Assessing Daily Patterns using Home Activity Sensors and Within Period Changepoint Detection

Simon A. C. Taylor

School of Mathematics, University of Edinburgh, Edinburgh, EH9 3FD.

E-mail: simon.taylor@ed.ac.uk

Rebecca Killick

Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF.

Jonathan Burr and Louise Rogerson

Intelesant Ltd., Manchester, M15 6JJ

Summary. We consider the problem of ascertaining daily patterns using passive sensors to establish a baseline for elderly people living alone. The data are whether or not some movement, or human related activity, has occurred in the previous 15 minutes. We seek to segment the broad patterns within a day, e.g, awake/sleep times or potentially more activity around meal-times. To address this problem we use changepoint detection which can segment the day into more/less active times. Traditional changepoint detection methods are inappropriate for this data as they fail to utilize the periodic nature of the data. The traditional assumption of conditional independence of the segments also hampers estimation of the within segment parameters. A new within-period changepoint detection scheme is proposed that instead assumes a circular perspective of the time axis. This permits the pooling of evidence of changepoint events from across multiple days. Inference is performed within the Bayesian framework by utilising the reversible jump Markov chain Monte Carlo sampler to explore the variable dimension parameter space. Simulations demonstrate that the sampler achieves high accuracy in approximating the posterior whilst being able to detect small segments. Application to four individuals from our industrial collaborator provides insights to their daily patterns.

Keywords: Changepoint analysis; Periodic time; Reversible jump Markov chain Monte Carlo; Home activity sensors;

1. Introduction

We live in a world where older generations are often living further away from their children and where over 4m people over 65 in the UK are living alone (ONS, 2019). In light of this reality, families are seeking a peace of mind that their elderly relatives are living well and maintaining an active lifestyle. This paper proposes a method to ascertain broad patterns of awake/sleep times and other strong within-day routines e.g., more activity around meal-times. This problem is motivated from an investigation with our industrial collaborator, Intelestent Ltd., into assessing the daily activity patterns from passive sensors within the home.

When people think about activity detection it is common to think of a wearable device that samples at 100-500 observations per second, but this relies upon the users wearing such devices 24-7. In contrast, our data arise from static infra-red sensors more commonly used in home security systems. The average daily activity, over 56 days, of the motion/non-motion measurements, in 15 minute bins, for four individuals are presented in Figure 1. These demonstrate a pattern of abrupt changes that can be attributed to certain regular activities such as sleep, morning routine, mealtime etc. Identification of the times of day where abrupt changes occur between different types of activity can be investigated using changepoints.

Changepoint analysis involves the detection and estimation of the time(s) within a non-stationary signal where the underlying data generation process experiences some abrupt shift. Typically, time is interpreted in a linear sense whereby the data generating process before and after each changepoint event are considered to be independent. This is the natural assumption for many changepoint methods that have been applied to a large number of fields (Truong et al., 2018). Contrary to existing methods, this paper proposes a method to investigate the within-period changepoint structure of periodic time series - where time is not linear.

Traditional methods to investigate the changepoint structure within a binary time series (see Chen and Gupta (2000)) assume that time is linear and so the estimated changepoint events are unlikely to align across days or may fail to be identified due to limited evidence. One method to ensure alignment of events could be to partition the time series into separate periods (days) and perform a multivariate analysis (Zhang et al., 2010). The main issue with this approach is that the arbitrary selection of a point within the period to partition the data, such as midnight, in effect artificially defines a changepoint event which may impede the estimation of nearby true events, such as when they go to sleep.

The limiting factor for the application of traditional changepoint methods is the underlying assumption that time is linear. This paper instead considers a circular perspective of the time axis to achieve the required conditional independence between segments whilst pooling evidence from multiple periods. The circular perspective also avoids the necessity to define a changepoint event at the start of each period as the axis wraps around on itself. It is worth emphasising that this paper applies the circular perspective to the time axis across a period rather than the response variable. This is distinctively different to Jammalamadaka and Sengupta (2001) and Price-Williams et al. (2017) where their objective was to estimate the mean shift in directional response data within the traditional linear time perspective. It is also distinctive from circular binary segmentation (Venkatraman and Olshen, 2007), which assumes that the start of the end of the time series have similar behaviour in order to estimate an anomalous segment (two changepoints), i.e. observing only a single cycle.

Methods to perform multiple changepoint inference have been developed for both Classical and Bayesian approaches. Notable algorithms within the Classical framework are binary segmentation (BS) (Scott and Knott, 1974) and the pruned exact linear time (PELT) method (Killick et al., 2012). Both methods utilise the conditional independence structure of the linear time perspective by either branch-

ing or sequentially trimming the time series into independent segments. In the Bayesian framework, filtering (Fearnhead, 2006) provides a similar mechanism for inference as BS and PELT that requires independence of parameters across consecutive changepoint events. Alternatively, hidden Markov models (HMM) (Rabiner, 1989) approach changepoint analysis under the assumption that there exists some unobserved process that switched between different states according to some unknown transition matrix. This structure permits re-occurrence of certain states and Pierson et al. (2018) propose the cyclic hidden Markov model (CyHMM) where the transition matrix is highly structured to provide a periodic mechanism. The two main disadvantages of this technique is that the number of hidden states must be known, although assessment via cross-validation can be performed, and the overall period length is stochastic. This is not the case for the home sensor time series as it presents a strong daily periodic pattern.

A further technique in approaching changepoint analysis within a Bayesian framework is the reversible jump Markov chain Monte Carlo (RJMCMC) (Green, 1995) sampler. Although not exclusively designed for changepoint analysis, the RJMCMC sampler is highly useful in this field. It is a generic method that permits inference across multi-dimensional spaces in which more/fewer parameters are required when considering the need for more/fewer segments to explain the available data.

The outline of this paper is as follows. Section 2 presents the within-period changepoint methodology by first setting out notation, terminology and the statistical model. The prior distribution for the unknown segment and changepoint parameters are subsequently specified along with derivation of their posterior distribution. The inference mechanism via the RJMCMC sampler is designed in Section 3, presenting the approach to navigate the variable dimension parameter space and derivation of the proposal distribution. The performance of the developed sampler is examined using simulations in Section 4. The case study in Section 5 examines home activity time series and discusses the potential insight into the daily patterns of four individuals. Section 6 concludes the paper with a discussion of the performance of the method in assessing within-period changepoint structure as well as potential future directions for research.

2. Methodology

This section lays the foundations of the changepoint model and details the inference mechanism within the Bayesian framework.

2.1. *Periodic binary changepoint model*

Let $\{X_t\}_{t=1,\ldots,T}$ be a sequence of binary random variables where each are independently Bernoulli distributed conditional on some unknown probability. We assume that this probability, $p(t)$, is a periodic piecewise constant function of time with a specified fixed period length of N (e.g., daily). In a single reference period the function consists of m segments taking the values $\phi_0, \ldots, \phi_{m-1}$ ending at the corresponding changepoint event $(0 <)\tau_0 < \ldots < \tau_{m-1}(\leq N)$. Periodicity is achieved
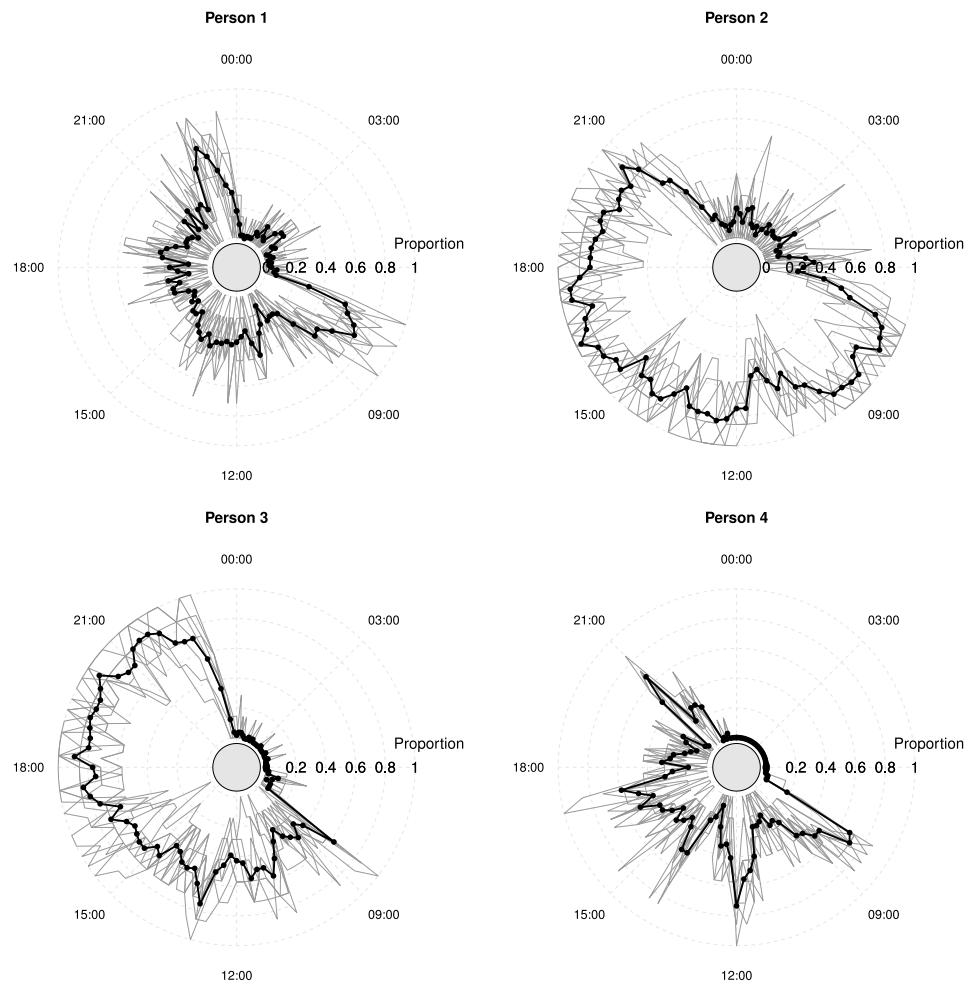
Fig. 1. Activity data from four individuals measured at 15 minute intervals over 56 days (8 weeks). Black: overall sample sample proportions, grey: sample proportions evaluated for each week.
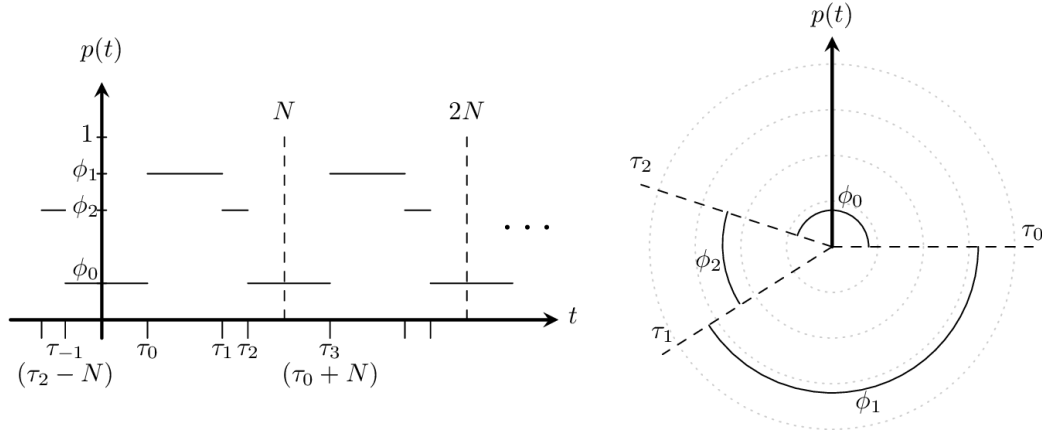
Fig. 2. Illustration of function $p(t)$ for $m = 3$ segments on the linear time axis (left) and the circular time axis (right).

by specifying that the function takes the value ϕ_0 after the τ_{m-1} event to match the beginning of the reference period. The changepoint events in other periods are notated according to the relationship $\tau_{i+jm} = \tau_i + jN$ for all $j \in \mathbb{Z}$. This means that the event τ_m describes the first changepoint in the next period whilst τ_{-1} represents the last event in the previous period, which are equivalent to $\tau_0 + N$ and $\tau_{m-1} - N$ on the linear time axis or τ_0 and τ_{m-1} within the circular reference period respectively. Therefore, we define the part intervals at the beginning and end of the period, namely $(0, \tau_0]$ and $(\tau_{m-1}, N]$, to represent the same within period segment.

It follows that the probability function is given by:

$$p(t) = \sum_{i=0}^{m-1} \phi_i \chi_{(\tau_{i-1}, \tau_i]}(t \bmod N), \tag{1}$$

where the indicator function $\chi_A(x)$ returns 1 if $x \in A$ or 0 otherwise, and the modulus operator $a \bmod N$ returns the remainder of a after division by N. Figure 2 illustrates the structure of $p(t)$ in both the linear and circular time axes.

An additional constraint is applied to the changepoint events such that the duration between any consecutive pair is greater than some minimum, $\tau_i - \tau_{i-1} \geq l$. The value for $l \geq 1$ is specified based on the context of the application. For example, in Section 5 it is assumed that individuals undertake a particular level of activity for at least an hour and so l is specified appropriately given data frequency.

$$p(t) = \sum_{i=0}^{m-1} \phi_i \chi_{(\tau_{i-1}, \tau_i]}(t \mod N), \tag{2}$$

In summary, the sequence of binary observations is assumed to be partitioned into m groups based on the relative temporal position within the period. On any

given segment, say $(\tau_{i-1}, \tau_i]$, the observations are independent and identically distributed Bernoulli random variables with probability ϕ_i irrespective of whether they occur in the same or different periods e.g., days. The target for inference are the number of segments that partition the period interval, m, the changepoint events, $\boldsymbol{\tau}_m = (\tau_0, \ldots, \tau_{m-1})$, and the segment probabilities, $\boldsymbol{\phi}_m = (\phi_0, \ldots, \phi_{m-1})$. It follows that the likelihood function based on an observed sequence $\boldsymbol{x} = (x_1, \ldots, x_T)$, for some $T \geq N$, is given by:

$$f(\boldsymbol{x}|\boldsymbol{\phi}_m, \boldsymbol{\tau}_m, m) = \prod_{i=0}^{m-1} \phi_i^{s_i}(1 - \phi_i)^{n_i - s_i}, \tag{3}$$

where $n_i = \sum_t \chi_{(\tau_{i-1}, \tau_i]}(t \mod N)$ and $s_i = \sum_t x_t \chi_{(\tau_{i-1}, \tau_i]}(t \mod N)$ respectively denote the number and sum of observations on the ith segment.

Note that if there is only one segment, $m = 1$, then the probability function is constant: $p(t) = \phi_0$ for all t. Consequently there are no changepoint events and so we denote $\boldsymbol{\tau}_1 = \emptyset$. Conversely, the most number of segments that can be fit into the period whilst respecting the minimum segment length constraint is $m_{\max} = \lfloor N/l \rfloor$, where $\lfloor y \rfloor$ denotes the integer part of y.

2.2. Prior distributions

The unknown number of segments m that span the period dictates the dimension of the model being defined. For parsimonious reasons it is desirable to place a higher preference on a simple model specification whenever the evidence from the data equally support two plausible cases. As such, the number of segments in the model is assigned a Poisson(1) prior distribution that is truncated to the interval $m = 1, \ldots, m_{\max}$.

Conditional on the total number of segments, the prior distribution for the changepoint positions is expressed with respect to the transformation $\delta_i = \tau_i - \tau_{i-1} - l$ representing the excess segment length after considering the minimum length constraint. The vector $\boldsymbol{\delta}_m = (\delta_0, \ldots, \delta_{m-1})$ therefore describes how the remaining $\Delta_m = N - lm$ time points within the period can be randomly partitioned into the m segments. A natural prior distribution for $\boldsymbol{\delta}_m$ is the Dirichlet-multinomial distribution with probability distribution function:

$$\pi(\boldsymbol{\delta}_m|m) = \frac{\Delta_m! \, \Gamma(m\gamma)}{\Gamma(\Delta_m + m\gamma) \, \Gamma(\gamma)^m} \prod_{i=0}^{m-1} \frac{\Gamma(\delta_i + \gamma)}{\delta_i!}. \tag{4}$$

The shape parameter γ describes the prior preference as to how the changepoint events are to be spread throughout the period interval. If $\gamma < 1$ then there is a prior preference to cluster the changepoint events, whereas $\gamma > 1$ expresses a preference for the changepoint to be more evenly spread. In balance, a specification of $\gamma = 1$ represents a uniform distribution. Note that (4) is a special case of the Dirichlet-multinomial distribution where a common shape parameter is specified for each element of $\boldsymbol{\delta}_m$. It is not possible to be more *a priori* specific about the

clustering/repelling of particular groups of changepoint events when the overall number of changepoints is unknown.

The prior distribution for the changepoint positions themselves is complete by specifying a discrete uniform distribution for an anchoring event, $\pi(\tau_A) = \frac{1}{N}$. This anchoring point allows all other changepoint events to be determined according to $\tau_i^* = \tau_A + \sum_{j=0}^{i}(\delta_j + l)$. The marginal prior $\pi(\tau_i|m)$ for each segment i is uniformly distributed and so there is a potential identifiability issue. This is addressed by permuting the vector of changepoints such that τ_0 corresponds to that which has the smallest reminder after division of the period length N. The Jacobian associated to the transformation $(\boldsymbol{\delta}_m, \tau_A) \to \boldsymbol{\tau}_m$ is equal to one.

The above prior specification for the changepoint vector $\boldsymbol{\tau}_m$ is only applicable when $m > 1$. In the unique case of a single segment model, $m = 1$, the probability function (1) is constant and no changepoint events exist. For completeness, a point-mass prior distribution on the empty set is applied: $\pi(\boldsymbol{\tau}_m|m = 1) = \chi_\emptyset(\boldsymbol{\tau}_m)$.

Conditional on the number of changepoint events, the prior distribution for the segment probabilities are independent uniform distributions on the unit interval: $\pi(\phi_i|\boldsymbol{\tau}_m, m) = \chi_{[0,1]}(\phi_i)$ for $i = 0, \ldots, m - 1$.

As we have now described the building blocks; the likelihood and the prior, the following section elicits the posterior distribution.

2.3. Posterior distribution

The aim for inference is to estimate the posterior distribution for the number and position of changepoint events and the segment probabilities:

$$\pi(\boldsymbol{\phi}_m, \boldsymbol{\tau}_m, m|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\phi}_m, \boldsymbol{\tau}_m, m)\pi(\boldsymbol{\phi}_m|m)\pi(\boldsymbol{\tau}_m|m)\pi(m) \tag{5}$$

By factorising the posterior as $\pi(\boldsymbol{\phi}_m, \boldsymbol{\tau}_m, m|\boldsymbol{x}) = \pi(\boldsymbol{\phi}_m|\boldsymbol{x}, \boldsymbol{\tau}_m, m)\pi(\boldsymbol{\tau}_m, m|\boldsymbol{x})$ it is clear from the likelihood function (3) and uniform prior that the conditional posterior for each of the segment probabilities follows the beta distribution:

$$\boldsymbol{\phi}_m|\boldsymbol{x}, \boldsymbol{\tau}_m, m \sim \text{Beta}(1 + s_i, 1 + n_i - s_i). \tag{6}$$

The marginal posterior for the number and position of changepoint events therefore follows from marginalising $\boldsymbol{\phi}_m$, and is concisely expressed as:

$$\pi(\boldsymbol{\tau}_m, m|\boldsymbol{x}) \propto \begin{cases} B\left(1 + \sum_{t=0}^{T} x_t, 1 + T - \sum_{t=0}^{T} x_t\right) & \text{if } m = 1, \\ \prod_{i=0}^{m-1}\left[\dfrac{B(1 + s_i, 1 + n_i - s_i)}{(\gamma + \delta_i)\ B(\gamma, \delta_i + 1)}\right] \dfrac{(m\gamma + \Delta_m)\ B(m\gamma, \Delta_m + 1)}{m!\ N} & \text{otherwise,} \end{cases} \tag{7}$$

where $B(a, b)$ denotes the beta function.

Although evaluation of (7) is trivial to compute for a specific changepoint vector, $\boldsymbol{\tau}_m$, the challenge to inference occurs because the size of the support for the posterior is potentially very large. The total number of possible changepoint vectors for a given period length, N, and minimum segment length l, is bounded from above by the higher-ordered Lucas number ${}^l L_N$ (Randic et al., 2008). For example, the size of

the posterior support for problems examined in Section 5, where $N = 96$ and $l = 4$, contains 27.34×10^{12} possible cases for $\boldsymbol{\tau}_m$. Thus the following section proposes an efficient way to explore the parameter space of the posterior (7) from which inference on the number and positions of changepoint events, and subsequently segment probabilities, can be evaluated.

3.   Reversible-jump Markov chain Monte Carlo sampler

The goal of the sampler is to approximate the posterior mass function (7) for the changepoint positions within N-periodic binary data where the segments are subject to a minimum length constraint. Direct evaluation of the posterior is not practically feasible because of the size of the sample space. In addition, the dimension of each sample, namely the number of segments to partition the period, is variable. Due to this, a reversible jump Markov chain Monte Carlo (RJMCMC) (Green, 1995) is developed in order to explore the parameter space.

3.1.   Proposal distribution

Given the current sample of the changepoint vector $\boldsymbol{\tau}_m$, a proposal vector $\boldsymbol{\tau}'_{m'}$ for which to consider transitioning to is generated according to the proposal distribution $q(\boldsymbol{\tau}'_{m'}|\boldsymbol{\tau}_m)$. There are many different ways that this proposal distribution could be constructed. The basic properties required are that i) the number of changepoints should be able to vary, and ii) changepoint locations can be added, deleted and moved. We choose to create a proposal distribution that satisfies these by acting on a single segment, i as follows.

$$q(\boldsymbol{\tau}'_{m'}|\boldsymbol{\tau}_m) = \sum_{i=0}^{m-1} q_1(i|\boldsymbol{\tau}_m)q_2(\boldsymbol{\tau}'_{m'}|\boldsymbol{\tau}_m, i) \tag{8}$$

Here $q_1(i|\boldsymbol{\tau}_m)$ generates a segment index based on the current partitioning of the period, and $q_2(\boldsymbol{\tau}'_{m'}|\boldsymbol{\tau}_m, i)$ describes the range of possible perturbations to make to the current changepoint vector with respect to the aforementioned sampled segment (add, delete, move).

The proposal segment index is uniformly sampled from the segment indices of the current sample, $q_1(i|\boldsymbol{\tau}_m) = 1/m$ for $i = 0, \ldots, m - 1$.

Before defining the proposal $q_2(\boldsymbol{\tau}'_{m'}|\boldsymbol{\tau}_m, i)$ it is necessary to first examine the method for augmenting the current changepoint vector in defining the support for the proposal distribution by restricting the search space to the current, one more or one fewer number of changepoints. Let the set $\mathcal{P}(\boldsymbol{\tau}_m, i)$ denote the support for q_2 that is based upon the current changepoint vector and the sampled segment according to q_1. As we are modelling circular time series, we must consider transitioning from $m = 1$ and $m > 1$ separately.

When the current sample defines the single segment model, $m = 1$, there are no changepoint events defined, $\boldsymbol{\tau}_1 = \emptyset$. It is clear that it is not possible to further remove changepoints and also that the current changepoint event is the possibility

for the single segment model. So, $\boldsymbol{\tau}'_{m'=1} = \emptyset$ must be within the proposal set $\mathcal{P}(\boldsymbol{\tau}_1, 0)$. To propose a move to the two-segment model, $m' = 2$, requires the introduction of two changepoint events. As there is currently no existing knowledge about the position of either changepoint event, the proposal set must contain all possible pairs of the positions that spans the $m = 2$ model. Hence, $\mathcal{P}(\boldsymbol{\tau}_1, 0) = \{\boldsymbol{\tau}'_{m'} | m' \leq 2\}$.

For the general $m > 1$ case, augmentation of the current changepoint vector can be viewed through the typical change or birth-death updating procedure (Green, 1995). In summary, the possibilities for proposing $\boldsymbol{\tau}'_{m'}$ from $\boldsymbol{\tau}_m$ would be to either move the changepoint event, τ_i, that corresponds to the sampled ith segment to a new position, combine the ith segment with the $(i + 1)$th segment by removing the event τ_i, or introduce a new changepoint event that partitions the sampled segment into two. This method would be applicable if there were no minimum segment length constraint, $l = 1$, but navigation of the parameter space using these update steps can be impeded when $l > 1$. For example, consider the case when $N = 10$ and the changepoint vectors $(2, 4, 9)$ and $(3, 9)$ are highly probable events. The transition between these two events must take at least two iterations under the aforementioned scheme. The number of shortest paths is four when $l = 1$ (via $(2, 9)$, $(4, 9)$, $(3, 4, 9)$ or $(2, 3, 9)$), which is reduced to two paths when $l = 2$ (via $(2, 9), (4, 9)$). If the transition probability to these intermediary events are low then there is a risk that the chain could get stuck at one of the highly probable events or that the most efficient path between the events involves more iterations and so the chain becomes highly autocorrelated.

These issues arise because of the pruning effect the minimum segment length condition has on the number of potential paths the RJMCMC could take. To increase the set of potential paths we consider an additional movement to the next changepoint event τ_{i+1} in the birth/death moves up to l timepoints. To elaborate further, consider the transition from $\boldsymbol{\tau}_m$ to $\boldsymbol{\tau}'_{m'}$ where $m' = m + k$ for $k \in \{-1, 0, 1\}$. We select ith segment according to q_1 in (8) as usual. Then the proposal $\boldsymbol{\tau}'_{m'}$ is drawn from the set $\mathcal{P}(\boldsymbol{\tau}_m, i)$ which is formed according to:

Death $(k = -1)$**:** If $m = 2$, then $\boldsymbol{\tau}'_{m'} = \emptyset$, otherwise $\boldsymbol{\tau}'_{m'} = (\ldots, \tau_{i-1}, \tau_{i+1} - u, \ldots)$ for $u = 0, \ldots, l - 1$.

Move $(k = 0)$**:** $\boldsymbol{\tau}'_{m'} = (\ldots, \tau_{i-1}, v, \tau_{i+1}, \ldots)$ for $v = \tau_{i-1} + l, \ldots, \tau_{i+1} - l$.

Birth $(k = 1)$**:** If $\tau_i - \tau_{i-1} > l$ and $\tau_{i+1} - \tau_{i-1} \geq 3l$ then $\boldsymbol{\tau}'_{m'} = (\ldots, \tau_{i-1}, w_1, \tau_i + w_2, \tau_{i+1}, \ldots)$ for $w_1 = \tau_{i-1} + l, \ldots, \tau_i + w_2 - l$ and $w_2 = \max(0, \tau_{i-1} - \tau_i + 2l), \ldots, \min(l - 1, \tau_{i+1} - \tau_i - l)$. Partition the sampled segment into two by introducing a new changepoint event whilst also considering movement of the sampled event, provided that the minimum length constraint is satisfied.

Figure 3 depicts the sample space for the proposal changepoints τ'_i and τ'_{i+1} in the birth step in relation to the current changepoint vector and the minimum segment length constraint. Note that when the minimum segment length condition is removed in the above scheme, i.e. setting $l = 1$, results in a set representing the standard birth-move-death proposals without perturbation.
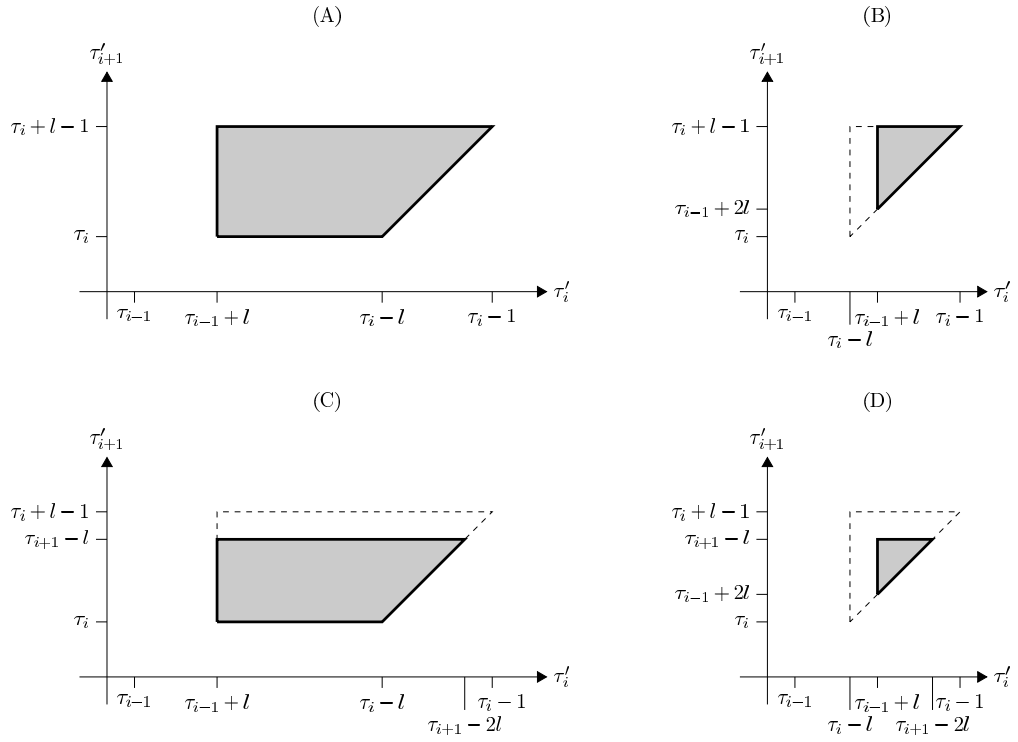
Fig. 3. Depictions of the sample space for the proposed changepoints τ_i' and τ_{i+1}' under the birth case, showing the dependence on the separation of the current changepoint values τ_{i-1}, τ_i and τ_{i+1} and a specified minimum segment length l. (A) $\tau_{i+1} + 2l \leq \tau_i \leq \tau_{i+1} - 2l$. (B) $\tau_i < \tau_{i-1} + 2l$. (C) $\tau_i > \tau_{i+1} - 2l + 1$. (D) $\tau_i < \tau_{i-1} + 2l$ and $\tau_i > \tau_{i+1} - 2l + 1$. The dashed regions represent the space corresponding to case (A) but which must be truncated due to the respective conditions such that the proposal changepoint vector satisfies the minimum length condition. Note that there are no possible proposal birth events when $\tau_i = \tau_{i-1} + l$ or $\tau_{i+1} < \tau_{i-1} + 3l$.

The size of the proposal set $\mathcal{P}(\boldsymbol{\tau}_m, i)$ according to the above rules is less than $(l+1)N$ for $m > 1$ or is less than $\frac{(N-l+1)^2}{2}$ when $m = 1$. The set size for the case studies in Section 5 is sufficiently small such that evaluations of the proportional posterior surface (7) for every $\boldsymbol{\tau}'_{m'} \in \mathcal{P}(\boldsymbol{\tau}_m, i)$ is not a costly computational operation. As such, the proposal distribution q_2 is defined to be proportional to the target posterior as follows:

$$q_2(\boldsymbol{\tau}'_{m'} | \boldsymbol{\tau}_m, i) = \frac{\pi(\boldsymbol{\tau}'_{m'} | \boldsymbol{x}) \chi_{\mathcal{P}(\boldsymbol{\tau}_m, i)}(\boldsymbol{\tau}'_{m'})}{Z(\boldsymbol{\tau}_m, i)} \tag{9}$$

with normalising constant $Z(\boldsymbol{\tau}_m, i) = \sum_{\mathbf{k}} \pi(\mathbf{k} | \boldsymbol{x}) \chi_{\mathcal{P}(\boldsymbol{\tau}_m, i)}(\mathbf{k})$. The following section considers the acceptance probability of our sampler.

3.2. Acceptance probability

The acceptance probability for transitioning to the proposal, $\boldsymbol{\tau}'_{m'}$, from the current sample, $\boldsymbol{\tau}_m$, based on the proposal distribution given in Section 3.1 is $\alpha(\boldsymbol{\tau}'_{m'}, \boldsymbol{\tau}_m) = \min[1, A(\boldsymbol{\tau}'_{m'}, \boldsymbol{\tau}_m)]$ where:

$$A(\boldsymbol{\tau}'_{m'}, \boldsymbol{\tau}_m) = \left[\frac{1}{m'} \sum_{i'=0}^{m'-1} \frac{\chi_{\mathcal{P}(\boldsymbol{\tau}'_{m'}, i')}(\boldsymbol{\tau}_m)}{Z(\boldsymbol{\tau}'_{m'}, i')} \right] \bigg/ \left[\frac{1}{m} \sum_{i=0}^{m-1} \frac{\chi_{\mathcal{P}(\boldsymbol{\tau}_m, i)}(\boldsymbol{\tau}'_{m'})}{Z(\boldsymbol{\tau}_m, i)} \right] . \tag{10}$$

Since the segment parameters have been marginalised and the changepoint event space is discrete, the determinant of the Jacobian matrix associated with the transdimensional proposals of the RJMCMC algorithm is one.

In the construction of the proposal set we note two properties for $m > 1$ and $m' > 1$: the set intersection across all segments contains only the current sample, $\cap_{i=0}^{m-1} \mathcal{P}(\boldsymbol{\tau}_m, i) = \{\boldsymbol{\tau}_m\}$; and for all other events there exists a unique segment, i', such that $\boldsymbol{\tau}_m \in \mathcal{P}(\boldsymbol{\tau}'_{m'}, i')$. The first property ensures that the current sample is always a possibility irrespective of which segment was sampled under q_1, and so it is clear that the acceptance ratio for the no-change proposal is 1. The second property arises due to the fact that the perturbation applied to the changepoint events in the birth and death proposals occur in opposite directions. This ensures that for each birth (death) proposal there exists a unique returning path to the original sample via a corresponding death (birth) case. As a consequence, the acceptance ratio in (10) simplifies because the indicator functions are all zero except for when segments i and i' correspond to the unique path between the events.

To complete, we consider the special case of $m = 1$ or $m' = 1$, i.e. transitioning into or out of the single segment model. Recall that both changepoints have to be discarded when transitioning to the single segment model, or two changepoints need to be sampled at any valid position when performing the reverse step. It is clear that both of the proposal set properties stated above are violated and the acceptance ratio must be evaluated in full.

Overall, the acceptance ratio (10) is as follows:

$$A(\boldsymbol{\tau}'_{m'}, \boldsymbol{\tau}_m) = \begin{cases} 1 & \text{if } \boldsymbol{\tau}'_{m'} = \boldsymbol{\tau}_m; \\ \frac{Z(\emptyset,0)[Z(\boldsymbol{\tau}_2,0)+Z(\boldsymbol{\tau}_2,1)]}{2Z(\boldsymbol{\tau}_2,0)Z(\boldsymbol{\tau}_2,1)} & \text{if } m = 1 \ \& \ m' = 2; \\ \frac{2Z(\boldsymbol{\tau}_2,0)Z(\boldsymbol{\tau}_2,1)}{Z(\emptyset,0)[Z(\boldsymbol{\tau}_2,0)+Z(\boldsymbol{\tau}_2,1)]} & \text{if } m = 2 \ \& \ m' = 1; \\ \frac{mZ(\boldsymbol{\tau}_m,i)}{m'Z(\boldsymbol{\tau}'_{m'},i')} & \text{otherwise,} \end{cases} \qquad (11)$$

for some unique i' such that $\boldsymbol{\tau}_m \in \mathcal{P}(\boldsymbol{\tau}'_{m'}, i')$.

The following section examines the performance of this sampler in simulations before considering an application to daily movements.

4. Simulation Study

This section examines the performance of the proposed sampler in approximating the posterior distribution for the within-period changepoint events in binary data. Section 4.1 assesses the Monte Carlo accuracy in estimating the posterior changepoint vector distribution for a data set generated from a simple model with a relatively small sample space where the true posterior can be analytically evaluated in full. The extent to which a changepoint event is detectable in terms of segment length and magnitude change in segment probabilities is investigated in Section 4.2. The third simulation study in Section 4.3 investigates the precision of the posterior when we vary the sample size across period length and number of observed periods.

4.1. Monte Carlo error

We evaluated the sampler's accuracy in approximating the posterior distribution in a simple scenario with $N = 24$ and $l = 4$ where the sample space consists of 2263 unique changepoint vectors. This example is small enough in size for the direct evaluation of the posterior in full. The data was generated from a three segment model with evenly spaced events at time points $\boldsymbol{\tau} = (8, 16, 24)$ with segment probabilities $\boldsymbol{\phi} = (0.25, 0.5, 0.6)$. Measurements from 30 complete cycles were generated giving a total of 720 observations.

The Monte Carlo uncertainty in approximating the posterior was evaluated from 1000 replications of the sampler based on a uniform spread hyperparameter $\gamma = 1$. Two independent chains were performed that were initiated at opposite ends of the sample space; namely the single segment model, $\boldsymbol{\tau}_1^{(1)} = \emptyset$, and a case where the changepoints are maximally dense e.g., $\boldsymbol{\tau}_6^{(2)} = (4, 8, 12, 16, 20, 24)$. The chains were iterated in batches of 10^4 until there was evidence of similarity in their frequency distributions. A convergence test was performed using the Weiß procedure (Deonovic and Smith, 2017) for discrete variable chains. This applies a conservative asymptotic Pearson chi-squared test which is adjusted for autocorrelation.

Bias and root mean square error (RMSE) estimates in Table 1 demonstrate that the sampler performs well in approximating the posterior probability for each

Table 1. Monte Carlo accuracy estimates in approximating the posterior probabilities from 1000 replications. Events for the top 80% are presented.

Event Positions	True Probability	Bias ($\times 10^{-3}$)	RMSE ($\times 10^{-3}$)
8, 16, 24	0.561	0.016	8.261
16, 24	0.087	-0.189	4.528
7, 16, 24	0.078	0.094	4.144
1, 8, 16	0.047	0.009	2.978
8, 16, 20, 24	0.031	-0.094	2.745

changepoint vector. Furthermore, the overall performance of the sampler in approximating the whole posterior is also good with low expected bias, -0.329×10^{-6}, and expected RMSE, 5.878×10^{-3} with respect to the true posterior distribution.

4.2. Estimation accuracy

The performance of the sampler in terms of estimation accuracy is assessed for a variety of two segment models, emulating wake/sleep. We set the period length $N = 96$, emulating 15 minute observations per day, and minimum segment length $l = 4$, sustained increased/decreased activity for an hour. Without loss of generality, the position of the first changepoint takes each of the time points $\tau_0 \in \{4, 8, 16, 24, 32, 48\}$ and the other is set at $\tau_1 = 96$. The first case, $\tau_0 = 4$, represents the shortest permitted under the minimum length constraint, whereas the last case, $\tau_0 = 48$, is positioned such that the lengths of the two segments are equal. The segment parameters were specified as $\phi = ((1 + d)/2, (1 - d)/2)$ which are centred about 0.5 but are separated by some distance $d \in \{0.0, 0.1, 0.2, 0.3, 0.4\}$. The first distance, $d = 0.0$, represents the single segment case irrespective of the position of τ_0.

We generated 1000 data sets consisting of 35 complete cycles, 2880 observations. The spread of changepoint events around the period was again specified as uniform with hyperparameter $\gamma = 1$. Note that for the later values of τ_0, $\gamma > 1$ is more appropriate and for the earlier values, $\gamma < 1$. Two chains were evaluated and tested for convergence using the same procedure described in Section 4.1. Table 2 summarises the posterior draws across the data sets for each model by three statistics: the average posterior probability for the true changepoint vector; the proportion of data sets where the posterior mode correctly identify the true changepoint positions; and the average size of the 95% highest posterior credible set (HPCS). This HPCS describes the smallest number of unique changepoint vectors that contain at least 95% of the posterior samples. The sampler performs extremely well in estimating the single segment model, $d = 0$; the truth is correctly identified as the mode in almost all cases and with high estimated posterior probability.

Turning to the changepoint results, as expected, when the change in segment probabilities is small there is little evidence for the change. The associated large coverage of the HPCS when $d = 0.1$ occurs due to high uncertainty in the location for both changepoint events, or in fact whether there are any changepoints. As expected, the ability to detect the truth improves with large changes in segment probabilities, d. Similarly, identifying the true changepoint events improves

Table 2. Estimation accuracy as measured by: $\hat{\pi}(\boldsymbol{\tau}^*)$, average posterior probability estimate of the true changepoint vector $\boldsymbol{\tau}^*$; $\%\{\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^*\}$, proportion of cases where the posterior mode, $\hat{\boldsymbol{\tau}}$, identifies the true changepoint vector and; average coverage as defined by the size of the 95% highest posterior credible set from 1000 independently generated two segment data sets for various second changepoint positions τ_0 and difference d is segment probability parameters.

τ_0		4	8	16	24	32	48
$d = 0.0$	$\hat{\pi}(\boldsymbol{\tau}^*)$	0.827	0.815	0.815	0.822	0.806	0.812
	$\%\{\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^*\}$	1.000	0.999	1.000	1.000	1.000	1.000
	Coverage	41	42	42	35	43	39
$d = 0.1$	$\hat{\pi}(\boldsymbol{\tau}^*)$	0.002	0.008	0.033	0.058	0.056	0.058
	$\%\{\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^*\}$	0.000	0.010	0.065	0.145	0.130	0.149
	Coverage	40	63	183	280	327	344
$d = 0.2$	$\hat{\pi}(\boldsymbol{\tau}^*)$	0.262	0.344	0.358	0.350	0.341	0.355
	$\%\{\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^*\}$	0.331	0.522	0.548	0.533	0.516	0.558
	Coverage	44	73	77	81	85	87
$d = 0.3$	$\hat{\pi}(\boldsymbol{\tau}^*)$	0.751	0.700	0.685	0.704	0.713	0.712
	$\%\{\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^*\}$	0.896	0.836	0.833	0.861	0.879	0.872
	Coverage	20	22	27	25	28	29
$d = 0.4$	$\hat{\pi}(\boldsymbol{\tau}^*)$	0.878	0.866	0.859	0.851	0.850	0.852
	$\%\{\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^*\}$	0.974	0.969	0.964	0.967	0.967	0.968
	Coverage	11	11	14	16	17	18

as the distance between the events increases too. The uncertainty in the posterior is broadly constant across well separated changepoints for large d, but there is a notable reduction in HPCS coverage when the changepoint events are close due to a small curtailing effect resulting from the minimum segment length condition. The broad constant feature, however, is most likely attributable to the assessment of two segment models where all observations fall on the current or following segment and therefore each providing information about both changepoint events irrespective of position.

To assess the sensitivity of the posterior estimates to the spread parameter we repeated the above study for $\gamma \in \{0.5, 1, 2\}$. These represent the prior preferences that the changepoint events are clustered, uniformly at random and more evenly spread around the period respectively. No sensitivity was identified with respect to the change in segment probabilities, but Figure 4 presents a subtle dependence of the posterior estimates when regarding changepoint positions. As may be expected, early positions, $\tau_0 = 4$, favour the prior spread of $\gamma = 0.5$ which results in subtly higher rates of identifying the truth and a smaller HPCS coverage. Conversely, the HPCS coverage when the changepoint events are as far apart as possible, $\tau_0 = 48$, is more likely on average to be smaller for $\gamma = 2$. This observed relationship is as expected due to the effect of γ on the prior as discussed in Section 2.2.
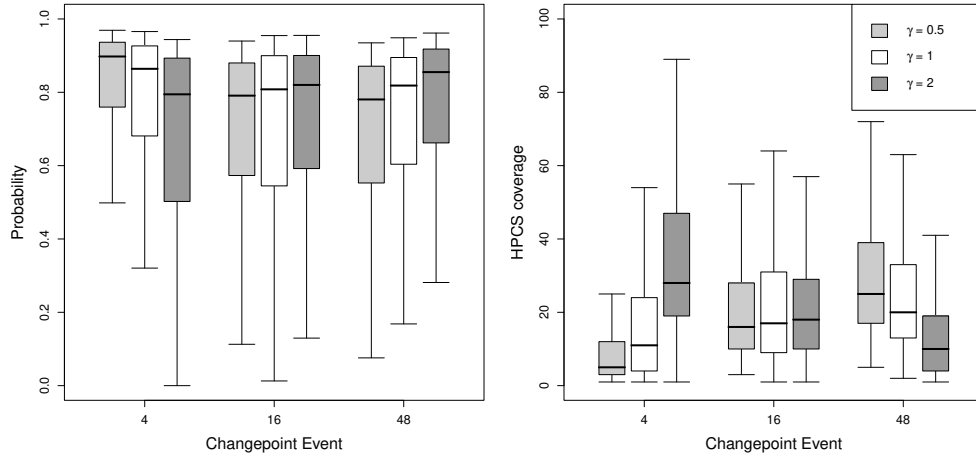
Fig. 4. Sensitivity of the prior hyperparameter, γ, on the average posterior probability for the true changepoint vector (left) and HPCS coverage (right) for various positions of the first changepoint, τ_0. The difference in segment parameter is $d = 0.3$ for all cases.

4.3. Period and sample sizes

Thus far we have considered a single $N = 96$ and 30 cycles of data. In this section we vary this to assess the effect on estimation accuracy. Motivated by our application, the increase in period length shall relate to a lower sampling frequency such as once, $N = 24$, or twice, $N = 48$, per hour. For each comparison we then vary the number of cycles to represent the same overall time frame. For consistency, the minimum segment constraint is appropriately scaled to $l = N/24$ i.e., sustained activity over an hour.

The data is sampled from a two segment model with evenly spaced changepoint events $\tau = (N/2, N)$ and segment probabilities $\phi = (0.3, 0.7)$. As with Section 4.2, 1000 data sets were generated independently for each scenario. The RJMCMC sampler was performed on each data set with a uniform prior spread $\gamma = 1$ and were iterated until convergence was satisfied.

Uncertainty in the posterior estimate over the different period lengths was assessed by evaluating the marginal distribution of the position of any changepoint event within the period. Table 3 presents the average 95% HPCS coverage of this distribution as a proportion of period length. It is clear that the uncertainty in the marginal posterior decreases with respect to both increases in period length and the observed number of cycles. Interestingly, across cases that have the same overall number of observations there is a slight reduction in the coverage proportions when the measurement frequency is high, i.e. for larger N.

Table 3. Coverage proportions of the marginal changepoint position distribution for various period lengths, N, and number of cycles. Values in parentheses denote the size of the data set for ease of comparison.

		Number of cycles			
N	7	14	21	28	35
24	61.5% (168)	46.3% (336)	37.2% (504)	30.5% (672)	25.9% (840)
48	46.0% (336)	28.8% (672)	19.1% (1008)	13.6% (1344)	10.6% (1680)
96	28.0% (672)	13.1% (1344)	7.4% (2016)	5.2% (2688)	4.3% (3360)

5. Case Study: Home Activity

The passive home sensor system developed by Intelesant Ltd. consists of a suite of sensors that monitor the activity within the home for the purpose of identifying abnormal patterns of behaviour. These abnormal patterns could indicate a decline in health or well-being in an older person. This section applies the periodic changepoint analysis developed in this paper to gain insight into baseline behaviour of individuals in determining their normal daily patterns of activity.

The data for this study was collected from four individuals over 56 days (summer 2018). Each data point indicates whether or not any of the sensors measured some sort of activity within a sequence of 15 minute intervals. Figure 1 summarises the data by averaging the activity over the 56 days. What is immediate is that the activity during the early hours is very low, indicative of sleeping, whereas the activity during daylight hours is more varied over time and individuals. When considering human activity, the requirement of consistent daily activity over the 8 week period can be seen as a strong assumption. Permutation tests presented in Appendix A based on the weekly estimated sample proportions, illustrated in Figure 1, all but person 3 present consistent long term behaviour. Closer inspection of person 3 identifies that there was only one week that exhibited a different pattern.

The period length for the analysis is specified at $N = 96$; 15 minute intervals and we are interested in daily patterns. It is assumed that any change in activity level must be sustained for at least one hour, hence we set $l = 4$. The Dirichlet hyperparameter is set at $\gamma = 1$ for all individuals to demonstrate no prior clustering preference of the changepoint events. Other choices for prior preference on clustering were investigated, but the posterior inference was robust to this decision. As in the simulations, two chains were initiated and tested for convergence. Estimates of the segment probability parameters were subsequently evaluated from the conditional beta posterior in (6).

The periodic changepoint method identifies two events for each individual representing a reduction in activity in the late evening and corresponding increase in activity in the morning that is indicative of the individual's sleep pattern.

The daytime behaviour for individual 1 is broadly low at a probability of 21%, but there is some evidence of a slight rise in activity to 30% around midday potentially attributable to preparing and eating lunch. Interestingly, a pair of changepoint events for individual 1 identify a short interval around 03:00 where the individual may occasionally get up in the night, representing a slight raise in the probability
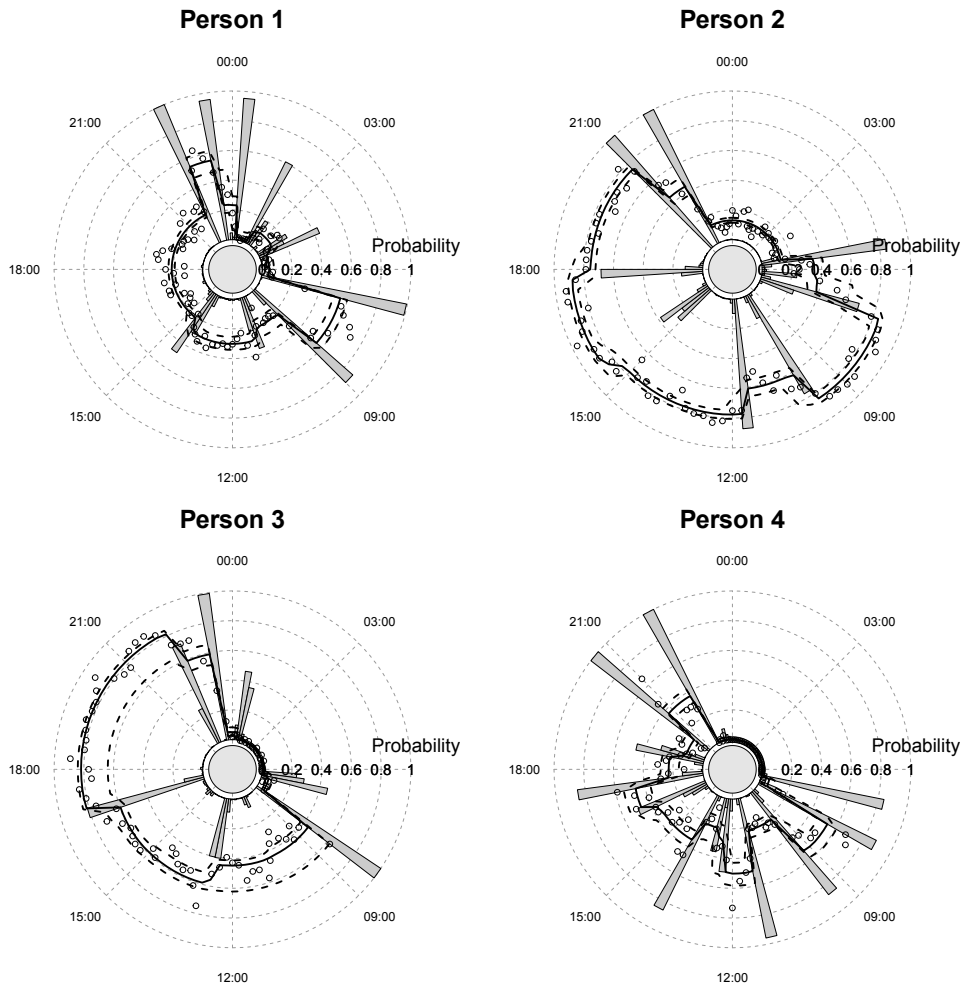
Fig. 5. Posterior activity estimates for four individuals. Points: averaged activity data over 56 days at 15 minute intervals. Radial barchart: estimated probability that a changepoint event occurs at that time point. Lines: pointwise median (solid) and 95% credible interval (dashed) of the segment probability parameter.

of sensor-detectable activity from 2% to 11%.

In contrast individual 2 is very active during daylight hours with the probability function varying between 61–88% at four distinct times. Similarly, individual 3 has a high daytime activity that increases from 44% to 82% at two events around 12:30 and 16:45. Individual 4 is the most varied in their daily activity in that the most probable number of changepoint events was identified to be 11 with a posterior probability of 49%. This is likely to be an overestimation for the number of within period changepoint events as some may be related slight variation in the probability function between weekdays (see Appendix A).

6. Discussion

This paper presents a novel approach for identifying within-period changepoint events for binary data. Unlike traditional changepoint methods, the proposed model permits the pooling of evidence across multiple periods by applying a circular rather than linear perspective of time. We developed a reversible jump Markov chain Monte Carlo sampler to traverse the complex sample space, where care is taken in defining the proposal distribution to account for the restriction imposed by the minimum segment length. Simulations demonstrate that the sampler performs well in approximating the posterior distribution, at least for problems with moderate sized support.

The inference capabilities of the proposed method were investigated via simulations with respect to the change in segment probabilities, segment duration, period length and overall number of observations. Posterior estimates correctly identified the truth across the vast majority of cases. However, potential inaccuracies to inference occurred principally for the anticipated scenarios where the change in the probability function across segments is small, and when segment durations are short or at the minimum segment length requirement. We also noted that the length of the data set has a significant impact on estimation accuracy as more days are pooled together to increase estimation accuracy. Similarly we noted a small improvement on inference accuracy occurs for longer period lengths in contrast to observing a larger number of periods.

The approach presented was motivated by an investigation of home activity data. We presented interesting insights into the regular periodic patterns of four individuals. The regular night-time period was easily identified for each individual as well as their broad daily patterns of activity that varied substantially across individuals. The resulting estimates provide a useful baseline as part of a longitudinal investigation where assessments of future periods provide insight into an individual's maintained or deteriorating daily behaviour.

In conclusion, the presented novel methodology for detecting periodic changepoint events provides a framework from which further extensions are possible. One such extension is to investigate non-binary data, which is achievable if the segment parameters can be marginalised from the joint posterior. Thus the presented methodology for approximating the posterior for the number of changepoints and their positions could be applied under the appropriately defined marginalised pos-

terior. Another possible extension would be to consider covariates such as the day of the week to allow individuals to have different probability patterns on different days.

Acknowledgements

A. Assessment of consistent daily pattern assumption

The requirement that the daily pattern of the probability function over a 2 month period is a strong assumption when regarding human activity. The degree of long term consistency for the individuals in this case study is illustrated in Figure 1 which presents the weekly estimated sample proportions calculated by $\hat{p}_w(t) = \frac{1}{7} \sum_{d=0}^{6} x_{N(d+7w)+t}$ for $w = 0, \ldots, 7$. The weekly patterns visually seems consistent with the sample proportions evaluated from all 56 days.

A more formal investigation was performed by considering the null hypothesis that the function $p_w(t)$ are identical for every week, against the alternative that there exists at least one week that is different. To assess the hypothesis the test statistic of the total elementwise absolute differences between the sample proportions between all unique week pairings as calculated: $W = \sum_{t=1}^{N} \sum_{w<w'} |\hat{p}_w(t) - \hat{p}_{w'}(t)|$. The value of this test statistic is small under the null hypothesis, otherwise W would be large. A permutation test was performed to evaluate the distribution of W under the null by resampling the binary observations across the days whilst maintaining the time-of-day to preserve any within period structure. From 10^4 samples, the one-tailed p-value for person 1–4 are 0.5651, 0.0793, 0.0282 and 0.7371 respectively. Only person 3 presents evidence of a varying week-by-week pattern, however re-evaluations of the permutation test where each week was excluded from the analysis in turn showed that only the 6th week was different (p-value: 0.2119 from excluding week 6, excluding any other week resulted in a value less than 5%). We found that the results from applying the proposed within period changepoint detection algorithm for person 3 excluding week 6 are very similar to those based on the full 56 days presented in Section 5, suggest that the difference does not affect the overall estimation.

A similar investigation was performed using weekday estimated sample proportions: $\hat{p}_d(t) = \frac{1}{8} \sum_{w=0}^{7} x_{N(d+7w)+t}$ for $d = 0, \ldots, 6$ with test statistic $D = \sum_{t=1}^{N} \sum_{d<d'} |\hat{p}_d(t) - \hat{p}_{d'}(t)|$. The one-tailed p-values from the permutation test for person 1–4 are 0.0428, 0.0587, 0.0094 and 0.0001 respectively. Although the null cannot be rejected for person 2 at the 5% level, it is clear that there is evidence for a weekday effect for the other individuals. The consequence of this on the analysis is discussed in Sections 5 and 6.

References

Chen, J. and A. K. Gupta (2000). Parametric statistical change point analysis: with applications to genetics, medicine, and finance. New York: Birkhäuser. 2

Deonovic, B. E. and B. J. Smith (2017). Convergence diagnostics for MCMC draws of a categorical variable. 12

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple change-point problems. Statistics and Computing 16(2), 203–213. 3

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82(4), 711–732. 3, 8, 9

Jammalamadaka, S. R. and A. Sengupta (2001). Topics in circular statistics, Volume 5. world scientific. 2

Killick, R., P. Fearnhead, and I. A. Eckley (2012). Optimal detection of change-points with a linear computational cost. Journal of the Americal Statistical Association 1107(500), 1590–1598. 2

ONS (2019, November). Families and households in the UK: 2019. Accessed: 27 November 2019. 1

Pierson, E., T. Althoff, and J. Leskovec (2018). Modeling individual cyclic variation in human behavior. In Proceedings of the 2018 World Wide Web Conference, pp. 107–116. International World Wide Web Conferences Steering Committee. 3

Price-Williams, M., N. Heard, and M. Turcotte (2017). Detecting periodic subsequences in cyber security data. In Intelligence and Security Informatics Conference (EISIC), 2017 European, pp. 84–90. IEEE. 2

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286. 3

Randic, M., D. Morales, and O. Araujo (2008). Higher-order Lucas numbers. Divulgaciones Matemáticas 16(2), 275–283. 7

Scott, A. J. and M. Knott (1974). A cluster analysis method for grouping means in the analysis of variance. Biometrics (30), 507–512. 2

Truong, C., L. Oudre, and N. Vayatis (2018). A review of change point detection methods. arXiv (arXiv:1801.00718). 2

Venkatraman, E. S. and A. B. Olshen (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23(6), 657–663. 2

Zhang, N. R., D. O. Siegmund, H. Ji, and J. Z. Li (2010). Detecting simultaneous changepoints in multiple sequences. Biometrika 97(3), 631–645. 2