the rejection region by level α test that meant to test the hypothesis of whether there is a change or not.

Another reason that a level α critical region is less concerned is that, when the null is rejected, the test statistic tends to be extremely far from the α upper quantile because we are only interested the periods that are highly active. Overall, our research question modifies the algorithm to detect locations of the activity bouts when using a test-based approach. The *binary segmentation procedure* applies the test recursively until no more changes are detected in any of the segments obtained from the change points already found. However, stopping criteria for the rejection of the null of no changes can be relaxed for the sake of finding more locations. The following is the summary of our modified strategy in binary segmentation procedure:

1. Use a relaxed α level like a 1.0 rather than a 0.05 or less. It is acceptable to have

many (false) change points.

2. If the null is rejected by a relaxed α (i.e., there is a change point), calculate the maximum likelihood test statistics in all minutes to evaluate the most likely location.

3. Stop search if

   - no more change points are detected at 0.1 level,
   - length of segment is less than or equal to 10 minute,
   - or until K-th binary segmentation, where K is a sufficient number of recursive steps of binary segmentation (e.g., K=10)

4. Combine two or more MVPA bouts if they are adjacent.

5. To decide the MVPA bouts (i.e., exercise duration), select bouts that are more than threshold while lasting more than 10 minutes.

3 Activity Data Example

The real data examples are from 2003-2004 National Health and Nutrition Examination Survey (NHANES) dataset. The NHANES data set is available at the website: `http://www.cdc.gov/nchs/nhanes/search/nhanes03_04.aspx`. The analysis here uses 763 daily profiles of 109 people (109×7 d), following Lee and Gill [12]'s data selection process. This dataset is also publicly available within statistical software R package *accelmissing*. Table 3 summarizes the characteristics of the dataset. From the table we see that the mix of demographics is nationally representable. The 2020 count per minute is used for MVPA threshold for this data.

Table 3: Summary of Data (No. of Participants = 109, N = 763 days)

Age(%)*		Sex (%)		BMI (%)		Race (%)	
Youth	38.5%	Male	50.5%	≤ 25	42.2%	White	44.0%
Adult	61.5%	Female	49.5%	> 25	57.8%	Others	56.0%

* Youth indicates 7-19 yrs and Adult indicates 20-85 yrs.

4 Results with real data

Four methods are applied for the detection of 10-minute bouts of moderate-to-vigorous physical activity (MVPA) as a physiologically meaningful exercise duration. (1) First, a threshold-based method identifies MVPA bouts when the count per minute exceeds 2020 for consecutive 10 minutes allowing two minutes deviation. This is the most common approach in practice. (2) Secondly, a change point method identifies MVPA bouts based on a likelihood ratio test statistic under a Normal distributional assumption. (3) Thirdly, a change point method applies a likelihood ratio test procedures to the relative ranks of the data. This method is referred to as a nonparametric test due to the fact that there is no distributional assumption of data and Mann-Whitney test statistic is involved in this

procedure. (4) Fourthly, a change point method under a Poisson distribution is included to present a situation where both mean and variance change.

In conclusion, four methods display different results in finding MVPA bouts. A numerical summary of one example is presented in Table 4, along with a graphical summary in Figure 3, where the different estimates of exercise duration (MVPA bouts) are suggested by different methods. For general conclusion, 763 daily profiles are inspected, and summaries are following.

Table 4: Differently estimated MVPA bouts by different methods from a daily profile

Method	minute start	minute end	MVPA bout(min)*	local mean	minutes > 2020	percent > 2020	clock start	clock end
(1)	729	744	16	3,952	13	81	12:08	12:23
			total : 16					
(2)	590	650	61	3,151	33	54	09:49	10:49
	727	742	16	3,916	12	75	12:06	12:21
			total : 77					
(3)	575	649	75	2,601	32	43	09:34	10:48
	726	747	22	3,202	13	59	12:05	12:26
			total : 97					
(4)	590	600	11	4,321	7	64	09:49	09:59
	628	637	10	5,373	7	70	10:27	10:36
	729	744	16	3,952	13	81	12:08	12:23
			total : 37					

* exercise duration

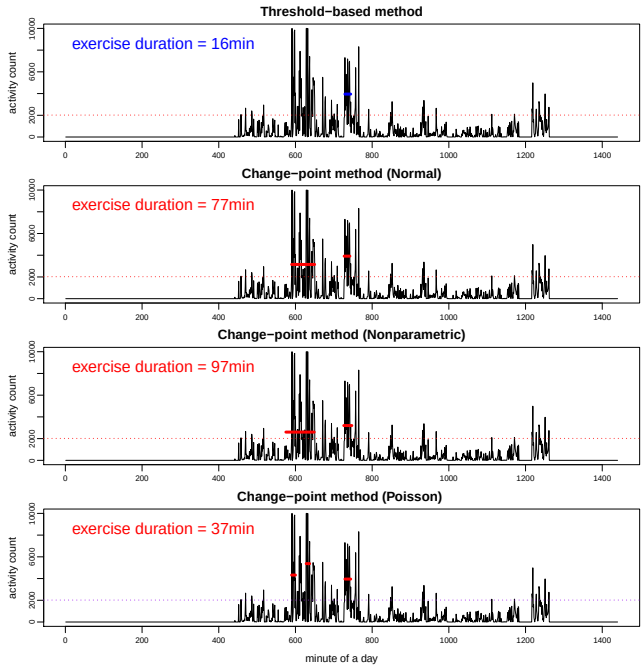


Figure 3: A day profile of an individual with an exercise. The y-axis displays count per minute from accelerometer data, and the x-axis displays 1440 minutes of a day.

(1) The nonparametric test statistic is disadvantageous for activity bout detection. The robustness to outliers of rank statistic is in fact against our intention to detect the relatively extreme values. Some obvious cases are displayed in Figure 4-5, in which four methods are compared simultaneously. From the visual impression, it is clear that there exists a MVPA, and it is expected for any method to find it easily. In the third panel, however, the nonparametric method fails to detect the exact locations for the MVPA by either overestimating or underestimating the duration. That is because using relative ranks works against detecting the locations where activity abruptly jumps.

(2) The threshold-based method shows less power for the detection of sparse moderate-to-vigorous activity. In Figure 3, the first panel by threshold-based method does not identify the period of a MVPA around 600 in x-axis while the other three methods detect it somewhat differently. The reason is that the two minutes restriction, not only allows two minutes deviation for continuous bouts, but also prevents the detection from sparse MVPA. For example, one hour exercise with half walking and half running can be exclusively regarded as a MVPA, but it is possible not to find so when the duration contains more than 2 minutes interruption between the runnings or even more sparse structure of time.

(3) The change-point approach under Normal tends to aggregate the time periods according to relative mean level. As shown in Figure 6, the second panel by change-point method under a Normal distribution tends to aggregate the time periods, resulting in expanding the exercise duration. This result also reflects the fact that this test has a weak power on the edges. Often, a lowered mean level due to the expanded bouts may result in dropping the MVPA bout as a whole because it does not reach the threshold as shown in Figure 7. These different results are not necessarily wrong depending on the research interest. This method is beneficial to find a sparse moderate activity, which can be missed by the threshold-based approach.

(4) The change-point approach under Poisson tends to find a short bouts being sensitive to variance changes as well as mean. From Figure 5-7, the fourth panel by change-point method under a Poisson distribution tends to detect a short burst bouts compared to other methods. When the results are conflicted among methods, the Poisson method gives neutral estimates overall while finding multiple local short bouts. That is because the test statistic under a Poisson is a one parameter statistic assuming the equality of mean and variance. Although this mean-variance relationship is incorrect, (in fact, typical over-dispersion problem appears in the accelerometer data), this test statistic allows us to detect a different quality of periods in terms of the variances as well as the means, which could be buried by the Normal test statistics.

These examples demonstrate the need for care in estimating exercise duration. Methods presented here by itself provides no information about the energy cost and must be supplemented by graphical diagnostics to obtain insights about practical exercise duration.

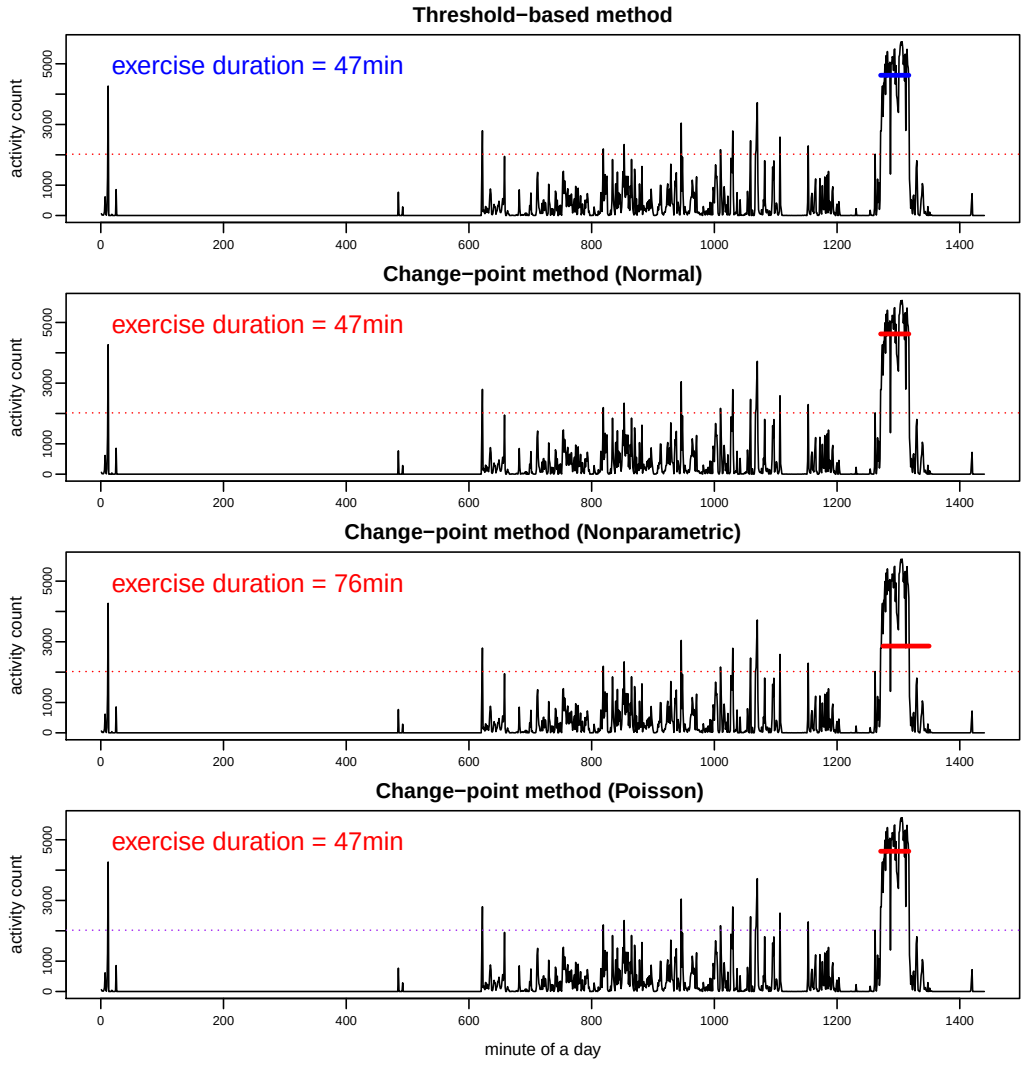


Figure 4: All methods but the nonparametric test method identify the same MVPA bouts of 47 minutes

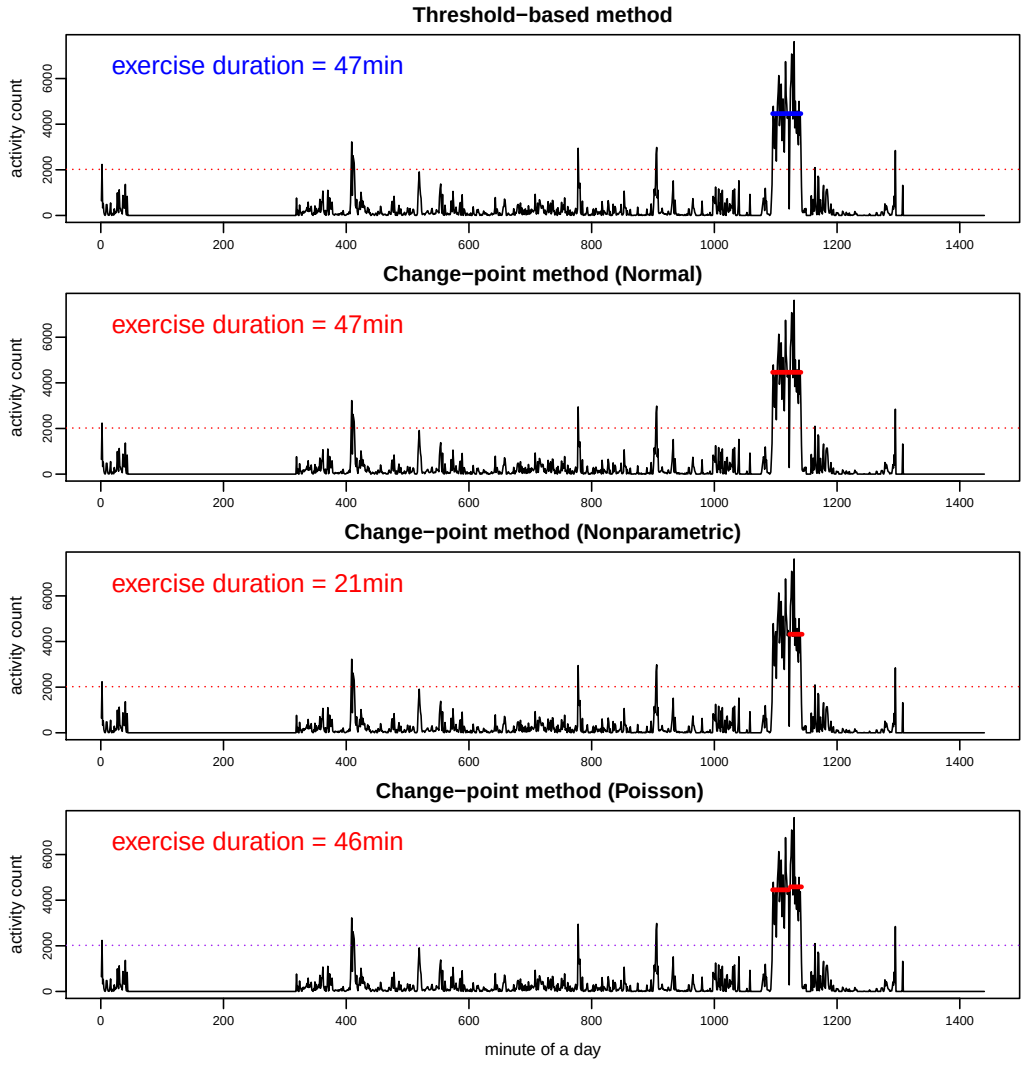


Figure 5: Another example of insensitive nonparametric test, a different day of the same individual in Figure 4. Only the nonparametric test approach (third panel) fails to detect the appropriate exercise duration.

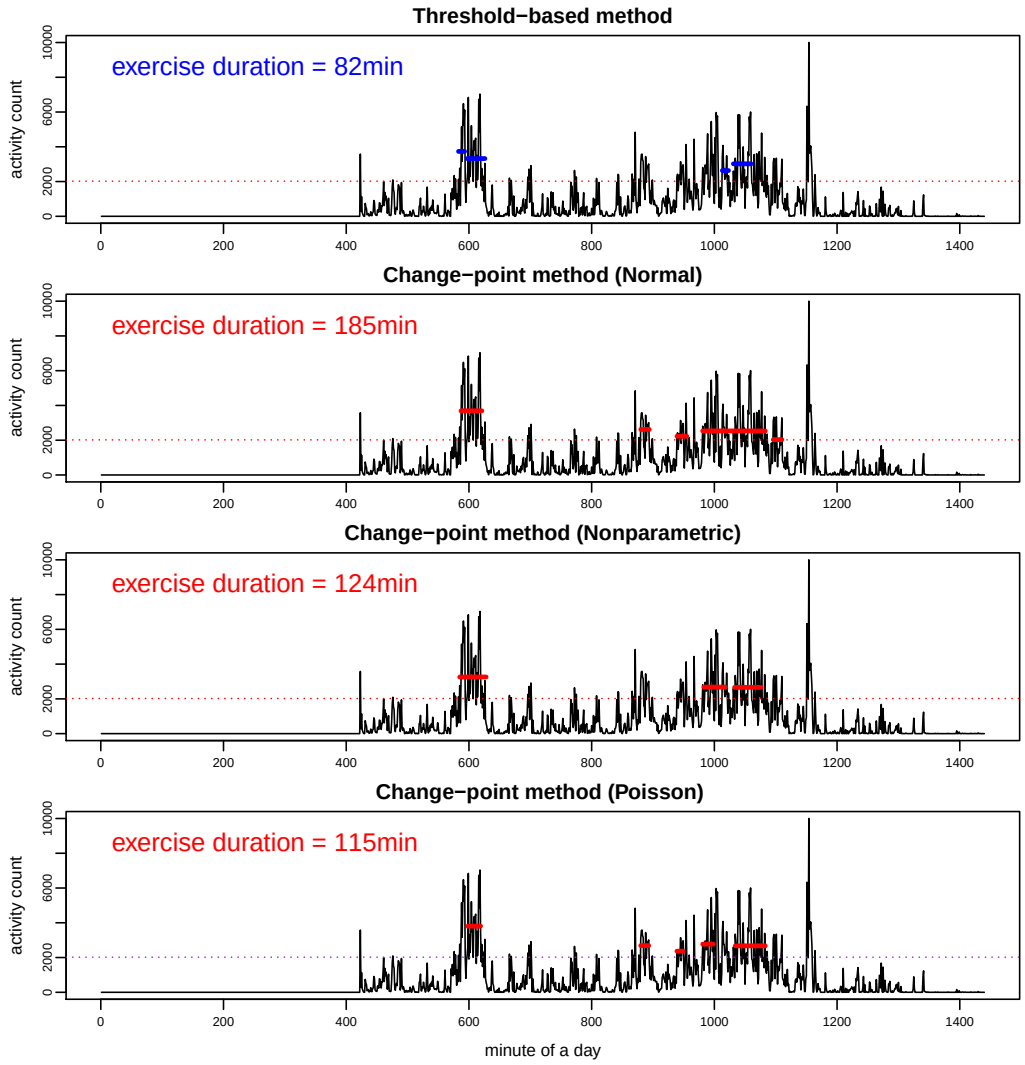


Figure 6: The threshold-based method (first panel) is less sensitive to sparse moderate activity around 1000 in x-axis. The change-point method under Normal (second panel) tends to aggregate the time periods, which may overestimate the exercise duration but detect sparse moderate activities. The change-point method under Poisson (fourth panel) finds local short burst bouts.

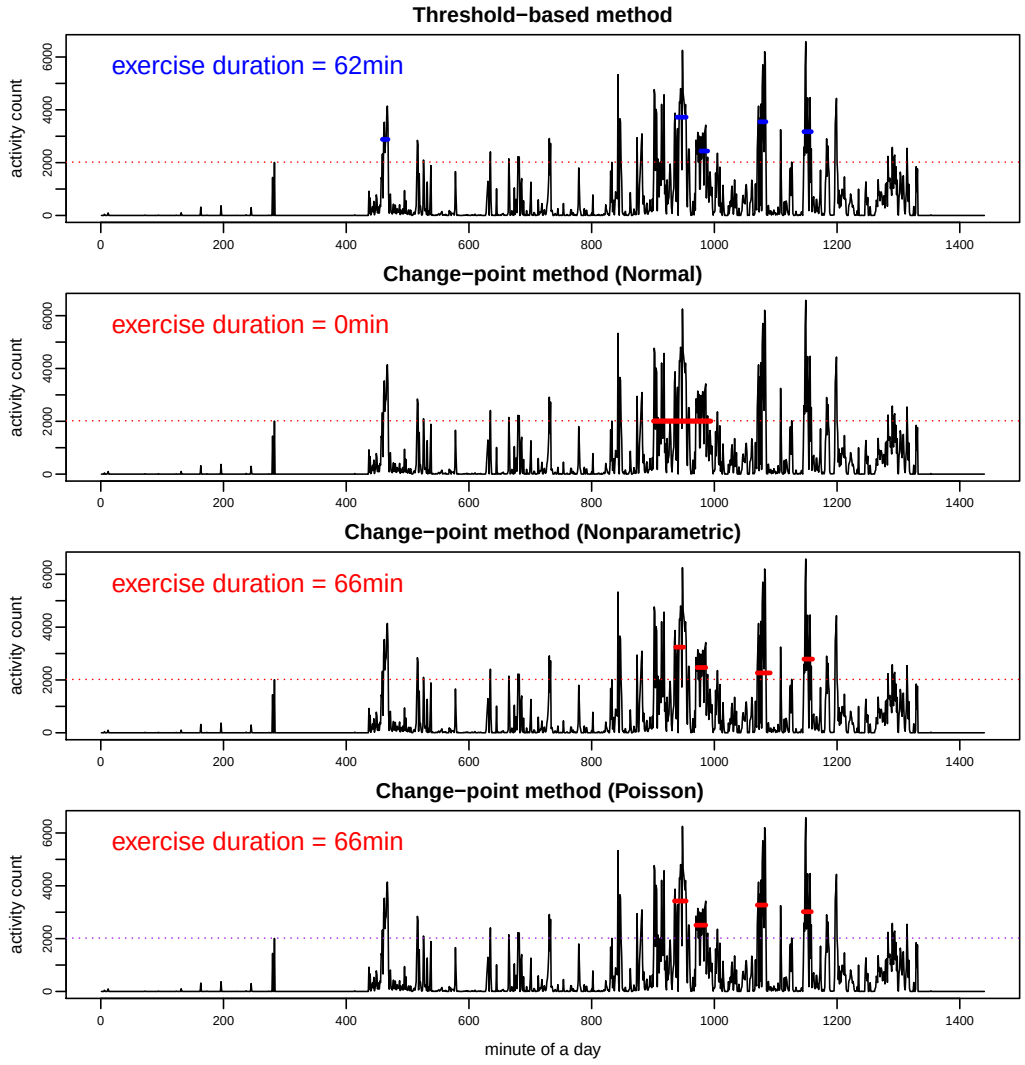


Figure 7: The change-point method under Normal (second panel) tends to find relatively long bouts due to the homogeneous mean test mechanism, as a result the sparse moderate activity may not reach the threshold.

5 Optimizing the Number of Change Points

Alternative to level-α test approach, one can find the number and positions of change points that optimize a cost function. The number of change points r is usually optimized with a Schwarz-like penalty. Appropriate choices are $2r$, $r\log n$, $2r\log\log n$, and $n(a+b\log(n/r))$ where a,b are data-driven (Lebarbier, 2005). For example, the objective function for the optimal number of change points under a Normal distribution (Yao, 1988) is given by

$$SC(r^*) = -\frac{n}{2}\log\hat{\sigma}_r^2 - r\log n,$$

and under a Poisson distribution is given by

$$SC(r^*) = \sum_{j=1}^{r+1} n_j\bar{y}_j\log\bar{y}_j - r\log n,$$

and we propose an estimator r^* to maximize this function. In our work, finding the optimal r or a choice of penalty function is not as meaningful. For the Normal test procedure, applying the optimal number of change points did not change the final detection of MVPA bouts. For the Poisson test procedure, there does not exist a maximum or a minimum of the objective function with aforementioned choices.

6 Discussion

Upon different results on real data, it is a natural question which method should be claimed to be "correct"? There is no unique method that is absolutely correct in all cases. Each method has a reason for a different location or number of change points. A threshold-based method is not justifiable in a mathematical sense, e.g., no justification for why 2 minutes allowance is appropriate, whereas a change point method is a test-based approach that mathematically justifies the locations or the number of change points under certain distributional assumption.

One of the fundamental differences between change point approach and threshold-based method is the use of threshold for the detection process. The threshold-based method is sensitive to an initial threshold value for the final detection. Meanwhile, change point approaches do not depend on this threshold itself although it involves when filtering the final set of MVPA bouts. Among methods, the nonparametric test method is not recommended in general. The Normal test can be useful to identify sparse moderate activities. The Poisson test is useful to identify the short intensive MVPA bouts.

In this paper, the exercise duration is measured by total length of MVPA bouts allowing a bit under-threshold deviation. This idea is useful for a physical activity guidelines to suggest a practical exercise time. In other situations, MVPA Bouts is not necessarily a synonym of exercise duration. It might be useful to just sum the minutes over threshold during the detected MVPA bouts if one is interested in collecting physiologically meaningful minutes. In all cases, it is a good idea to analyze the bouts with graphical diagnostics. Also, biological information such as Oxygen consumption (VO_2) during the bouts should be examined, not only a minute unit, through the lab experiments.

References

1. Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, 6(661-675).

2. Alhassan, S., Lyden, K., Howe, C., Kozey-Keadle, S., Nwaokelemeh, O., and Freedson, P. S. (2012). Accuracy of accelerometer regression models in predicting energy expenditure and METs in children and youth. *Pediatric exercise science*, 24(4):519–536.

3. Bonomi, A. G. and Plasqui, G. (2012). Divide and conquer: Assessing energy expenditure following physical activity type classification. *Journal of applied physiology*, 112(5):932.

4. Crouter, S. E., Kuffel, E., Haas, J. D., Frongillo, E. A., and Bassett, Jr., D. R. (2010). A refined 2-regression model for the actigraph accelerometer. *Medicine and Science in Sports and Exercise*, 42(5):1029–1037.

5. Erdman, C. and Emerson, J. W. (2008). A fast bayesian change point analysis for the segmentation of microarray data. *Bioinformatics*, 24(9):2143–2148.

6. Freedson, P. S., Melanson, E., and Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine and Science in Sports and Exercise*, 30(5):777–781.

7. Hawkins, D. M. (1979). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72:180–186.

8. Heil, D. P. (2006). Predicting activity energy expenditure using the actical activity monitor. *Research quarterly for exercise and sport*, 77(1):64–80.

9. Hinkley, D. V. (1972). Inference about the change point in a sequence of random variables. *Biometrika*, 57:1–17.

10. Killick, R., Eckley, I. A., Jonathan, P., and Ewans, K. (2010). Detection of changes in the characteristics of oceanographic time-series using statistical change point analysis. *Ocean Engineering*, 37(13):1120–1126.

11. Kozey, S. L., Lyden, K., Howe, C. A., Staudenmayer, J., and Freedson, P. S. (2010). Accelerometer output and MET values of common physical activities. *Medicine and Science in Sports and Exercise*, 42(9):1776–1784.

12. Lee, J. A. and Gill, J. (2016). Missing value imputation for physical activity data measured by accelerometer. *Statistical Methods in Medical Research*, March. preprint.

13. Lyden, K., Kozey, S. L., Staudenmayer, J. W., and Freedson, P. S. (2011). A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *European journal of applied physiology*, 111(2):187–201.

14. Masse, L. C., Fuemmeler, B. F., Anderson, C. B., Matthews, C. E., Trost, S. G., Catellier, D. J., and Treuth, M. (2005). A comparison of four reduction algorithms on selected outcome variables. *Medicine and Science in Sports and Exercise*, 37(11 Suppl):S544–S554.

15. Matthews, C. E. (2005). Calibration of accelerometer output for adults. *Medicine and Science in Sports and Exercise*, 37(11 Suppl):S512–S522.

16. Metzger, J. S., Catellier, D. J., Evenson, K. R., Treuth, M. S., Rosamond, W. D., and Seiga-Riz, A. M. (2008). Patterns of objectively measured physical activity in the united states. *Medicine and Science in Sports and Exercise*, 40(4):630–638.

17. Muggeo, V. M. R. and Adelfio, G. (2010). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27:161–166.

18. Nam, C. F. H., Aston, J. A. D., and Johansen, A. M. (2012). Quantifying the uncertainty in change points. *Journal of Time Series Analysis*, 33(5):807–823.

19. National Health and Nutrition Examination Survey (2003-4). SAS programs for analyzing NHANES 2003-2004 accelerometer data. `http://epi.grants.cancer.gov/nhanes_pam`. Last update on Feb 03 2016.

20. Page, E. S. (1955). A test for a change in a parameter occuring at an unknown point. *Biometrika*, 42:523–526.

21. Pettit, A. N. (1979). A non-parametric approach to the change point problem. *Applied Statistics*, 28:126–135.

22. Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meterorology and Climatology*, 46(6):900–915.

23. Rothney, M. P., Schaefer, E. V., Neumann, M. M., Choi, L., and Chen, K. Y. (2008). Validity of physical activity intensity predictions by Actigraph, Actical and RT3 accelerometers. *Obesity*, 16(8):1946–1952.

24. Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of Statistics*, 3(1):98–108.

25. Zeileis, A., Shah, A., and Patnaik, I. (2010). Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics and Data Analysis*, 54(6):1696–1706.