Biometrics Manuscript ID BIOM000 , 1–??
000 000

DOI: 000

Changepoint Detection on Daily Home Activity Pattern: A Sliced Poisson Process Method

Israel Martínez-Hernández

School of Mathematical Sciences, Lancaster University, Lancaster, United Kingdom.
email: i.martinezhernandez@lancaster.ac.uk

and

Rebecca Killick

School of Mathematical Sciences, Lancaster University, Lancaster, United Kingdom.
email: r.killick@lancaster.ac.uk

Summary: The problem of health and care of people is being revolutionized. An important component of that revolution is disease prevention and health improvement from home. A natural approach to the health problem is monitoring changes in people's behaviour or activities. These changes can be indicators of potential health problems. However, due to a person's daily pattern, changes will be observed throughout each day, with, e.g., an increase of events around meal times and fewer events during the night. We do not wish to detect such within-day changes but rather changes in the daily behaviour pattern from one day to the next. To this end, we assume the set of event times within a given day as a single observation. We model this observation as the realization of an inhomogeneous Poisson process where the rate function can vary with the time of day. Then, we propose to detect changes in the sequence of inhomogeneous Poisson processes. This approach is appropriate for many phenomena, particularly for home activity data. Our methodology is evaluated on simulated data. Overall, our approach uses local change information to detect changes across days. At the same time, it allows us to visualize and interpret the results, changes, and trends over time, allowing the detection of potential health decline.

Key words: B-spline basis; Changepoints detection; PELT; Segmentation; Sequence of inhomogeneous Poisson processes.

1. Introduction

This paper is part of a vision to revolutionize health and care in the community for 2050 (QUEST, 2024). When imagining the community settings and homes of 2050, we envisage multiple sensors and media features in our smart environment. Data will be collected on a large scale, at different frequencies (e.g., seconds or minutes), and from different sources; some of these are being collected now. Statistical methods are required to analyze these complex and large datasets. There are several different types of sensors and goals for statistical analysis of these sensors. Some recent work includes Wifi disturbances (Usman et al., 2022), fatigue using RF sensing (Cooper et al., 2024), vital sign using radar (Elsayed et al., 2024), and changes in gait (Austin et al., 2011) – all using passive sensing.

One type of data in the sensing context is a collection of random event times, known as temporal point process: $N(t) = \sum_{j\geq 1} 1_{\{t_j \leq t\}}$, where $\{t_j\}$ are the event times, and $N(t)$ represents the total number of events up to time t (see, e.g., Baddeley, 2007). The most commonly used point process is the temporal Poisson process, which is characterized by the mean measure $\lambda(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{N(t)\}$, where $\lambda(t)$ is known as the intensity function. Loosely speaking, $\lambda(t)$ represents the probability that there will be an event within a "small" time interval $[t, t + \mathrm{d}t]$. If this function is not constant over t, then the Poisson process is known as an inhomogeneous Poisson process (IHPP). Here, we propose a novel statistical methodology to detect changes in event times using a new paradigm of data analysis. Specifically, we propose a changepoint methodology for a sequence $\{N_i(t)\}$ of IHPPs, where each N_i is defined on a specific period of time (e.g., 24 hours).

Our proposal is motivated by home activity data provided by Howz (2024). The company collects data on household activities of older people. The dataset consists of a collection of times, representing activities performed by a household, e.g., using the kitchen, walking through the hallway, and opening the back door, see top left plot of Figure 1. The rate of events will vary throughout each day due to a household's daily pattern, with, e.g., fewer events during the night, and increases around meal times. Notice that these within-day changes are part of the regular routines. We do not wish to detect such changes, but rather changes in the daily behaviour pattern from one day to the next. For example, from the top left plot of figure 1, we observe the same behaviour (no "change") between day 1 and 2. In contrast, there appears to be a change between day 2 and 6. On day 6, activities start hours later than on day 2. Similarly, day 6 and day 36 vary. These across-day changes are

more informative than the within-day changes in monitoring people's health. To that end, we treat the set of event times on a given day as a single observation. That is, for each day i, the process $N_i(t)$, $t \in [0, 24]$ is assumed to be a single observation and represents the data for day i. We model this observation as a realization of an IHPP whose rate function can vary with the time of day. Figure 1, top right plot, shows all realizations of N_i with the Howz data. We then wish to detect changes in this sequence $\{N_i\}$ of processes. Specifically, if N_i and N_{i+1} are realizations of the same process, then there is no change. In contrast, there will be a change if they are realizations from two different processes. This induces a segmentation of $\{N_i\}$, as shown in the bottom plot of Figure 1, where each segment represents realizations of the same IHPP. Overall, our proposed methodology uses a concept similar to functional data analysis.

To the best of our knowledge, this is the first paper to propose detecting changes in the sequence $\{N_i\}$ of temporal processes. Previous research studied changes on the IHPP defined on the total time interval where the events are observed, i.e., $t \in [0, T]$, where T represents the total observation time (e.g., one month or one year). In the latter framework, Shen and Zhang (2012); Chernoyarov et al. (2018); Ng and Murphy (2019) all propose methods with slight variations in assumptions or fitting methods. Most of the research papers put changepoints within the intensity function $\lambda(t)$. For example, for a single change point, it is assumed that $\lambda(t)$ has the form $\lambda(t) = \lambda^{(1)}(t)\mathbb{1}\{0 \le t \le \tau_1\} + \lambda^{(2)}(t)\mathbb{1}\{\tau_1 < t \le T\}$. That is, time is linear or a univariate vector. However, in many phenomena, linear changes may not indicate a change in the process itself (noted above).

In summary, this paper proposes a new methodology for detecting changepoints in a sequence of IHPPs. We consider each period (day) to be an observation from an IHPP. The corresponding intensity functions are modelled nonparametrically using finite basis functions. Taking advantage of this low-rank representation, we use a penalized cost approach to detect changepoints.

The remainder of our paper is organized as follows: In Section 2, we present our methodology and the changepoint model. Then, in Section 3, we describe how to estimate the locations and the number of changepoints. Also, we describe how to model and estimate the intensity functions. In Section 4, we conduct a simulation study to evaluate the performance of the proposed methodology. We evaluate the accuracy of detecting changepoints and the accuracy of changepoint locations under different simulation settings. In Section 5, we analyze sensor data of the daily activities of a household. In Section 6, we present some discussion.

2. Methodology

In this section, we detail our proposed method to detect changepoints for event observations.

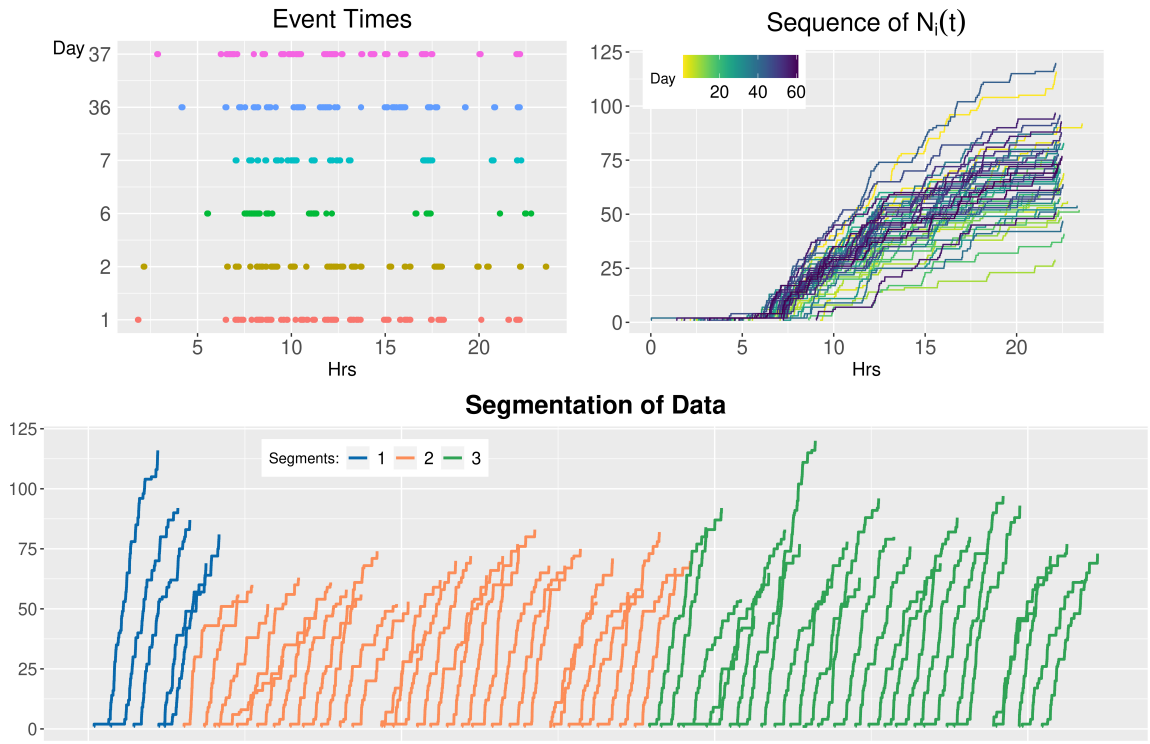

Figure 1: Example of our proposal with Howz data. The top left plot shows six periods of event times. The top right plot shows the complete sequence of IHPPs. Finally, we seek to partition i into consecutive regions and obtain a segmentation of $\{N_i\}$, as shown in the second row of the plot. Each segment should correspond to a specific intensity function.

2.1 *Preliminaries*

Assume that we observe events at times $0 < t_1, \ldots, t_l$ with $t_1 < \ldots < t_l < \delta$ in the interval $[0, \delta]$, where δ represents the duration of a specific period, e.g., 24 hours. We assume that the cumulative count events over time is a realization of an IHPP with intensity function $\lambda : [0, \delta] \to \mathbb{R}^+$. That is, $N(t) := \max\{j : t_j \leq t\}$ is a realization of an IHPP on $[0, \delta]$. With this assumption, for each $t \in [0, \delta]$, $N(t)$ is a random variable representing the random number of points observed in the interval $[0, t]$, and $N(t)$ is Poisson distributed with mean $\int_0^t \lambda(u)\mathrm{d}u$.

Now, assume that we observe time events for several periods. Let n be the total number of periods (e.g., the total number of days) for which we have data. Then, our data is the sequence of IHPPs $\{N_i(t) : t \in [0, \delta], i = 1, \ldots, n\}$, with corresponding intensity functions $\{\lambda_i(t) : t \in [0, \delta], i = 1, \ldots, n\}$. Notice that the counting process N_i only represents events for period i. Figure 1, top right, illustrates realizations of $\{N_i\}$.

2.2 *Model definition*

We now seek to partition $i = 1, \ldots, n$ into consecutive regions. Each segment will be represented by one intensity function. This is, $\{N_i\}$ will be grouped by assuming that they are realizations from the same underlying IHPP.

For $i \leq j$, we denote by $\mathbf{N}_{i:j}$ the set of observed processes from index i to j, i.e., $\mathbf{N}_{i:j} = \{N_i, \ldots, N_j\}$, where $\mathbf{N}_{i:i} = \{N_i\}$. Our model contains m changepoints at positions $\boldsymbol{\tau}_{1:m} = (\tau_1, \ldots, \tau_m)$ with $\tau_1 < \tau_2 < \ldots < \tau_m$. The changepoint vector $\boldsymbol{\tau}_{1:m}$ is a subset of $(1, 2, \ldots, n-1)$. These m changepoints split the data into $m+1$ segments. Let $\tau_0 := 0$ and $\tau_{m+1} := n$, the kth segment contains the trajectory of the processes $\mathbf{N}_{(\tau_{k-1}+1):\tau_k}$, $k = 1, \ldots, m+1$. The statistical problem is estimating the number of changepoints m and their locations.

For data that are scalars, several methods exist to estimate multiple changepoints. One approach is to combine a single changepoint identification method with a binary search (Scott and Knott, 1974). This iteratively applies the method to different subsets of the data. This has a cheap computational cost $O(n \log n)$ but does not guarantee to find the optimal solution (Eckley et al., 2011). An alternative approach is to minimize a cost function for m changepoints. An exhaustive search guarantees the optimal solution but involves considering 2^{n-1} solutions, which is computationally challenging. Recent dynamic programming algorithms have been proposed to overcome this challenge, see, e.g., Killick et al. (2012), and Maidstone et al. (2017). Here, we extend the exact search approach from (Killick et al., 2012), to the sequence of IHPPs $\{N_i\}$.

3. Estimation Method

To estimate the number and locations of changepoints, we have to specify the parameter space for the intensity functions, λ_i, $i = 1, \ldots, n$. We describe the estimation method for these before discussing how to estimate the number, m and locations, $\boldsymbol{\tau}$, of changepoints.

3.1 *Model for the intensity function*

To ensure that λ_i takes values in $\mathbb{R}^+$, we assume that $\lambda_i(t) = \exp\{W_i(t)\}$, where W_i are elements of a space $\mathcal{H}$ of functions defined on the compact interval $[0, \delta]$. The functions λ_i describe how events occur in the time interval $[0, \delta]$, so we would like to model them with minimal restrictions to capture any possible structure. We use a nonparametric approach. Notice that λ_i (or W_i) are defined in an intrinsically infinite-dimensional space, making the estimation procedure challenging. To overcome this challenge, we represent each unknown function W_i as a linear combination of known basis functions. Here, we use cubic B-spline basis function because of its ability to represent local properties and their numeral properties, although other basis functions could equally be used. The number of knots determines the number of the B-spline basis functions. Let $0 = \tilde{t}_1 = \cdots = \tilde{t}_4 < \tilde{t}_5 < \ldots < \tilde{t}_P < \tilde{t}_{P+1} = \cdots = \tilde{t}_{P+4} = \delta$ be the knots in $[0, \delta]$. The number of basis functions will be P. The knot placement will depend on the problem being studied, the most common approaches used are the quantile-based and the equally-spaced knots (Ruppert et al., 2003). For our application, we use the quantile-based method, and this is fixed for all $i = 1, \ldots, n$. Let $\boldsymbol{\psi}(t) = (\psi_1(t), \ldots, \psi_P(t))$ be the associated B-spline basis. Thus, we represent each function W_i as

$$W_i(t) = \sum_{p=1}^{P} w_{p,i}\psi_p(t) = \mathbf{w}_i^\top \boldsymbol{\psi}(t), \tag{1}$$

where $\mathbf{w}_i$ is the vector weights. This low-rank representation is widely used in many contexts, such as functional data analysis, and it can represent complicated shapes of the intensity function. The vector weights $\mathbf{w}_i$ are obtained using the maximum likelihood estimation method by optimising (3) below. With the model (1) for a single period i, we consider how to discriminate between different N_i across $i = 1, \ldots, n$.

3.2 *Penalized cost approach*

The penalized cost approach has two components; a cost function $\mathcal{C}(\mathbf{N}_{i:j})$ associated with a segment of data $\mathbf{N}_{i:j}$, $i \leq j$, and a penalty term β to prevent overfitting. In practice, the common cost functions used are the square-error-loss function (see, eg, Lavielle and Moulines, 2000), cumulative sums (see, e.g., Inclán and Tiao, 1994), and minus twice the log-likelihood (see, eg, Horvath, 1993). For β, the most common choices include Schwarz information criterion (SIC; Schwarz, 1978) and modified Bayesian information criterion (MBIC; Zhang and Siegmund, 2007). Then, the penalized cost function for a segmentation is defined as

$$Q(\mathbf{N}_{1:n}; \boldsymbol{\tau}_{1:m}) = \sum_{k=1}^{m+1} \{\mathcal{C}(\mathbf{N}_{(\tau_{k-1}+1):\tau_k}) + \beta\}. \tag{2}$$

where $\beta > 0$ is the penalty by introducing a changepoint into the model.

Since $\{N_i\}$ is a sequence of IHPPs, the negative log-likelihood is the cost function. In this case $\mathcal{C}(\mathbf{N}_{(\tau_{k-1}+1):\tau_k}) = -\max_\lambda \log L(\mathbf{N}_{(\tau_{k-1}+1):\tau_k}|\lambda)$, where L is the likelihood of the sequence $\mathbf{N}_{(\tau_{k-1}+1):\tau_k}$. For a segment, assume the processes N_i within are independent and identically distributed. Then, $\mathcal{C}(\mathbf{N}_{(\tau_{k-1}+1):\tau_k}) = \min_\lambda \sum_{i=\tau_{k-1}+1}^{\tau_k} \gamma(N_i;\, \lambda)$, where γ is minus the log-likelihood of an IHPP defined in $[0, \delta]$.

Let λ_k be the intensity function of the underlying IHPP for the kth segment. Let us assume that W_i has the form (1). Then,

$$\gamma(N_i;\, \lambda_k) = \int_0^\delta \exp\left\{\sum_{p=1}^P w_{p,k}\psi_p(t)\right\} \mathrm{d}t - \sum_{j=1}^{n_i}\sum_{p=1}^P w_{p,k}\psi_p(t_{i_j}), \tag{3}$$

where n_i is the number of events observed in $[0, \delta]$ for process N_i, and $\{t_{i_1}, \ldots, t_{i_{n_i}}\}$ are the corresponding intraperiod times where events are observed. Thus, the cost for segment k is obtained by minimizing the summation of (3) on the index $i = \tau_{k-1} + 1, \ldots, \tau_k$. Explicitly,

$$\mathcal{C}(\mathbf{N}_{(\tau_{k-1}+1):\tau_k}) = (\tau_k - \tau_{k-1}) \int_0^\delta \exp\left\{\sum_{p=1}^P \widehat{w}_{p,k}\psi_p(t)\right\} \mathrm{d}t - \sum_{i=\tau_{k-1}+1}^{\tau_k}\sum_{j=1}^{n_i}\sum_{p=1}^P \widehat{w}_{p,k}\psi_p(t_{i_j}), \tag{4}$$

where $\widehat{w}_{p,k}$ is such that $\log \widehat{\lambda}_k(t) = \widehat{\mathbf{w}}_k^\top \boldsymbol{\psi}(t)$ is the intensity function that minimizes the sum of the loss functions for the kth segment. Similarly, $\widehat{\lambda}_k(t)$ is the "best" intensity function that represents the data from segment k. So each segment has an estimated intensity $\widehat{\lambda}_k(t)$.

To estimate the number and location of changepoints, we minimize (2):

$$\min_{1 \le \tau_1 < \ldots < \tau_m \le n-1} Q(\mathbf{N}_{1:n}; \boldsymbol{\tau}_{1:m}) \tag{5}$$

Note this minimization is over all possible changepoints and can be computationally challenging. We adopt the PELT method to overcome this challenge, as described in the following.

3.3 *Minimizing the penalized cost*

To minimize the cost function (5), we adopt the PELT method (Killick et al., 2012); a modification of the optimal partitioning (OP) procedure (Jackson et al., 2005).

The basis of OP method is a recursion for the minimum cost of segmenting the sequence of data $\mathbf{N}_{1:q}$, with $q < n$. Let $\mathcal{S}_q = \{\boldsymbol{\tau} : 0 = \tau_0 < \tau_1 < \ldots < \tau_m < \tau_{m+1} = q\}$ be the set of all possible changepoint vectors for data $\mathbf{N}_{1:q}$. Let $F(q)$ be the associated cost of the solution of $Q(\mathbf{N}_{1:q}, \boldsymbol{\tau})$, and define $F(0) = -\beta$. Then, we have that

$$F(q) = \min_{\boldsymbol{\tau} \in \mathcal{S}_q} \left[\sum_{k=1}^m \left\{\mathcal{C}(\mathbf{N}_{(\tau_{k-1}+1):\tau_k}) + \beta\right\}\right] = \min_{p\,:\,p<q} \left\{F(p) + \mathcal{C}(\mathbf{N}_{p+1:q}) + \beta\right\}. \tag{6}$$

Thus, to obtain the minimal cost for the data $\mathbf{N}_{1:q}$, one minimizes over all possible values for the most recent changepoint prior to q. Although this recursion reduces the computational cost from $O(2^n)$ to $O(n^2)$, it is still challenging for large datasets. To gain computational advantages, we adopt the PELT method whose key idea is to discard values of p that cannot never be a minimum in (6). The condition of discarding a candidate changepoint is as follows.

Let $p < q < r$ be three time points. Assume that $\mathcal{C}(\mathbf{N}_{(p+1):q}) + \mathcal{C}(\mathbf{N}_{(q+1):r}) \le \mathcal{C}(\mathbf{N}_{(p+1):r})$. If

$$F(p) + \mathcal{C}(\mathbf{N}_{(p+1):q}) \ge F(q), \tag{7}$$

the candidate p can never be the optimal last changepoint prior to r. Thus, if (7) holds, p is discarded in (6) for all indices larger than q. This can reduce the computational cost significantly; if changepoints occur regularly the computational cost is $O(n)$.

Finally, the recursion (6) is solved using the pruning step for $q = 1, \ldots, n$, and thus we obtain $F(n)$, the minimum value of (2). The set of changepoints at $F(n)$ is the estimator of the changepoint positions within the data. Despite the pruning, PELT remains an exact optimisation method. Thus an exhaustive search and PELT would return the same solution but PELT would complete the optimization orders of magnitude faster ($O(n)$ verses $O(2^n)$).

3.4 *Selecting P*

Notice that P, number of basis functions, is fixed across all segments, and it is assumed to be known in all the above equations. In practice, the value of P needs to be defined. This is a challenge, and it depends on the data. For some basis functions, it is possible to set $P = 1$ and consider a constant basis function, implying that the Poisson process is homogeneous. One way to select P is using AIC. For a given P, estimate individual intensity functions λ_i, $i = 1, \ldots, n$, via maximum likelihood and (1). Obtain the average AIC values over n, $\widehat{\mathrm{AIC}}(P)$. Now, repeat this process, varying P, and set $\widehat{P} = \min_P \widehat{\mathrm{AIC}}(P)$. On average, this should accurately represent the intensity functions. Then, estimate changepoints using $\hat{P}$.

4. Simulation Study

We investigate the performance of our proposed method, NHPP-PELT, under different scenarios. First, we assess the accuracy of detecting the presence of changepoints; summarizing the results using the ROC curves. Second, we demonstrate the accuracy of estimating the changepoint positions. For each scenario, we use cubic splines to represent the intensity function and we

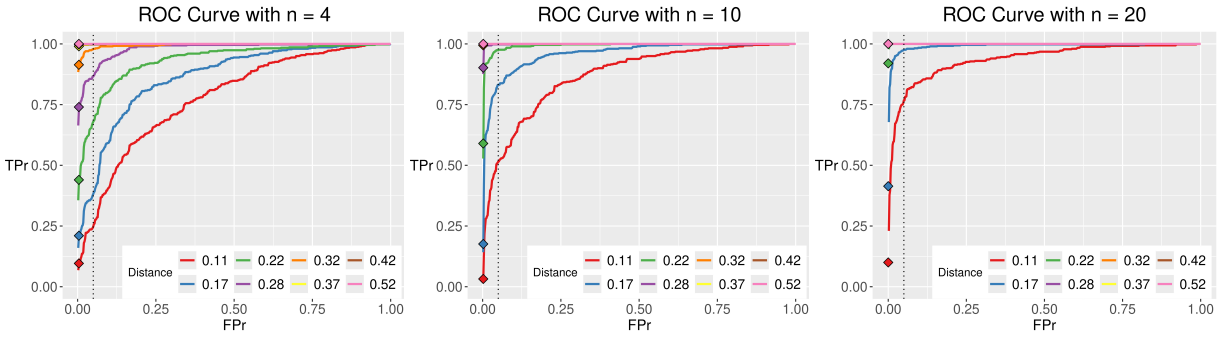

Figure 2: ROC curves for single changepoint detection. The diamond-shaped dots represent the values when using the SIC penalty. The dotted vertical line represents the 0.05 FPr. For $n \geq 10$ and for all distances (except $d = 0.11$ with $n = 10$), a 0.05 proportion of false positives corresponds to a proportion of true positives bigger than 0.8.

compute the average number of correctly estimated and falsely detected $\{\tau_k\}$. Histograms of $\widehat{\tau}_k$ values and boxplots of $\widehat{m}$ values are provided. Additional simulation results for data generated with an ETAS model can be found in the Web Appendix B.2.

4.1 *Simulation setting*

We simulate data as a sequence of IHPPs with m changepoints: $N_1(t), \ldots, N_{\tau_1}(t), N_{\tau_1+1}(t), \ldots, N_{\tau_2}(t), \ldots, N_{\tau_m+1}(t), \ldots, N_n(t)$, $t \in [0, 24]$. The m changepoints represent changes in the intensity functions across days. Thus, we define $m+1$ intensity functions $\{\lambda_0(t), \lambda_1(t), \ldots, \lambda_m(t)\}$, and $N_{\tau_k+1}, \ldots, N_{\tau_{k+1}}$ are simulated from the IHPP with intensity function λ_k, $k = 0, 1, \ldots, m$, defined as

$$\lambda_k(t) = 20\{\phi(t; \mu_k, 2) + \phi(t; \mu_k + 8, \sqrt{8})\}, \quad t \in [0, 24]. \tag{8}$$

Here $\phi(t; \mu_k, \sigma)$ denotes the density of a normal random variable with mean μ_k and standard deviation σ. The different shapes of this intensity function are visualized in Web Figure 1.

To quantify the performance of our method, we define the degree of change using the Hellinger distance, (see Reiss, 1993, Chap. 3). This is convenient in this case of horizontal shift changes but other distances could be used, such as the L_2 distance. See Web Appendix B.1 for changes in the magnitude of the intensity functions. We are interested in the magnitude of change of two consecutive segments, so focus on the distance between λ_k and λ_{k+1}.

The Hellinger distance is defined as $d(\lambda_k, \lambda_{k+1}) = \frac{1}{\sqrt{2}} \left\{ \int \{\tilde{\lambda}_k^{1/2}(t) - \tilde{\lambda}_{k+1}^{1/2}(t)\}^2 \mathrm{d}t \right\}^{1/2}$, where $\tilde{\lambda}_k(t) := \lambda_k(t)/a_k$, with $a_k = \int \lambda_k(u)\mathrm{d}u$. Notice that $d(\cdot, \cdot)$ takes values on $[0, 1]$. If $d(\lambda_k, \lambda_{k+1}) \approx 0$, then $\lambda_k \approx \lambda_{k+1}$ (small change). If $d(\lambda_k, \lambda_{k+1}) = 1$, then λ_k and λ_{k+1} have disjoint domains (large change). In our simulation, we consider different values of d, varying from small to large magnitude changes (see Web Figure 1) alongside varying m and n.

To estimate the intensity functions for each simulation, we use the quantile-based method to place the knots and set $P = 5$ as the number of basis functions in (1). This value is selected based on the AIC criterion (Section 3.4). Each simulation set is replicated 500 times.

4.2 *Accuracy on detecting changepoints*

We simulate data with a single changepoint to look at detection accuracy. We consider sample sizes $n = 4, 10, 20$. These sample sizes are small to illustrate the performance of our method and results are expected to improve as n grows. For the ROC curves (Figure 2), false positive rates (FPr) are calculated from no changepoint simulations and true positive rates (TPr) from data with a single changepoint at $n/2$. ROC curves are obtained by varying the penalty (threshold) value for the detection of a change.

For no changepoints, we simulate $\{N_i(t)\}$ with intensity function $\lambda_0(t)$ (8) with $\mu_0 = 3$. For one changepoint, the first segment $\mathbf{N}_{1:n/2}$ has intensity function $\lambda_0(t)$, and the second segment $\mathbf{N}_{(n/2+1):n}$ has intensity function $\lambda_1(t)$, where $\mu_1 = 4, 4.5, 5, 5.5, 6, 6.5, 7, 8$. For these values of μ_1, the distance $d(\lambda_0, \lambda_1)$ is $0.11, 0.17, 0.22, 0.28, 0.32, 0.37, 0.42, 0.52$, respectively.

The left panel of Figure 2 shows the results when the sample size is $n = 4$. Changes are accurately detected except for $d = 0.11$ and 0.17. This is expected since the magnitude of change is small, coupled with a small sample size. Despite this, we can conclude that for $n = 4$, our method has good overall performance. Increasing the sample size across periods improves results for all change magnitudes. For example, for $d = 0.11$, for $n = 10$, a 0.05 FPr corresponds to a 0.52 TPr. Whereas for $n = 20$, a 0.05 FPr corresponds to a 0.78 TPr.

In conclusion, for small change magnitudes ($d = 0.11$ or 0.17), our method performs well for n larger than 10. If the magnitude of changes is moderate or large ($d \geq 0.22$), the performance of our method is good even if the number of periods is as small as $n = 4$.

Table 1: Average number of true positives over 500 simulations. In parenthesis, the average segment length for each scenario is indicated. Overall, as d or n grows, the result improves.

	$d = 0.17$	$d = 0.22$	$d = 0.28$	$d = 0.32$	$d = 0.37$	$d = 0.42$
m = 2						
$n = 24$ (8)	0.56	1.36	1.79	1.98	2.00	2.00
$n = 50$ (16)	1.41	1.80	1.94	2.00	2.00	2.00
$n = 100$ (33)	1.77	1.90	1.98	2.00	2.00	2.00
$n = 200$ (66)	1.68	1.86	1.96	2.00	2.00	2.00
m = 3						
$n = 24$ (6)	1.29	1.69	2.29	2.67	2.94	3.00
$n = 50$ (12)	1.56	2.39	2.86	2.94	2.99	3.00
$n = 100$ (25)	1.57	2.67	2.87	2.96	3.00	3.00
$n = 200$ (50)	1.80	2.80	2.85	2.95	3.00	3.00
m = 6						
$n = 24$ (3)	2.01	2.24	2.57	3.14	4.37	5.17
$n = 50$ (7)	2.23	3.38	4.34	5.40	5.98	6.00
$n = 100$ (14)	3.14	5.03	5.84	5.96	6.00	6.00
$n = 200$ (28)	4.79	5.47	5.80	5.94	6.00	6.00
m = 12						
$n = 24$ (2)	1.16	3.10	3.94	4.16	4.77	5.45
$n = 50$ (3)	3.73	4.26	4.82	5.51	7.16	8.38
$n = 100$ (7)	4.35	5.37	6.64	8.23	10.94	11.56
$n = 200$ (15)	6.31	8.95	10.94	11.62	12.00	12.00

4.3 *Accuracy of changepoint positions*

Turning to estimated changepoint positions, we simulate data with $n = 24, 50, 100, 200$, and $m = 2, 3, 6, 12$ changepoints. For each value of m, we define the changepoint locations $\boldsymbol{\tau}_{1:m}$ as fixed proportions (floors) of the number of periods n: $(0.2, 0.6)$, $(0.6, 0.7, 0.8)$, $(0.12, 0.25, 0.45, 0.6, 0.75, 0.9)$, and $(0.09, 0.18, 0.23, 0.36, 0.4, 0.5, 0.57,\ 0.65, 0.72, 0.78, 0.83, 0.9)$ for $m = 2, 3, 6$, and 12, respectively. These proportions contain a variety of short and long segments. The smallest segment has one data point and corresponds to the case $m = 12$ and $n = 24$, the largest segment contains 120 points and corresponds to the scenario $m = 3$ and $n = 200$. Given n and $\boldsymbol{\tau}_{1:m}$, we simulate IHPPs $\{\mathbf{N}_{1:\tau_1}, \ldots, \mathbf{N}_{\tau_{m-1}:\tau_m}\}$, where the segment $\mathbf{N}_{\tau_{k-1}:\tau_k}$ has intensity function λ_k. As before, we consider intensity functions such that $d(\lambda_{k-1}, \lambda_k)$ is constant for all $k = 1, \ldots, m$. We use $d = 0.17, 0.22, 0.28, 0.32, 0.42$ and for each value, we define $\{\lambda_0, \ldots, \lambda_m\}$ using combinations of μ_k in (8), such that $d(\lambda_{k-1}, \lambda_k) = d$, for $k = 1, \ldots, m$. For example, for $m = 3$ and $d = 0.22$, we define $(\mu_0, \mu_1, \mu_2, \mu_3) = (6, 7.9, 9.8, 11.8)$. Then, we estimate m and $\{\tau_i\}_{i=1,\ldots,\hat{m}}$ for each dataset simulated. The penalty β is defined as $\beta = (P + 1)\log(n)$ (SIC penalty).

We expect an increase in detection accuracy for larger d. Table 1 shows the average number of changepoints correctly estimated, and Table 2 shows the average number of false positives.

For Table 1, we say that τ_k is correctly estimated if a changepoint is estimated in the interval $(\tau_k - \log n, \tau_k + \log n)$, $k = 1, \ldots, m$. This is a window width of $6.3, 7.8, 9.2, 10.5$ for $n = 24, 50, 100, 200$ respectively. The $\log(n)$ choice is due to the best theoretical rate for consistency of changepoint detection algorithms (Tickle et al., 2020). If one interval intersects another interval, then the length of the intervals involved is reduced equally such that there is no longer intersection. If more than one changepoint is estimated in the interval, one is a true positive, and the remainder false positives. Changepoints estimated more than $\log n$ points from the closest true changepoint are counted as false positives. Table 2, summarises both types of false positives (see Web Table 1 for the split values). In Table 1, we want $\hat{m}$ to be close to m and observe that, as d and n grow, the result improves for all values of m. Recall the use of small sample sizes here to demonstrate the limits of our approach.

Let us analyze the result for each scenario in more detail. For $m = 2$, we observe that the true changepoints are correctly estimated in almost all cases of different sample sizes and magnitude of changes, except for $n = 24$ with $d = 0.17$. The reason may be the sample size of the first segment, which is four, since if the segment sample size increases to ten ($n = 50$), the result gets better. We obtain the same conclusions for $m = 3$ and $m = 6$. If some of the segment sample sizes on the data are smaller or equal to ten (e.g., $m = 3$ and $n = 100$), half of the true changepoints can be missed when the magnitude of change is small, $d = 0.17$. The result improves if the segment sample sizes increase or if the magnitude of change increases. For example, $n = 200$ and $d = 0.22$, the average number of correctly estimating the true changepoints when $m = 3$ is 2.80, and when $m = 6$ is 5.47. Now, let us analyze the scenario with $m = 12$ changepoints. In this case, the method has a poor performance when $n = 24$ and 50; this is mainly because the segment sample sizes vary from 2 to 7. At $n = 100$, we have a good performance with a magnitude of change $d \geq 0.38$. Results improve when $n = 200$, e.g., on average 8.95 true changepoints are correctly

Table 2: Average number of false positives over 500 simulations. In parenthesis, the average segment length for each scenario is indicated. Overall, the method does not produce too many false positives.

	$d = 0.17$	$d = 0.22$	$d = 0.28$	$d = 0.32$	$d = 0.37$	$d = 0.42$
m = 2						
$n = 24$ (8)	0.04	0.02	0.05	0.07	0.13	0.14
$n = 50$ (16)	0.19	0.28	0.45	0.45	0.41	0.51
$n = 100$ (33)	0.85	1.34	1.72	1.57	1.48	1.66
$n = 200$ (66)	2.86	3.78	4.24	3.88	3.55	3.66
m = 3						
$n = 24$ (6)	0.06	0.04	0.03	0.04	0.07	0.25
$n = 50$ (12)	0.50	0.24	0.33	0.35	0.40	0.67
$n = 100$ (25)	1.75	1.30	1.55	1.84	1.55	1.86
$n = 200$ (50)	4.62	3.81	4.27	5.09	4.07	4.12
m = 6						
$n = 24$ (3)	0.00	0.01	0.00	0.00	0.00	0.01
$n = 50$ (7)	0.05	0.06	0.04	0.03	0.04	0.10
$n = 100$ (14)	0.34	0.32	0.27	0.25	0.33	0.69
$n = 200$ (28)	1.28	1.47	2.12	2.28	1.92	2.86
m = 12						
$n = 24$ (2)	0.00	0.00	0.00	0.00	0.00	0.00
$n = 50$ (3)	0.00	0.00	0.00	0.00	0.00	0.01
$n = 100$ (7)	0.02	0.02	0.01	0.01	0.01	0.19
$n = 200$ (15)	0.78	0.32	0.19	0.56	0.30	1.12

estimated with magnitude of change $d = 0.22$ and 10.94 with $d = 0.28$. In the latter scenario, the smallest segment sample size is 10.

To summarize Table 1, our method has good performance in all cases where the segment contains "enough" observations for the size of change in each scenario. Based on our single changepoint simulation studies, that is, at least 10 observations for each segment for small values of d (a complex scenario) and around 5 observations for medium values of d.

Now, let us analyze the results for Table 2. We describe the results according to the sample sizes. We have that, for $n = 24$ and 50, the average false positives are less than 0.67 for all distances and all values of m. Specifically, occasionally only one false positive is observed.

For $n = 100$, we observe that, for $m = 2$ and 3, on average, between 0.85 and 1.86 false positives are observed. Specifically, most of the time, we observed one false positive and sometimes two false positives for all distances. For $m = 6$ and 12, the average number of false positives is less than one. Finally, for $n = 200$, for $m = 2, 3$, and 6, the average number of false positives varies from 1.28 to 5.09, being $m = 3$ the scenario where more false positives are observed (on average 4 or 5). While for $m = 12$, frequently only one and very occasionally two false positives are observed.

Overall, for sample sizes $n = 24, 50$, and 100, we observe 1 or 2 false positives on average for all values of m and all distances. Whereas for $n = 200$, we obtain around 3 to 5 false positives on average with $m = 2$ and $m = 3$, two on average with $m = 6$, and around 1 with $m = 12$. Based on these results and given the complexity of the cases, we think that our method performs well.

In addition, Web Figures 2 and 3 present boxplots of the values of $\widehat{m}$ histograms of the $\widehat{\tau}$ values, respectively. Overall, compared with the results from scalar data, the method is sensitive to small changes and does not produce too many FPs.

5. Data Application

In this section, we analyse sensor data measuring daily activity using the approach described in Section 3. In the following, we only present the results for one person. Web Appendix C presents two more analyses of two different people.

5.1 *Data description*

Data are obtained from Howz (Howz, 2024), and they represent sensor activations by a single older person living in a house. The sensors are triggered when any movement is detected, e.g., in the bedroom, hallway, kitchen, bathroom, and living room. In addition, the sensors capture specific activities that are performed around the house, e.g., using a kettle, a toaster, and a microwave, opening the fridge door, the back door, the front door, and the main door. These sensors record the time they are activated for 61 days. Figure 1 (top left) shows six days of the raw data. We assume that there is no weekend effect. This is reasonable as these are retired adults who have a similar routine every day of the week (see also Web Appendix C.2 for a justification of this assumption).

Detecting changes is important since it allows us to detect potential health problems. In normal conditions, a person is expected to perform their regular routines. For example, waking up, having breakfast, lunch, and dinner around the same time every day. If one day the waking up is pushed back by 30 or 60 minutes, or if the person skips some regular routines, then

this is a change. Depending on the change, it could be interpreted as an indicator of possible health problems, and our goal is to detect these changes. This is challenging because of the inherent periodicity component of the data (24 hrs periodicity). Taylor et al. (2021) assumes that data are Bernoulli time series and detects changes around the circular axis. Changes detected are within period, e.g., wake-up times. These changes cannot be used as an indicator of possible health problems. Instead, we consider identifying changes across days. To achieve this goal, we assume that the number of activities for each day is a realization of an IHPP (top right plot of Figure 1). Note that the number of activities per day is variable; 29 is the minimum, 120 is the maximum, and 70.43 is the average. Since regular daily routines will have similar sensor activations, it is natural to assume that they will have the same intensity function. Whereas if there is a change in the regular routines, the corresponding intensity function should be different.

5.2 *Model specifications*

We assume the model specification as given in Section 2 with $i = 1, \ldots, 61$ days, and $t \in [0, 24]$. For model diagnostic, see Web Appendix C.1.

We segment this sequence according to its intensity functions. We represent the intensity functions as in (1), with $\{\psi_p(t)\}$ being cubic B-spline basis functions. We note that other basis functions could be used here. Based on the AIC criterion, we set $P = 5$. Finally, we define β, the penalty incurred when introducing a changepoint, as $\beta = 6\log(61)$ (SIC penalty). This penalty considers the number of parameters to be estimated when introducing a new intensity function and the sample size. Whilst this standard penalty works well for this application, it may not work in all applications. There are several approaches to penalty selection within the changepoint literature with two common data-driven approaches. The first is the "elbow" approach common in, e.g., choice of significant principal components, one plots the number of changepoints selected against the penalty parameter and looks for the "elbow" in the curve (Lavielle, 2005). This approach works better for signals with a larger number of changepoints to populate the curve. A second approach inspired by supervised clustering uses labelled data (Hocking et al., 2020). Here, segments of data are labelled as "contains a changepoint" or "no changepoint," and a penalty is chosen which balances ensuring that the former segments contain a changepoint and the latter do not. Naturally, this requires a sizeable set of labelled data to provide a sensible choice.

5.3 *Results*

Applying our method, we obtain $\hat{m} = 2$ changepoints at 5, and 35. That is, the segmentation is $\mathbf{N}_{1:5}$, $\mathbf{N}_{6:35}$, and $\mathbf{N}_{36:61}$. Figure 1 shows the segmentation, and Figure 3 (first row) shows the three different intensity functions of the Poisson processes for each segment.

The first difference we observe in the segments is the different number of activities throughout the day. In the first segment, the person is much more active than in the second segment. The number of activities then increases again in the third segment. These overall changes are clearly observed in the plots shown in Figure 3. Let us take a closer look at the difference between two consecutive segments. For this, we present histograms in Figure 3 (second row), which show the proportion of daily events within a segment and grouped per hour. We observed that activities are concentrated from 7 to 1 pm for segment one, with peaks at 8 and 12. Thus, the maximum of $\widehat{\lambda}_0$ is at around 11 am. Whereas for segment two, activities are concentrated from 7 am to 9 am and 12 pm. This gives the maximum of $\widehat{\lambda}_1$ around 8.

The other significant difference between segment one and segment two is there are proportionately more activities around 10 o'clock in segment 1. We can conclude that, after five days, this person decreased their activity and also changed their behaviour, specifically at hours 10am and 1pm. Now, let us compare segments two and three. In segment three, activities increase again. In these two segments, the person has the same morning routine and changes after midday. In segment 2, afternoon, the person's activity decreases drastically and then remains relatively constant, with a slight peak at 5 and 10 pm. In terms of the intensity function, this represents a steady slope after the maximum value is reached, as we can appreciate in Figure 3 (first row, middle plot). On the other hand, in segment 3, we observe more of an increase in activity after 2 pm. This remains relatively constant until 6 pm. This is a more uniform decline after the peak in the intensity function. For segments two and 3, we conclude that the main difference is the activities in the afternoon. In segment 3, more activities are performed from 2 pm to 6 pm than in segment 2.

Finally, we verified with Howz if these two detected changes are related to their findings. Upon visual inspection by Howz, 2 out of 4 other metrics that Howz uses would signal the first changepoint, with a further metric changing at the second changepoint. More specifically, the second changepoint appears to be driven by a change in the time of their start-of-day whilst the first is by a change in their end-of-day routines.

From a medical viewpoint, changes in daily activities may be related to e.g., memory loss or fatigue, since these may cause a person to skip some daily routine activities. When a change is detected in the dataset, the company contacts the person to ask some questions. If necessary, the person is advised to seek medical advice. In general, changepoint detection methods and continuous monitoring of daily activities can help a person's well-being and enhance care and support. It also supports the older person's ability to maintain an independent living.

6. Conclusion and Discussion

An important issue today is the development of new technologies and methodologies to support independent living and healthy ageing of older people. In-home sensors are unobtrusive and future-oriented in collecting data that can be used to study behavioural changes. Tracking behaviour changes provides a valuable monitoring tool. Along this line, we have proposed a new approach to model and detect changes in sensor data measuring daily activity.

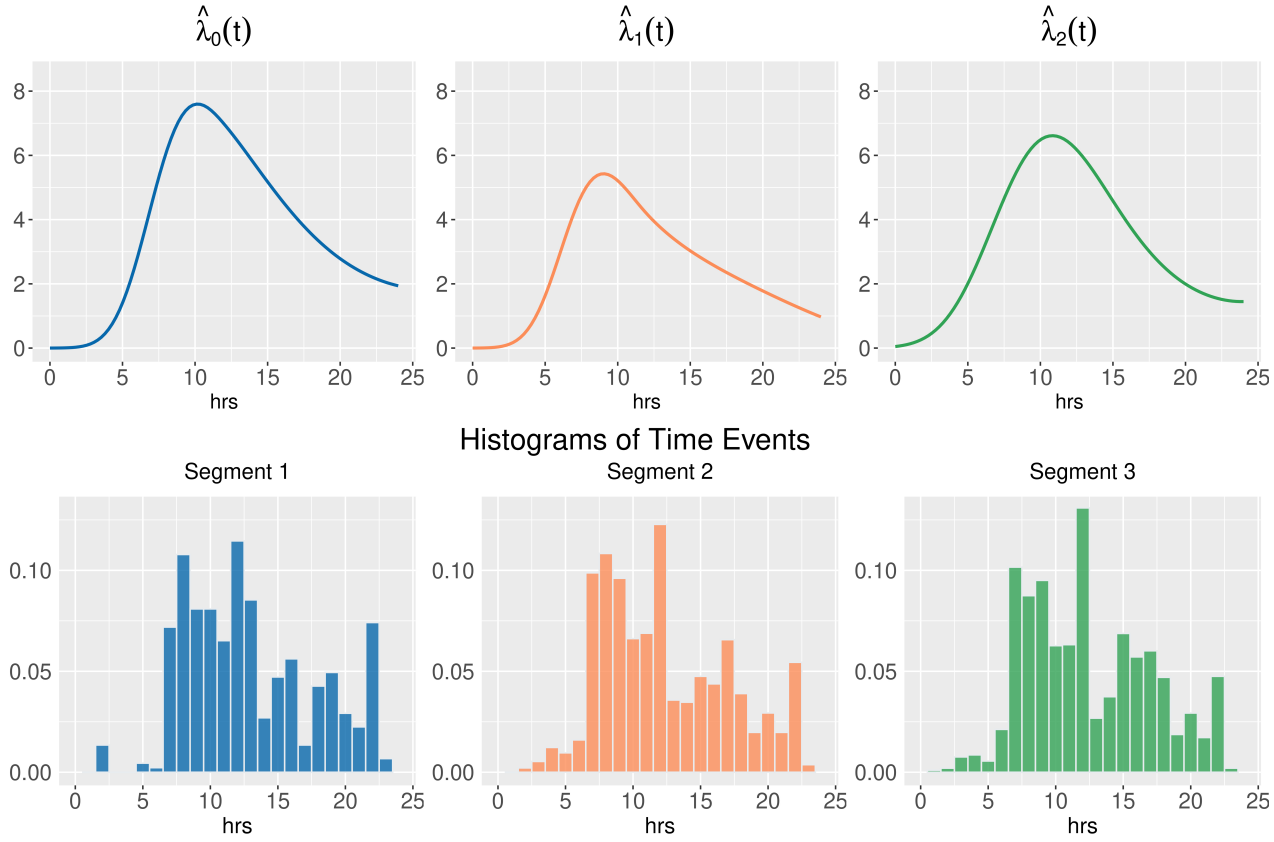

Figure 3: First row: Intensity function of each segment. The maximum of $\widehat{\lambda}_0$, $\widehat{\lambda}_1$, and $\widehat{\lambda}_2$ are attained around 11 am, 8 am, and 9 am, respectively. Second row: Histograms of event times for the three segments. The number of activities are grouped per hour.

Our approach assumes that changes within a day are part of a person's daily routine and focuses on changes across daily behaviour. We assume that the set of event times within a given day is a single observation. This observation is represented by a continuous temporal process. Then, our changepoint model is defined on the sequence of temporal processes. Here, we used B-spline basis functions to represent the daily processes. However, note that other basis functions (e.g., Fourier, wavelets) could be used as appropriate for a given application, especially if continuity at the start/end of each day is desirable.

Our approach can be extended in different directions. For example, one can cluster the sequence of processes according to different alarm levels (habitual behaviours, nonhabitual behaviour, etc.). Also, it allows us to focus on specific periods or intervals of time –for example, daily, weekly or nighttime routines. If data is collected at different houses, we could also cluster households using the individual intensity functions. Moreover, our approach can combine data from different sources and at different resolutions. Another extension would be to use Cox processes or Hawkes processes instead of inhomogeneous Poisson processes; one needs "only" to change the cost function, but this could involve computational challenges. Alternatively, we could consider multivariate processes (including spatial structure). For example, each sensor can be modelled by a point process, and then changes can be detected in the multivariate point process. Again, one could extend the proposed method by modifying the cost function. All these possible extensions are left for future work.

A limitation of the method is that it requires observing a whole period. For example, if a period represents 24 hours, data collected until mid-day cannot be used since we will not be able to estimate the intensity function. Another limitation is that the proposed method requires a "large" sample size to detect small changes. For example, from Figure 2 with $d = 0.11$, a sample size bigger than 20 is needed to accurately detect a changepoint with lower false positive rates. Although this is a very small change (see Web Figure 1), and in this case, any method will require a large sample size to obtain accurate results. A more realistic small change would be $d = 0.17$ ($d = 0.22$); in this case, the method requires a sample size bigger than 10 (6) to accurately detect a changepoint with a lower false positive rate (around 0.05). Thus, we recommend the user set the minimum size of a segment according to the context (this is an argument in the `R` function that estimates m and $\boldsymbol{\tau}$).

We conclude that our methodology presents a valuable alternative to detect change points when data are a sequence of point processes. The corresponding R code to estimate the changepoints will be available as part of the *changepoint* R package (Killick et al., 2022). Alternatively, R code can be found on the author's GitHub.

Acknowledgments

The authors are grateful to Ben Norwood from Howz for providing the dataset and helpful discussions around the results. Also, the authors thank the associate editor and the referees for their thoughtful and constructive comments and suggestions. Finally, the authors acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC) grant EP/T021020/1.

Supplementary Materials

Web Appendices, Tables, Figures, and code referenced in Sections 4 and 5 are available with this paper at the Biometrics website on Oxford Academic.

Data Availability

The data used in this paper are available from Howz (2024). Restrictions apply to the availability of these data, which were used under license in this paper.

References

Austin, D., T. L. Hayes, J. Kaye, N. Mattek, and M. Pavel (2011). Unobtrusive monitoring of the longitudinal evolution of in-home gait velocity data with applications to elder care. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6495–6498.

Baddeley, A. (2007). Spatial point processes and their applications. In Stochastic geometry, Volume 1892 of Lecture Notes in Math., pp. 1–75. Springer.

Chernoyarov, O. V., Y. A. Kutoyants, and A. Top (2018). On multiple change-point estimation for Poisson process. Commun. Stat. Theor. 47(5), 1215–1233.

Cooper, J., W. Taylor, M. Imran, and Q. Abbasi (2024). Smart fatigue monitoring using RF sensing employing a deep learning convolutional neural network. In IEEE International Symposium on Antennas and Propagation and ITNC-USNC-URSI Radio Science Meeting.

Eckley, I. A., P. Fearnhead, and R. Killick (2011). Analysis of changepoint models. In Bayesian Time Series Models, pp. 205–224. Cambridge Uni Press.

Elsayed, M., N. Ghadban, Q. Abbasi, and J. L. Kernec (2024). Case study of radar-based vital signs monitoring and the effect of target aspect angle. In IEEE International Symposium on Antennas and Propagation and ITNC-USNC-URSI Radio Science Meeting.

Hocking, T. D., G. Rigaill, P. Fearnhead, and G. Bourque (2020). Constrained dynamic programming and supervised penalty learning algorithms for peak detection in genomic data. J. Mach. Learn. Res. 21(87), 1–40.

Horvath, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. Ann. Stat. 21(2), 671–680.

Howz (2024). Lloyd street north Manchester Science Park, Manchester, M15 6SE. https://www.howz.com [Accessed: September 2024].

Inclán, C. and G. C. Tiao (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. J. Am. Stat. Assoc. 89(427), 913–923.

Jackson, B. et al. (2005). An algorithm for optimal partitioning of data on an interval. IEEE Sig. Proc. Let. 12(2), 105–108.

Killick, R., P. Fearnhead, and I. A. Eckley (2012). Optimal detection of changepoints with a linear computational cost. J. Amer. Stat. Assoc. 107(500), 1590–1598.

Killick, R., K. Haynes, and I. A. Eckley (2022). changepoint: An R package for changepoint analysis. R package version 2.2.4.

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. Sig. Proc. 85(8), 1501–1510.

Lavielle, M. and E. Moulines (2000). Least-squares estimation of an unknown number of shifts in a time series. Journal of Time Series Analysis 21(1), 33–59.

Maidstone, R., T. Hocking, G. Rigaill, and P. Fearnhead (2017). On optimal multiple changepoint algorithms for large data. Statistics and Computing 27, 519–533.

Ng, T. L. J. and T. B. Murphy (2019). Estimation of the intensity function of an inhomogeneous Poisson process with a change-point. Can. J. Stat. 47(4), 604–618.

QUEST (2024). QUEST Project. https://healthcare-quest.com [Accessed: September 2024].

Reiss, R.-D. (1993). A course on point processes. Series in Statistics. Springer-Verlag.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). Semiparametric Regression, Volume 12 of Series in Statistical and Probabilistic Mathematics. Cambridge Uni Press.

Schwarz, G. (1978). Estimating the dimension of a model. Ann. Stat. 6(2), 461–464.

Scott, A. J. and M. Knott (1974). A cluster analysis method for grouping means in the analysis of variance. Biometrics 30(3), 507–512.

Shen, J. J. and N. R. Zhang (2012). Change-point model on nonhomogeneous Poisson processes with application in copy number profiling by next-generation DNA sequencing. The Annals of Applied Statistics 6(2), 476–496.

Taylor, S. A. C., R. Killick, J. Burr, and L. Rogerson (2021). Assessing daily patterns using home activity sensors and within period changepoint detection. JRSS C 70(3), 579–595.

Tickle, S. O., I. A. Eckley, P. Fearnhead, and K. Haynes (2020). Parallelization of a common changepoint detection method.

J. Comp. Graph. Stat. 29(1), 149–161.

Usman, M., J. Rains, T. Cui, M. Khan, J. Kazim, M. Imran, and Q. Abbasi (2022). Intelligent wireless walls for contactless in-home monitoring. Light: Sci. App. 11.

Zhang, N. R. and D. O. Siegmund (2007). A modified Bayes information criterion with applications to the analysis of CGH data. Biometrics 63(1), 22–32.