Submitted to the Annals of Applied Statistics

DOUBLY-ONLINE CHANGEPOINT DETECTION FOR MONITORING HEALTH STATUS DURING SPORTS ACTIVITIES

BY MATTIA STIVAL[1], MAURO BERNARDI[1] AND PETROS DELLAPORTAS [2,3]

[1] *Department of Statistical Sciences, University of Padova, mattia.stival@unipd.it*

[2] *Department of Statistical Science, University College London*

[3] *Department of Statistics, Athens University of Economics and Business*

We provide an online framework for analyzing data recorded by smart watches during running activities. In particular, we focus on identifying variations in the behavior of one or more measurements caused by changes in physical condition, such as physical discomfort, periods of prolonged detraining, or even the malfunction of measuring devices. Our framework considers data as a sequence of running activities represented by multivariate time series of physical and biometric data. We combine classical changepoint detection models with an unknown number of components with Gaussian state space models to detect distributional changes between a sequence of activities. The model considers multiple sources of dependence due to the sequential nature of subsequent activities, the autocorrelation structure within each activity, and the contemporaneous dependence between different variables. We provide an online Expectation-Maximization (EM) algorithm involving a sequential Monte Carlo (SMC) approximation of changepoint predicted probabilities. As a byproduct of our model assumptions, our proposed approach processes sequences of multivariate time series in a doubly-online framework. While classical changepoint models detect changes between subsequent activities, the state space framework coupled with the online EM algorithm provides the additional benefit of estimating the real-time probability that a current activity is a changepoint.

1. Introduction. Running is one of the most popular and practiced sports worldwide, with almost 60 million people having participated in running, jogging, and trail running in 2017 in the United States (Statista, 2020a). Increasingly more runners use smart watches and devices that record their workouts, allowing for performance analysis and the planning of future workouts. Worldwide smart watch shipments volume as estimated by Statista (2020b) were 74 million units in 2018, 97 million units in 2019, 115 million units in 2020, with an expected growth to over 258 million units by 2025. Apps and wearables are driving the next digital health and fitness revolution, in which intelligent and automatic real-time control and monitoring tools will become extremely relevant (Statista, 2020c). Indeed, it is expected that in the near future, and in some cases even now, smart watches may be used as medical monitoring devices, providing support at an individual level to health-care consumers (Free et al., 2013; Singh et al., 2018) and, more importantly, to users with different levels of health literacy, communication, and data skills (Siqueira do Prado et al., 2019; Vitabile et al., 2019). The spectrum of available and potential measurements by smart watches includes information on movement, heart rate, blood oxygenation and pressure, and glucose (García-Guzmán et al., 2021; PKvitality, 2020). Our contribution provides a modeling framework to analyze, in an online fashion, data recorded from smart devices during running activities. In particular, we focus on identifying variations in the behavior of one or more measurements caused by

Keywords and phrases: real-time health monitoring, smart watches, Online Expectation Maximization, Sequential Monte Carlo.

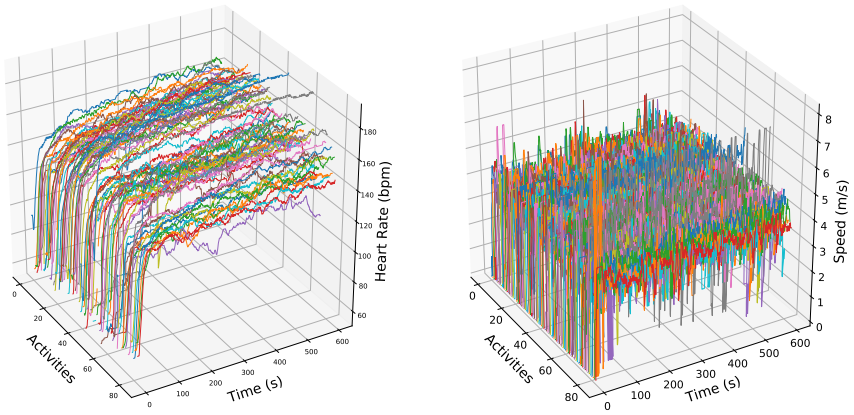

FIG 1. *A sequence of activities performed by one athlete from our dataset*

changes in physical condition such as physical discomfort, periods of prolonged de-training, or even the malfunction of measuring devices (Schneider et al., 2018).

The use of wearable technologies and sensor data for medical problems is gaining increasing interest from the statistical community, see for example Huang et al. (2019); de Chaumaray, Marbac and Navarro (2020); Qian et al. (2020). The difficulty in monitoring performances due to the presence of disturbing factors, such as environmental conditions or other within-activity sources of variability, is widely accepted; see, for example Schneider et al. (2018). A valuable contribution to this field was provided by Frick and Kosmidis (2017), who developed an R (R Core Team, 2020) package that allows for both basic and advanced retrospective analysis of data collected from smart devices. Unlike previous works on this type of data, we focus on online inference because it highlights the important aspect of smart devices related to the monitoring activities as they are carried out (Bourdon et al., 2017).

Recent literature in sports science and medicine points out the need to make decisions by evaluating the personal medical history, the long- and short-term training goals of the athlete, and the time course of training schedules (Pelliccia et al., 2021; Schneider et al., 2018). We address these issues by utilizing data collected as a sequence of *activities*, where each activity represents a part of the training session. The relevant measurements that we will consider in this study are heart rate (bpm, beats per minute) and speed (m/s, meters per second), whereas other common variables that can be incorporated in our proposed methodology are cadence (spm, steps per minute) and the runner's geographical position (latitude, longitude, and altitude). Figure 1 shows a sample of the data, consisting of 85 consecutive warm-up activities performed by one athlete during which the heart rate and speed are monitored over time. For all the activities, after a sudden increase, the heart rate curves seem to slowly evolve around a trend, while the speed levels change slowly during the activity.

For one activity, all collected information is represented by a multivariate time series, with complex dependence structures that make the extraction of the underlying signal a non-trivial statistical problem. Our inferential framework is *doubly-online* in the following sense. First, we identify changepoints in a *between-online* setting, in which activities are processed sequentially when a new one is fully observed. This permits to divide activities into subsequent segments and update the information on the unknown parameters at the end of each activity. We also consider a *within-online* setting, which refers to the online data processing of one activity. During a run, having information on the behavior difference between the current and the previous activities may be translated into motivational feedback or a potential alert before

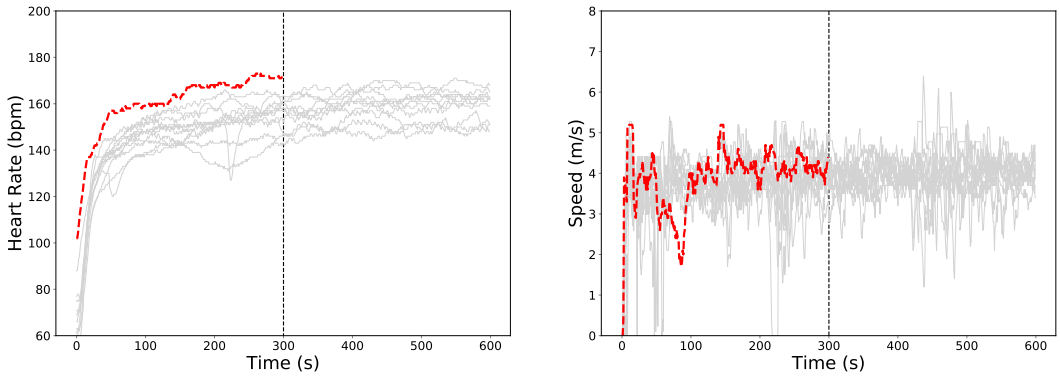

FIG 2. *An example of the within-online setting. The red dashed lines indicate the current monitored activity, while gray lines denote previous activities. The vertical line marks the time at which our algorithm provides the posterior probability that the current activity is a changepoint.*

the end of the activity. Figure 2 shows the within-online setting for data collected by one runner in our dataset. The red lines are associated to one new activity, monitored by the athlete after five minutes of running and characterized by high effort, although the speed behavior seems to be similar to those in the previous activities (shown in gray). Our algorithm provides an online probabilistic quantification of the changepoint uncertainty by delivering the posterior probability of a behavioral change occurrence at any time point of the activity. In the case of Figure 2, the runner is interested in the behavior change at minute 5 of the current activity. The instant when the runner monitors the status was chosen for exemplary reasons, with the aim of showing the interest of the athlete in having information before the end of the activity. However, this interest may vary and be located in different parts of the activity.

We model the set of observed activities as a multivariate state space model (Durbin and Koopman, 2012; Shumway and Stoffer, 2017) and we adapt to this framework classical changepoint modeling, which allows for the online detection of an a priori unknown set of changepoints between activities, see Chib (1998); Fearnhead and Liu (2007); Caron, Doucet and Gottardo (2012); Yildirim, Singh and Doucet (2013). Changepoint detection is a relevant problem in many fields of science, ranging from industrial process control, health monitoring, cybersecurity, and machine learning (see, e.g., Aminikhanghahi and Cook, 2017; Titsias, Sygnowski and Chen, 2020; Xie et al., 2021; Haynes, Fearnhead and Eckley, 2017). Early works on changepoints identification coupled with state space model can be also found in Kitagawa (1987) and Song (2007). Our approach differs in that we solve a problem of changepoint signal extraction in which the double sequential nature—between and within activities—of the data-generating process is preserved. The key idea is that we leverage the data on the past history of the athlete as a benchmark for identifying standard behaviors and deviations, providing relevant information about the performance as new data are collected. In our application, making online inferences on a sequence of activities before the last one is fully observed is clearly of paramount importance. The literature on the changepoint detection problem is very large, and alternative approaches have been proposed for high dimensional frameworks, mostly based on dimensionality reduction techniques (see, e.g., Samé and Govaert, 2017; Grundy, Killick and Mihaylov, 2020). Such approaches, although potentially usable in the between-online setting, in which the observations for identifying changepoints consist of entire activities represented by multiple multivariate time series, are not directly applicable in the within-online setting, in which there is the need to preserve the dual sequential nature of the data. The literature on online learning is vast and fast increasing, with new contributions and techniques that deal with both high dimensional and challenging scientific

problems; see, for example, Cappé and Moulines (2009), Denevi et al. (2019) and Luo and Song (2020). We contribute to this literature by proposing a new state-space-based algorithm for changepoint detection in a sequence of time series by adopting the online Expectation-Maximization (EM) algorithm developed by Yildirim, Singh and Doucet (2013). The nature of our problem requires taking into account three sources of dependence: one that inherits the sequential nature of subsequent activities, one that considers the autocorrelation structure within each activity, and one that models the contemporaneous dependence between variables. As a byproduct of our model assumptions and the online inferential procedure, our approach processes sequences of data in a doubly-online framework. While classical changepoint models detect distributional changes in a sequence of activities (i.e., multivariate time series), our state space model coupled with the online EM approach provides the additional benefit of estimating the probability that a single activity is a changepoint during a run.

The paper is organized as follows. Sections 2 and 3 describe the statistical model and the algorithm to perform inference respectively. Section 4 illustrates a small simulation study that assesses our algorithm performance with synthetic examples and Section 5 presents the results of our model using real data collected by an athlete. We conclude with Section 7, which describes the model's limitations and possible future developments. Proofs and details on the available dataset are provided in the supplementary material.

2. The model. For each runner, we observe the data $\mathbf{y}_{1:N,1:T}$, composed of N ordered activities that are represented by P-dimensional time series at T time points. An activity can be thought of as a running session taking place on different days; T defines the duration of each activity, which is considered , for simplicity, to be equal for all activities, and P denotes the number of smart device measurements, such as heart rate and speed. Our interest lies in modeling the data online and identifying changepoints during each activity, using information on both previous activities and previous recordings during the current activity. We build our model by first introducing an N-dimensional latent vector $S_{1:N} = (S_1, \dots, S_N)$ such that $S_1 = 1$ and $S_n - S_{n-1} = 1$ if a changepoint occurs at the n-th ($n > 2$) activity. The vector $S_{1:N} = (S_1, \dots, S_N)$ divides the activities into S_N contiguous *segments*, in which activities belonging to different segments are assumed to be independent of each other. The segments $S_{1:N}$ are modeled using a discrete state space Markov chain with transition probability $p(S_n|S_{n-1}) = \lambda$ if $S_n = S_{n-1} + 1$, for $0 < \lambda < 1$.

Assume that the activity n belongs to segment s. We model its measurements at time t by a state space representation with measurement equation

$$\mathbf{y}_{n,t} = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\theta}}^{(S)} & \mathbf{Z}_{\boldsymbol{\theta}}^{(A)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{n,t} \end{bmatrix} + \boldsymbol{\epsilon}_{n,t}, \tag{1}$$

with $\boldsymbol{\epsilon}_{n,t} \overset{iid}{\sim} N_P(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, and state equation

$$\begin{bmatrix} \boldsymbol{\alpha}_{t+1}^{(s)} \\ \boldsymbol{\alpha}_{n,t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\boldsymbol{\theta}}^{(S)} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{\boldsymbol{\theta}}^{(A)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{n,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_t^{(s)} \\ \boldsymbol{\eta}_{n,t} \end{bmatrix}, \tag{2}$$

with $\boldsymbol{\eta}_t^{(s)} \overset{iid}{\sim} N_M(\mathbf{0}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, $\boldsymbol{\eta}_{n,t} \overset{iid}{\sim} N_K(\mathbf{0}, \boldsymbol{\Delta}_{\boldsymbol{\theta}})$, and $\boldsymbol{\alpha}_1^{(s)} \overset{iid}{\sim} N_M(\hat{\boldsymbol{\alpha}}_{1|0}^{(S)}, \mathbf{P}_{1|0}^{(S)})$ independent of $\boldsymbol{\alpha}_{n,1} \overset{iid}{\sim} N_K(\hat{\boldsymbol{\alpha}}_{1|0}^{(A)}, \mathbf{P}_{1|0}^{(A)})$. The subscript $\boldsymbol{\theta}$ is used throughout to highlight which parts of the model depend on, or are a function of, an unknown parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, which is the object of inference in the model. In this part of model specification, both the parameter $\boldsymbol{\theta}$ and the parameter space $\boldsymbol{\Theta}$ are deliberately left generic so that a general model that can be used with different type of measurements is presented. One could think to $\boldsymbol{\theta}$ as the typical

parameter involved in state space models, composed, for example, of autoregressive coefficients and covariance matrices of disturbances and errors. In the above specification, $\boldsymbol{\alpha}_t^{(s)}$ are vectors of dimensions $M \times 1$ that denote the dynamic segment-specific latent features, which are supposed to be independent of any other $\boldsymbol{\alpha}_t^{(s')}$, for any $s \neq s'$. Together with $S_{1:N}$, the segment-specific latent features $\boldsymbol{\alpha}_t^{(s)}$ account for the dependence between subsequent activities. The activity-specific latent features $\boldsymbol{\alpha}_{n,t}$ are vectors of dimension $K \times 1$ that capture temporal dependencies that are unrelated to the performance of the athlete and describe negligible factors or disturbing aspects associated with the activities. These vectors are assumed to be independent of $\boldsymbol{\alpha}_t^{(s)}$ and any other $\boldsymbol{\alpha}_{n',t}$, with $n' \neq n$. The elements $\hat{\boldsymbol{\alpha}}_{1|0}^{(S)}$, $\hat{\boldsymbol{\alpha}}_{1|0}^{(A)}$, and $\mathbf{P}_{1|0}^{(S)}, \mathbf{P}_{1|0}^{(A)}$ are the initial means and covariance matrices of the segment-specific and activity-specific latent features, respectively. With no information on the initial states, we adopt the diffuse state initialization technique, in which these means and covariances are independent of $\boldsymbol{\theta}$, and the latter are supposed to be large (Durbin and Koopman, 2012). The elements $\mathbf{Z}_{\boldsymbol{\theta}}^{(S)}$, $\mathbf{Z}_{\boldsymbol{\theta}}^{(A)}$, $\mathbf{T}_{\boldsymbol{\theta}}^{(S)}$, and $\mathbf{T}_{\boldsymbol{\theta}}^{(A)}$ are non-stochastic design matrices with dimensions $P \times M$, $P \times K$, $M \times M$, and $K \times K$, respectively. The superscripts (S) and (A) indicate that these matrices are linked to the segment-specific latent states $\boldsymbol{\alpha}_t^{(s)}$ and activity-specific latent states $\boldsymbol{\alpha}_{n,t}$, respectively. These matrices are shared across different segments and different activities, and may depend on $\boldsymbol{\theta}$. Their specification is left undefined and depends on the specific application and behavior of the variables being considered, as it is typical in state space modeling (see, e.g., Durbin and Koopman, 2012). Coupled with the design matrices, the covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$, and $\boldsymbol{\Delta}_{\boldsymbol{\theta}}$ of dimensions $P \times P$, $M \times M$, and $K \times K$, respectively, capture any contemporaneous dependencies between different elements of the model, such as the entries of the error component $\boldsymbol{\epsilon}_{n,t}$ of dimensions $P \times 1$ or those of the disturbance vectors $\boldsymbol{\eta}_t^{(s)}$ and $\boldsymbol{\eta}_{n,t}$ of dimensions $M \times 1$ and $K \times 1$, respectively. In general, the covariance matrices are full and unstructured; however, depending on the application, they may have a specific structure and involve a small number of elements of $\boldsymbol{\theta}$. Note that the design matrices of the model and the covariance matrices are fixed both with respect to different time instants (i.e. t) and different activities (i.e. n). Generalizing the model to consider time- or activity-dependent matrices is possible, conditional on whether these are known or non-stochastic. This aspect, although it allows to enlarge the number of models that can be considered within the same framework, is outside the scope of this work, whose central themes are the between and the within-online settings for online monitoring of sports activities. More details on these generalizations, however, can be found in Section 7 and in the supplementary material (Section S5). In the next Subsection we derive some alternative model formulations that are useful both from a computational point of view and for obtaining the model likelihood.

2.1. *Working model and likelihood.* Let $\boldsymbol{\alpha}_{1:T}^{1:N} = (\boldsymbol{\alpha}_{1:N,1:T}, \boldsymbol{\alpha}_{1:T}^{(1:S_N)})$ be a vector storing both the segment-specific and the activity-specific latent features. It is possible to write the augmented likelihood for the model, which has the conditional independence structure

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}, \boldsymbol{\alpha}_{1:T}^{1:N}, S_{1:N}) = p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T} | \boldsymbol{\alpha}_{1:T}^{1:N}, S_{1:N}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:T}^{1:N} | S_{1:N}) p(S_{1:N}), \tag{3}$$

where $p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:T}^{1:N} | S_{1:N}) = p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:T}^{(1:S_N)} | S_{1:N}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:N,1:T})$. To obtain the likelihood of the observed process, we need to integrate out both $\boldsymbol{\alpha}_{1:T}^{1:N}$ and $S_{1:N}$. To do so, we first note that, conditional on $S_{1:N}$, it is possible to consider S_N independent segments that are described by S_N independent segment-specific state space models. More specifically, consider the model specified in Equations (1) and (2) and condition on $S_{1:N}$. Assume also that the s-th

segment ranges between the j_s-th and the k_s-th activity, so that its length is $m_s = k_s - j_s + 1$. Under our model assumptions, the segment s can be modeled using the following equations:

$$\begin{bmatrix} \mathbf{y}_{j_s,t} \\ \mathbf{y}_{j_s+1,t} \\ \vdots \\ \mathbf{y}_{k_s,t} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\theta}}^{(S)} & \mathbf{Z}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{Z}_{\boldsymbol{\theta}}^{(S)} & \mathbf{0} & \mathbf{Z}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \vdots \\ \vdots & \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{Z}_{\boldsymbol{\theta}}^{(S)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}_{\boldsymbol{\theta}}^{(A)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{j_s,t} \\ \boldsymbol{\alpha}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\alpha}_{k_s,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{j_s,t} \\ \boldsymbol{\epsilon}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\epsilon}_{k_s,t} \end{bmatrix}, \tag{4}$$

$$\begin{bmatrix} \boldsymbol{\alpha}_{t+1}^{(s)} \\ \boldsymbol{\alpha}_{j_s,t+1} \\ \boldsymbol{\alpha}_{j_s+1,t+1} \\ \vdots \\ \boldsymbol{\alpha}_{k_s,t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\boldsymbol{\theta}}^{(S)} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{T}_{\boldsymbol{\theta}}^{(A)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{j_s,t} \\ \boldsymbol{\alpha}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\alpha}_{k_s,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_t^{(s)} \\ \boldsymbol{\eta}_{j_s,t} \\ \boldsymbol{\eta}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\eta}_{k_s,t} \end{bmatrix} \tag{5}$$

for $\boldsymbol{\epsilon}_{j_s:k_s,t} = (\boldsymbol{\epsilon}'_{j_s,t}, \boldsymbol{\epsilon}'_{j_s+1,t}, \dots, \boldsymbol{\epsilon}'_{k_s,t})' \overset{iid}{\sim} N_{m_s P}(\mathbf{0}, \mathbf{I}_{m_s} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, $\boldsymbol{\eta}_{j_s:k_s,t} = (\boldsymbol{\eta}'_{j_s,t}, \boldsymbol{\eta}'_{j_s+1,t} \dots, \boldsymbol{\eta}'_{k_s,t})' \overset{iid}{\sim} N_{m_s K}(\mathbf{0}, \mathbf{I}_{m_s} \otimes \boldsymbol{\Delta}_{\boldsymbol{\theta}})$, $\eta_t^{(s)} \sim N_M(\mathbf{0}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, and $\boldsymbol{\alpha}_{j_s:k_s,1} = (\boldsymbol{\alpha}'_{j_s,1}, \boldsymbol{\alpha}'_{j_s+1,1} \dots, \boldsymbol{\alpha}'_{k_s,1})' \sim N_{m_s K}(\mathbf{1}_{m_s} \otimes \hat{\boldsymbol{\alpha}}_{1|0}^{(A)}, \mathbf{I}_{m_s} \otimes \mathbf{P}_{1|0}^{(A)})$, $\boldsymbol{\alpha}_1^{(s)} \sim N_M(\hat{\boldsymbol{\alpha}}_{1|0}^{(S)}, \mathbf{P}_{1|0}^{(S)})$, independent of each other and with fixed hyper-parameters.

Conditional on segments $S_{1:N}$, Equations (4) and (5) specify a state space model such that both the segment-specific and activity-specific latent features can be integrated out by means of a Kalman filter routine. By integrating out these latent features in Equation (3) we obtain the contribution of the s-th segment to the likelihood conditional on $S_{1:N}$ given by

$$\log p_{\boldsymbol{\theta}}(\mathbf{y}_{j_s:k_s,1:T} | S_{j_s:k_s}) = -\frac{1}{2} \sum_{t=1}^{T} \left(m_s P \log(2\pi) + \log |\mathbf{F}_{s,t}| + \boldsymbol{\upsilon}'_{j_s:k_s,t} (\mathbf{F}_{s,t})^{-1} \boldsymbol{\upsilon}_{j_s:k_s,t} \right) \tag{6}$$

where both the innovations vectors $\boldsymbol{\upsilon}_{j_s:k_s,t}$ and their respective covariance matrices $\mathbf{F}_{s,t}$ are outputs of the Kalman filter routine, reviewed in the supplementary material (Section S3). Thus, the likelihood is conditional on the segments, but no longer on the segment- and activity-specific latent features. The conditional likelihood depends clearly on the unknown parameter $\boldsymbol{\theta}$ through $\boldsymbol{\upsilon}_{j_s:k_s,t}$ and $\mathbf{F}_{s,t}$, which are functions of the data, the design matrices, and the covariance matrices involved in the state space model, for which the subscript $\boldsymbol{\theta}$ has been omitted for simplicity of notation. While the model specification above is intuitively driven by the mechanism that generates the data, it is useful to connect it with the way Yildirim, Singh and Doucet (2013) specified a model because we will adopt their inferential strategy in the next section. Specifically, instead of $S_{1:N}$, we can define a latent vector $D_{1:N} = (D_1, \dots, D_N)$ such that D_n represents the delay of from the last changepoint defined through the following recursion

$$D_n | D_{n-1} = \begin{cases} D_{n-1} + 1 & \text{if} \quad S_n = S_{n-1} \\ 1 & \text{if} \quad S_n = S_{n-1} + 1 \end{cases},$$

with $D_1 = 1$, and we note the information equivalence between $D_{1:N}$ and $S_{1:N}$. We can then express the conditional likelihood of the observed process as

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T} | D_{1:N}) = \prod_{n=1}^{N} G_{\boldsymbol{\theta},n}^{D}(D_n), \tag{7}$$

where the *potentials* are defined as

$$G^D_{\boldsymbol{\theta},n}(D_n) = p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:T}|D_{1:n}, \mathbf{y}_{1:(n-1),1:T}) = \begin{cases} \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:T}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),1:T}|D_{n-1})} & \text{if } D_n = D_{n-1}+1 \\ p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:T}|D_n) & \text{if } D_n = 1 \end{cases},$$

with $j = n - D_n + 1$. Notice that the potential $G^D_{\boldsymbol{\theta},n}(D_n)$ is nothing more than the individual contribution of activity n to the conditional likelihood of the observed process, provided that the first $n-1$ activities have already been observed and the index of the last changepoint is known by means of D_n. The likelihoods involved in the potentials can be easily calculated through the use of Kalman filter routines, as in Equation (6), in which, for activity n, the activities to be considered in the respective segment are determined by D_n. Knowing either $D_{1:N}$ or $S_{1:N}$ is equivalent, while if we consider only the marginal D_n instead of $S_{j:n}$ with $j = \max(1, S_n - D_n + 1)$, we lose the information on the number of the segment the n-th activity belongs to. We do not consider the random variable D^s_n, which highlights both the delay with respect to the last changepoint and the segment to which the activity belongs to. Since our primary interest is the early changepoint detection, all the provided results rely on an underlying exchangeability assumption between segment-specific features, which simplifies the mathematical treatment.

The likelihood of the observed process is given by $p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}) = \mathrm{E}_{\boldsymbol{\theta}}\big[\prod_{n=1}^N G^D_{\boldsymbol{\theta},n}(D_n)\big]$ where the expectation is taken with respect to $D_{1:N}$. This likelihood represents the target to maximize for obtaining an estimate of the unknown parameter $\boldsymbol{\theta}$, which drives the behavior of the observed process. The parameter $\boldsymbol{\theta}$ is involved in the model specification of both the segments-specific, and the activity-specific temporal dynamics during the activities.

3. Estimation and changepoint detection.

3.1. *From batch to online EM algorithms.* Our interest lies in $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big[p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T})\big]$ via the EM algorithm introduced by Dempster, Laird and Rubin (1977). An exact online EM algorithm for linear and Gaussian state space models was introduced by Elliott, Ford and Moore (2002). Here, we review and adapt to our setting the online EM algorithm by Yildirim, Singh and Doucet (2013), involving a Sequential Monte Carlo (SMC) approximation step, developed for a large class of changepoints models.

Let $\hat{\boldsymbol{\theta}}_{it}$ be the the estimate of the maximizer at the it-th iteration of the EM algorithm. At iteration $it+1$ the expectation step of the offline EM algorithm computes

$$Q_{1:N}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{it}) = \mathrm{E}_{\hat{\boldsymbol{\theta}}_{it}}\big[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}, \boldsymbol{\alpha}^{1:N}_{1:T}, D_{1:N})|\mathbf{y}_{1:N,1:T}\big] \tag{8}$$

$$= \mathrm{E}_{\hat{\boldsymbol{\theta}}_{it}}\Big[\log p(D_{1:N}) + \mathrm{E}_{\hat{\boldsymbol{\theta}}_{it}}\big[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}, \boldsymbol{\alpha}^{1:N}_{1:T}|D_{1:N})|D_{1:N}, \mathbf{y}_{1:N,1:T}\big]|\mathbf{y}_{1:N,1:T}\Big] \tag{9}$$

The expected value in Equation (8) is computed with respect to both $D_{1:N}$ and the latent features $\boldsymbol{\alpha}^{1:N}_{1:T}$, considered jointly, and involves the log-density augmented for both latent variables. Equation (9) involves an external and an internal expectation, which are computed with respect to the random variables $D_{1:N}$ and $\boldsymbol{\alpha}^{1:N}_{1:T}|D_{1:N}$, respectively, given the entire set of data $\mathbf{y}_{1:N,1:T}$. The subscript $1{:}N$ in $Q_{1:N}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{it})$ indicates that all the observations up to activity N are used. Moreover, $Q_{1:N}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{it})$ depends on $\boldsymbol{\theta}$ through the functional form of the augmented likelihood $p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}, \boldsymbol{\alpha}^{1:N}_{1:T}, D_{1:N})$. The true parameter $\boldsymbol{\theta}$ is substituted by its estimate $\hat{\boldsymbol{\theta}}_{it}$ when the expected values are computed at iteration $it+1$. Once this expectation is computed, the maximization step solves

$$\hat{\boldsymbol{\theta}}_{it+1} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big[Q_{1:N}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{it})\big] = \boldsymbol{\Lambda}(\mathcal{Q}_{1:N}) \tag{10}$$

with $\boldsymbol{\Lambda} : \mathcal{Q}_{1:N} \to \boldsymbol{\Theta}$, and $\mathcal{Q}_{1:N}$ being the r-dimensional set of sufficient statistics. The two steps are repeated until a set of stopping rules are satisfied, which allows to iteratively grow the function $Q_{1:N}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{it})$ and, consequently, the likelihood of the observed process. The offline EM algorithm requires the ability to compute both the E-step in Equation (8) and the M-step in Equation (10) in closed form or through the use of a finite set of elementary operations, involving the expectation of the set of r sufficient statistics $\mathcal{Q}_{1:N}$.

To adapt the EM algorithm to the online setting, we define the *individual contribution* of activity n to $Q_{1:n}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ as

$$\begin{aligned}\iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T}) := &\log p(D_{1:n}) - \log p(D_{1:(n-1)}) \\ &+ \mathrm{E}_{\boldsymbol{\theta}'}\left[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:n,1:T}, \boldsymbol{\alpha}_{1:T}^{1:n} | D_{1:n}) | \mathbf{y}_{1:n,1:T}, D_{1:n}\right] \\ &- \mathrm{E}_{\boldsymbol{\theta}'}\left[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:T}, \boldsymbol{\alpha}_{1:T}^{1:(n-1)} | D_{1:(n-1)}) | \mathbf{y}_{1:(n-1),1:T}, D_{1:(n-1)}\right],\end{aligned}$$

with $\iota_{\boldsymbol{\theta}'}(\mathbf{y}_{1,1:T}) = \mathrm{I}(D_1 = 1) + \mathrm{E}_{\boldsymbol{\theta}'}\left[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1,1:T}, \boldsymbol{\alpha}_{1:T}^{1} | D_1) | \mathbf{y}_{1,1:T}, D_1\right]$, for any value $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$. The expression for $\iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})$ is nothing else but the difference between the argument of the external expected value in Equation (9) computed using the observations up to activity n and the same argument calculated using with observations up to activity $n-1$, in which the expectations are taken with respect to the latent features $\boldsymbol{\alpha}_{1:T}^{1:n}$ and $\boldsymbol{\alpha}_{1:T}^{1:(n-1)}$ involved in the respective state space models. Although not easy to interpret, the construction of $\iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})$ mimics the definition of the conditional likelihood in terms of the potentials in Equation (7) and allows to write the expression of $Q_{1:N}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ as the expected value with respect to $D_{1:N}$ of a sum of N functionals, i.e. $Q_{1:N}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathrm{E}_{\boldsymbol{\theta}'}\left[\sum_{n=1}^{N} \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T}) | \mathbf{y}_{1:N,1:T}\right]$, and therefore its sequential evaluation as new activities are observed.

We adopt the stochastic approximation proposed by Yildirim, Singh and Doucet (2013) based on a forward smoothing technique, see for example Kantas et al. (2015). By setting $\mathbf{T}_1(D_1, \boldsymbol{\theta}) = \iota_{\boldsymbol{\theta}}(\mathbf{y}_{1,1:T})$, and defining

$$\mathbf{S}_n(D_{1:n}, \boldsymbol{\theta}') := \sum_{j=1}^{n} \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{j,1:T})$$

$$\begin{aligned}\mathbf{T}_n(D_{1:n}, \boldsymbol{\theta}') :&= \sum_{D_{1:(n-1)} \in \mathcal{D}_{1:(n-1)}} \mathbf{S}_n(D_{1:n}, \boldsymbol{\theta}') p_{\boldsymbol{\theta}}(D_{1:(n-1)} | \mathbf{y}_{1:(n-1),1:T}, D_n) \\ &= \sum_{D_{n-1} \in \mathcal{D}_{n-1}} \left[\mathbf{T}_{n-1}(D_{1:(n-1)}, \boldsymbol{\theta}') + \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})\right] p_{\boldsymbol{\theta}}(D_{n-1} | \mathbf{y}_{1:(n-1),1:T}, D_n), \qquad (11)\end{aligned}$$

we are able to evaluate $\mathbf{T}_n(D_{1:n}, \boldsymbol{\theta}')$ sequentially. It can also be shown that

$$Q_{1:n}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathrm{E}_{\boldsymbol{\theta}'}\left[\sum_{j=1}^{n} \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{j,1:T}) | \mathbf{y}_{1:n,1:T}\right] = \sum_{D_n \in \mathcal{D}_n} \mathbf{T}_n(D_{1:n}, \boldsymbol{\theta}') p_{\boldsymbol{\theta}}(D_n | \mathbf{y}_{1:n,1:T})$$

allowing, subject to knowing $\mathbf{T}_n(D_{1:n}, \boldsymbol{\theta}')$ and $p_{\boldsymbol{\theta}}(D_n | \mathbf{y}_{1:n,1:T})$, to also obtain $Q_{1:n}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ sequentially for any activity n that has been fully observed.

Let γ_n be a step-size decreasing function such that $0 < \gamma_n < 1$, $\sum_{n=1}^{\infty} \gamma_n = \infty$, $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$. The stochastic approximation of Equation (11) proposed by Yildirim, Singh and Doucet (2013) becomes

$$\begin{aligned}\mathbf{T}_{\gamma,n}(D_{1:n}; \hat{\boldsymbol{\theta}}_{n-1}) = &\sum_{D_{n-1} \in \mathcal{D}_{n-1}} \left[(1-\gamma_n)\mathbf{T}_{\gamma,n-1}(D_{1:(n-1)}; \hat{\boldsymbol{\theta}}_{n-2})\right. \\ &\left. + \gamma_n \iota_{\hat{\boldsymbol{\theta}}_{n-1}}(\mathbf{y}_{n,1:T})\right] p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_{n-1} | \mathbf{y}_{1:(n-1),1:T}, D_n), \qquad (12)\end{aligned}$$

which leads to

$$\mathcal{Q}_n = \sum_{D_n \in \mathcal{D}_n} \mathbf{T}_{\gamma,n}(D_{1:n}; \hat{\boldsymbol{\theta}}_{n-1}) p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n | \mathbf{y}_{1:n,1:T}),$$

which is used for obtaining $\hat{\boldsymbol{\theta}}_n$, in substitution of $\mathcal{Q}_{1:N}$ in Equation(10). The algorithm requires the ability to compute online the approximations $p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_{n-1}|\mathbf{y}_{1:(n-1)}, D_n)$ and $p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:(n-1),1:T})$, obtained here by an SMC approximation, as described in the next subsection. The algorithm is listed in the supplementary material, together with details regarding the computations (Sections S1 and S3).

3.1.1. *SMC approximation of the predicted probabilities.* The between-online setting processes activity the data sequentially, whenever an activity has been fully observed. The purpose of the between-online setting is to leverage existing proposals in the literature for online parameter estimation and changepoint identification, which is useful both for retrospective performance analysis and as an analysis tool in the within-online setting. We review here the principles underlying the algorithm proposed by Yildirim, Singh and Doucet (2013) and derive the computations that lead to our algorithm for changepoint detection, details of which are given in the supplementary material (Section S1). Suppose that $p_{\boldsymbol{\theta}}(D_{n-1}, |\mathbf{y}_{1:(n-1),1:T})$ is known. The quantity

$$\begin{aligned} p_{\boldsymbol{\theta}}(D_n, |\mathbf{y}_{1:(n-1),1:T}) &= \sum_{D_{n-1} \in \mathcal{D}_{n-1}} p_{\boldsymbol{\theta}}(D_n, D_{n-1}|\mathbf{y}_{1:(n-1),1:T}) \\ &= \sum_{D_{n-1} \in \mathcal{D}_{n-1}} p(D_n|D_{n-1}) p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}) \end{aligned} \tag{13}$$

can be used to derive exactly

$$\begin{aligned} p_{\boldsymbol{\theta}}(D_{n-1}|D_n, \mathbf{y}_{1:(n-1),1:T}) &= \frac{p_{\boldsymbol{\theta}}(D_n, D_{n-1}|\mathbf{y}_{1:(n-1),1:T})}{\sum_{D'_{n-1} \in \mathcal{D}_{n-1}} p_{\boldsymbol{\theta}}(D_n, D'_{n-1}|\mathbf{y}_{1:(n-1),1:T})} \\ &= \frac{p_{\boldsymbol{\theta}}(D_n|D_{n-1}) G_{\boldsymbol{\theta},n}(D_{n-1}) p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-2),1:T})}{\sum_{D'_{n-1} \in \mathcal{D}^{n-1}} p_{\boldsymbol{\theta}}(D_n|D'_{n-1}) G_{\boldsymbol{\theta},n}(D'_{n-1}) p_{\boldsymbol{\theta}}(D'_{n-1}|\mathbf{y}_{1:(n-2),1:T})} \end{aligned}$$

and

$$p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:n,1:T}) = \frac{G^D_{\boldsymbol{\theta},n}(D_n) p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})}{\sum_{D'_n \in \mathcal{D}_n} G^D_{\boldsymbol{\theta},n}(D_n) p_{\boldsymbol{\theta}}(D'_n, |\mathbf{y}_{1:(n-1),1:T})}$$

where $G^D_{\boldsymbol{\theta},n}(D_n) = p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:T}|D_n, \mathbf{y}_{1:(n-1),1:T})$. It is important to note that, although the involved quantities can be obtained exactly, computing Equation (13) has complexity $O(n)$, as $p(D_n|D_{n-1}) \neq 0$ for $2(n-1)$ combinations of (D_n, D_{n-1}). Hence, the online exact computation for a large panel of activities may be impractical in many situations, as the complexity increases with new activities.

Let $\eta^B_{n-1}(D_{n-1})$ be a particle approximation of $p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-2),1:T})$, composed of B particles with support $\mathcal{D}^B_{n-1} = \{d^1_{n-1}, \ldots, d^{B-1}_{n-1}, d^B_{n-1}\}$ composed by the particles themselves. Consider then the *augmented support* $\mathcal{D}^{B\star}_n$ of dimension $2B$ defined as

$$\mathcal{D}^{B\star}_n = \{(1, d^1_{n-1}), (d^1_{n-1}+1, d^1_{n-1}), \ldots, (1, d^B_{n-1}), (d^B_{n-1}+1, d^B_{n-1})\}.$$

An approximation of $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})$ can be obtained by sampling B independent particles from $\mathcal{D}^{B\star}_n$ with weight $W(D_n, D_{n-1}) \propto p(D_n|D_{n-1}) G^D_{\boldsymbol{\theta},n-1}(D_{n-1}) \eta^B_{n-1}(D_{n-1})$, and then marginalizing with respect to D_{n-1}. Let $\mathcal{D}^B_{(n,n-1)} = \{(d^1_n, d^1_{n-1}), \ldots, (d^B_n, d^B_{n-1})\}$

be the B sampled particles. The approximation of $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})$ is $\eta_n^B(D_n) = \sum_{b=1}^B \delta_{D_n}(d_n^b, d_{n-1}^b)$, with support $\mathcal{D}_n^B = \{d_n^1 \ldots, d_n^B\}$, where $\delta_{D_n}(d_n^b, d_{n-1}^b) = 1$ if $D_n = d_n^b$, and 0 otherwise. Moreover, $p_{\boldsymbol{\theta}}(D_{n-1}|D_n, \mathbf{y}_{1:(n-1),1:T})$ is approximated by

$$p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}, D_n) = \frac{p(D_n|D_{n-1}) G^D_{\hat{\boldsymbol{\theta}}_{n-1},n-1}(D_{n-1}) \eta^B_{n-1}(D_{n-1})}{\sum_{D'_{n-1} \in \mathcal{D}^B_{n-1|n}} p(D_n|D_{n-1}) G^D_{\hat{\boldsymbol{\theta}}_{n-1},n-1}(D'_{n-1}) \eta^B_{n-1}(D_{n-1})},$$

and $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:n,1:T})$ by

(14)

$$p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:n,1:T}) = \frac{\sum_{D_{n-1} \in \mathcal{D}^B_{n-1}} G^D_{\hat{\boldsymbol{\theta}}_{n-1},n}(D_n) p(D_n|D_{n-1}) \eta^B_{n-1}(D_{n-1})}{\sum_{(D'_n, D'_{n-1}) \in \mathcal{D}^B_{(n,n-1)}} G^D_{\hat{\boldsymbol{\theta}}_{n-1},n}(D'_n) p(D'_n|D'_{n-1}) \eta^B_{n-1}(D'_{n-1})},$$

over the supports $\mathcal{D}^B_{(n,n-1)}$ and $\mathcal{D}^B_n$, respectively, where the index $\hat{\boldsymbol{\theta}}_{1:(n-1)}$ highlights the fact that the approximations are obtained via a sequence of parameter's updates.

3.1.2. *Maximization step and inner expectations.* The maximization step in Equation (10) that attempts to solve $\frac{\partial Q_{1:N}(\boldsymbol{\theta},\boldsymbol{\theta}')}{\partial \boldsymbol{\theta}} = \mathbf{0}$ requires the computation of the derivative with respect to the elements $\mathbf{Z}^{(S)}_{\boldsymbol{\theta}}$, $\mathbf{Z}^{(A)}_{\boldsymbol{\theta}}$, $\mathbf{T}^{(S)}_{\boldsymbol{\theta}}$, $\mathbf{T}^{(A)}_{\boldsymbol{\theta}}$, $\boldsymbol{\Delta}_{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, and $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$, before applying the chain rule to obtain the derivative with respect to $\boldsymbol{\theta}$. These computations involve a finite set of elementary operations and the knowledge of both the inner and the outer expectations in Equation (9). In the online setting, the SMC approximation allows to compute the outer expectation conditional on the available data, while the inner expectation can be obtained by considering that, conditioned on $S_{1:n}$, the model for the s–th segment in Equations (4) and (5) is a linear Gaussian state space model.

These quantities can be generally obtained by standard Kalman recursions, such as the Kalman smoother and the lagged smoother proposed, for example, by Durbin and Koopman (2012) and Shumway and Stoffer (2017). Indeed, let us condition on $S_{1:n}$ or, equivalently, on the sequence of delays $D_{1:n}$. By the independence assumption between activities of different segments, it can be shown that $\iota_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:T})$ depends only on the activities that belong to the last segment. Let us define $L_{\boldsymbol{\theta}'}(\mathbf{y}_{j:n,1:T}) = \mathrm{E}_{\boldsymbol{\theta}'}\left[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:T}, \boldsymbol{\alpha}^{j:n}_{1:T}|D_n)|\mathbf{y}_{j:n,1:T}, D_n\right]$, with $j = \max(1, n - D_n + 1)$. The quantity $\iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})$ is exactly

$$\iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T}) = \begin{cases} 1 - \lambda + L_{\boldsymbol{\theta}'}(\mathbf{y}_{j:n,1:T}) - L_{\boldsymbol{\theta}'}(\mathbf{y}_{j:(n-1),1:T}) & \text{if } j = \max(1, n - D_n + 1) < n \\ \lambda + L_{\boldsymbol{\theta}'}(\mathbf{y}_{j:n,1:T}) & \text{if } j = \max(1, n - D_n + 1) = n, \end{cases}$$

where $L_{\boldsymbol{\theta}'}(\mathbf{y}_{j:n,1:T})$ depends on the expectations $\mathrm{E}_{\boldsymbol{\theta}'}\left[\boldsymbol{\alpha}^{j:n}_t|D_n, \mathbf{y}_{j:n,1:T}\right]$, $\mathrm{E}_{\boldsymbol{\theta}'}\left[(\boldsymbol{\alpha}^{j:n}_t)(\boldsymbol{\alpha}^{j:n}_t)'|\right.$ $\left.D_n, \mathbf{y}_{j:n,1:T}\right]$, and $\mathrm{E}_{\boldsymbol{\theta}'}\left[(\boldsymbol{\alpha}^{j:n}_{t+1})(\boldsymbol{\alpha}^{j:n}_t)'|D_n, \mathbf{y}_{j:n,1:T}\right]$, which are computed using the standard Kalman filtering, smoothing, and lagged smoothing routines(see, e.g., Shumway and Stoffer, 2017; Durbin and Koopman, 2012), reviewed also in the supplementary material (Section S3).

3.1.3. *Computational cost in the between-online setting.* The computational complexity for obtaining the SMC approximation is linear with respect to the number of particles B, as we may need to evaluate at most B distinct quantities in the numerator of Equation (14). Furthermore, since B is fixed during the estimation procedure, the complexity of our changepoint identification method is also linear with respect to the number of activities N. This is one of the main benefits of adopting approaches based on the SMC approximations for

changepoint identification, as they allow to lower the computational complexity of the procedure, which otherwise increases quadratically with N (see, e.g. Yildirim, Singh and Doucet, 2013; Fearnhead and Liu, 2007; Adams and MacKay, 2007). The evaluation of $G^D_{\hat{\boldsymbol{\theta}}_{n-1},n}(D_n)$ in Equation (14) represents another possible bottleneck for online changepoint identification. Indeed, to obtain $G^D_{\hat{\boldsymbol{\theta}}_{n-1},n}(D_n)$, the evaluation of the likelihood in Equation (6) by means of a Kalman filter routine whose complexity depends on the system considered, the delay D_n, the number of measurements P, and the length T of the activities is required. This computation has linear cost with respect to T, but requires the inversion of the matrix $\mathbf{F}_{s,t}$ for each time points t, which has $O(P^3 D_n^3)$ complexity. It is worth noting also that, when the number of measurements is large, the computational burden of the proposed method can be further reduced by observing that our specification for the dynamic evolution of the multiple time series is enough general to nest factor models (in the spirit of Durbin and Koopman, 2012) and related efficient computation techniques, such as the reduction by transformation technique by Jungbacker and Koopman (2008). Nevertheless, we aware that, in practice, the number of measurements collected by current smartwatches is typically limited, as it involves no more than ten variables (see Bourdon et al., 2017). On the contrary, the magnitude of D_n is a random quantity, and depends on how much the observed process is segmented by distinct changepoints. A more segmented process leads to lower delays D_n, and consequently to a lower computational cost. In general, improvements and speed-ups can be achieved by linear algebra tricks, or by considering ad hoc techniques for linear state space models (see, e.g., the univariate treatment of multivariate time series in Durbin and Koopman, 2012). Finally, the complexity of the maximization step is strongly related on the considered model specification model considered. Since the computation of the computing the overall computational cost for the online EM algorithm also involves random quantities (such as the number of distinct particles and delays), we also investigate these aspects by performing additional simulation, reported in the supplementary material (Section S2), accompanying the paper.

3.2. *Monitoring new activities in the within-online setting.* The ability to monitor the presence of a changepoint during activity n is given by the need of computing, on the fly, $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{n,1:t},\mathbf{y}_{1:(n-1),1:T})$ for any $t<T$. Note that the activities $\mathbf{y}_{1:(n-1),1:T}$ have already been observed completely, $\mathbf{y}_{n,1:t}$ is the n-th activity that is being observed, and the interest resides in checking whether $D_n=1$ or not. This allows knowledge of the status of the athlete during an activity, while also accounting for their already observed past. The direct use of the Bayes formula gives

(15)
$$p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{n,1:t},\mathbf{y}_{1:(n-1),1:T}) = \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n,\mathbf{y}_{1:(n-1),1:T})p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})}{\sum_{D'_n\in\mathcal{D}^n} p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D'_n,\mathbf{y}_{1:(n-1),1:T})p_{\boldsymbol{\theta}}(D'_n|\mathbf{y}_{1:(n-1),1:T})}.$$

An approximation of the predicted probability $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})$ is given by the SMC approach used by our algorithm so that we now consider the element $p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n,\mathbf{y}_{1:(n-1),1:T})$. We note that

(16)
$$p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n,\mathbf{y}_{1:(n-1),1:T}) = \begin{cases} \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t},\mathbf{y}_{j:(n-1),1:T}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),1:T}|D_n)} & \text{if } D_n>1 \\ p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n) & \text{if } D_n=1 \end{cases}$$

where $j=\max(1,n-D_n+1)$, can be computed by means of Kalman filters evaluations. Indeed, if $D_n>1$,

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t},\mathbf{y}_{j:(n-1),1:T}|D_n) = p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:t}|D_n)p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),(t+1):T}|D_n,\mathbf{y}_{j:n,1:t}),$$

where $p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:t}|D_n)$ is evaluated by a filtering routine up to time t with data of activities with indices that range between j and n, and $p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),(t+1):T}|D_n, \mathbf{y}_{j:n,1:t})$ is evaluated going forward with Kalman filters that treat the element $\mathbf{y}_{n,(t+1):T}$ as missing. The need to evaluate $p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),(t+1):T}|D_n, \mathbf{y}_{j:n,1:t})$ at any time point requires the ability to perform $T-t$ step ahead Kalman filter evaluations, highlighting the potential computational problem of evaluating the likelihood for long time series (large T) and early stages (small t). One simple solution is to approximate Equation (16) with

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n, \mathbf{y}_{1:(n-1),1:T}) \propto \begin{cases} \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}, \mathbf{y}_{j:(n-1),1:(t+k)}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),1:(t+k)}|D_n)} & \text{if } D_n > 1 \\ p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n) & \text{if } D_n = 1 \end{cases}, \tag{17}$$

with $k = \min(T-t, k^\star)$ and $k^\star \geq 0$ known, assuming that

$$\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+k+1):T}|\mathbf{y}_{1:n,1:t}, \mathbf{y}_{1:(n-1),(t+1):(t+k)}, D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+k+1):T}|\mathbf{y}_{1:(n-1),1:(t+k)}, D_n)} \propto 1,$$

for any D_n and k fixed in advance. This means that whenever a new activity is observed, one needs to simply use a finite number of competing Kalman filters with fixed parameters, where their number is given by the number of different unique particles in the SMC approximation of $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})$.

In principle, the role of $k^\star$ is to go forward with Kalman filters and to evaluate information that is subsequent to time t but that has already been observed before the n-th activity. However, choosing a large $k^\star$ implies the need to proceed with Kalman filter evaluations even many instants after t. This could be a problem for contexts in which it is necessary to obtain real-time feedback quickly. A large $k^\star$ allows to go very far ahead with Kalman filter evaluations, thereby slowing down the computations. Setting $k^\star = 0$ is a practical choice to avoid slow-downs in computations. It is interesting to note that although the information regarding observations after t for activities prior to the n-th are not considered by the Kalman filters, they are used in the derivation of $\eta_n^B(D_n)$.

3.2.1. *Computational cost in the within-online setting.* The interest of the within-online is to monitor the posterior probability of changepoint in Equation (15) while the activity is being carried out. In the practical context, there is a trade-off between feasibility of the approach and accuracy of the results obtained. Given the possible limitations of current devices in storing and managing intense routines, it is worth highlighting some strategies that can help their computing power management.

We first distinguish between quantities that can be calculated before a new activity starts and those that must be necessarily calculated while the new activity is being performed. The former are those quantities that do not involve any data of the new activity $\mathbf{y}_{n,1:T}$, i.e. the predicted probability in Equation (15) obtained by means of the SMC approximation $\eta_n^B(D_n)$, and the denominator of Equation (16) when $D_n > 1$. In our treatment, the evaluation of both these quantities is done with $\boldsymbol{\theta}$ held fixed and setting $\boldsymbol{\theta}$ to $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{n-1}$, as a by-product of the online EM in the between-online setting. In this sense, the between-online setting can be considered as a stepping stone for monitoring activities in the within-online setting, as it reduces and simplifies the quantities to be calculated while a new activity is collected.

The quantities that need to be calculated with a new activity are the numerator in Equation (16), i.e. $p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}, \mathbf{y}_{j:(n-1),1:T}|D_n)$, when $D_n > 1$, and $p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n)$ if $D_n = 1$. The number of distinct elements to be evaluated is at most B, that is the number of particles used to approximate $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})$. In practice, the presence of ties among the particles permits the evaluation of only $\widetilde{B} \leq B$ of these elements, that can be evaluated with separate routines, also allowing for naive parallel computing. It is worth pointing out that the SMC

approximation in the between-online setting plays another important role in our approach, as it determines the number of activities that are required to be stored in the devices. Indeed, when evaluating $p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}, \mathbf{y}_{j:(n-1),1:T}|D_n)$, only the last D_n activities are used for the computation, so that we may need to store the data of at most $\bar{d}_n^B = \max(\mathcal{D}_n^B)$ activities, where $\mathcal{D}_n^B$ is the support of $\eta_n^B(D_n)$, highlighting an additional benefit of adopting the approach by Yildirim, Singh and Doucet (2013).

A final computational complexity aspect is related to the evaluation of the numerator $p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}, \mathbf{y}_{j:(n-1),1:T}|D_n)$ in Equation (17). For a given system of Equations (1) and (2), this complexity depends on, among other factors, t, T, P, and D_n. While the same reasoning of the between-online setting applies for D_n and P, we note that evaluating $p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}, \mathbf{y}_{j:(n-1),1:T}|D_n)$ from scratch requires T Kalman filter steps, as we may need to run a Kalman filter up to time t with data of both the past activities and the new one to obtain $p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:t}|D_n)$, and then to continue up to time T with only past activities to get $p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),(t+1):T}|D_n, \mathbf{y}_{j:n,1:t})$. In practice, while $p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:t}|D_n)$ can be obtained once per time instance t, the quantity $p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),(t+1):T}|D_n, \mathbf{y}_{j:n,1:t})$ needs to be updated every time a new observation is present. In the within-online setting, not stopping earlier Kalman filter evaluations implies quadratic complexity in T, as $T + \sum_{t=1}^{T}(T-t) = T + T(T-1)/2$ Kalman step evaluations are used. By adopting the proposed strategy, this complexity becomes linear in T (precisely $T + k^\star(T - k^\star - 1) + k^\star(k^\star + 1)/2$ Kalman step evaluations are used).

4. Simulation studies: abilities in identifying changepoints. We investigate here the performance of our proposed changepoint detection algorithm for the between- and within-online settings via a series of simulated data scenarios. In particular, we illustrate that beyond changepoint identification, and unlike other potentially competitive alternatives (see, e.g. Xie et al., 2021), our methodology can monitor online the probability of a changepoint during the activities. We fixed $N = 1000$, $T = 60, 120, 200$, $P = 2$, $S = 50$ randomly chosen changepoints and variances $\sigma_\epsilon^2 = 1$, $\sigma_\alpha^2 = 0.05$, $\sigma_d^2 = 5$, and $\rho = 0.8$. With $\boldsymbol{\alpha}_0^{(s)} = \mathbf{0}_{2P}$, $\boldsymbol{\Psi}_0 = \begin{bmatrix} 1/3 & 0.5 \\ 0.5 & 1 \end{bmatrix}$, and $\boldsymbol{\alpha}_{n,0} = \mathbf{0}_P$, we generated the shared states for each segment according to $\boldsymbol{\alpha}_{t+1}^{(s)} = \mathbf{I}_P \otimes \begin{bmatrix} 0.95 & 1 \\ 0 & 0.90 \end{bmatrix} \boldsymbol{\alpha}_t^{(s)} + \boldsymbol{\xi}_t^{(s)}$, $\boldsymbol{\xi}_t^{(s)} \sim N_{2P}(\mathbf{0}_{2P}, \sigma_\alpha^2(\mathbf{I}_P \otimes \boldsymbol{\Psi}_0))$ and the activity-specific states according to $\boldsymbol{\alpha}_{n,t+1} = \rho \cdot \boldsymbol{\alpha}_{n,t} + \boldsymbol{\xi}_{n,t}$, $\boldsymbol{\xi}_{n,t} \sim N_P(\mathbf{0}_P, \sigma_d^2 \mathbf{I}_P)$. We then generated the observations $\mathbf{y}_{n,t} = \left[\mathbf{I}_P \otimes \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{I}_P\right] \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{n,t} \end{bmatrix} + \boldsymbol{\epsilon}_{n,t}$, $\boldsymbol{\epsilon}_{n,t} \sim N_P(\mathbf{0}_P, \sigma_\epsilon^2 \mathbf{I}_P)$.

We set $\lambda = 0.5$ and estimated $\boldsymbol{\theta} = (\sigma_\epsilon^2, \sigma_\alpha^2, \sigma_d^2, \rho)$. In addition, we set $k^\star = 0$ since using a small $k^\star$ is a necessary practical choice when T is large. As $k^\star$ increases, the proposed algorithm for the within-online setting becomes infeasible for large T. Alternative specifications of $k^\star$ are investigated in the supplementary material together with alternative specifications of λ (Section S4).

We estimate (i) the changepoints in the between-online setting by utilizing Equation (14) and testing $p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n = 1|\mathbf{y}_{1:n,1:T}) > \delta$ for some threshold δ and (ii) the probability of activity n of being a changepoint before it ends in the within-online setting according to $p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n = 1|\mathbf{y}_{n,1:t}, \mathbf{y}_{1:(n-1),1:T})$. Figure 3 depicts the behavior of sensitivity and specificity as the length of the time series increases, leaving the remaining elements of the models unchanged. We deal with the usual trade-off between sensitivity and specificity by noting that in our application, maximizing sensitivity —which is minimizing the number of activities that are wrongly classified as negative— is more important, as it may indicate possible activity problems. This is naturally controlled by the threshold δ, see Figure 3. It is reassuring

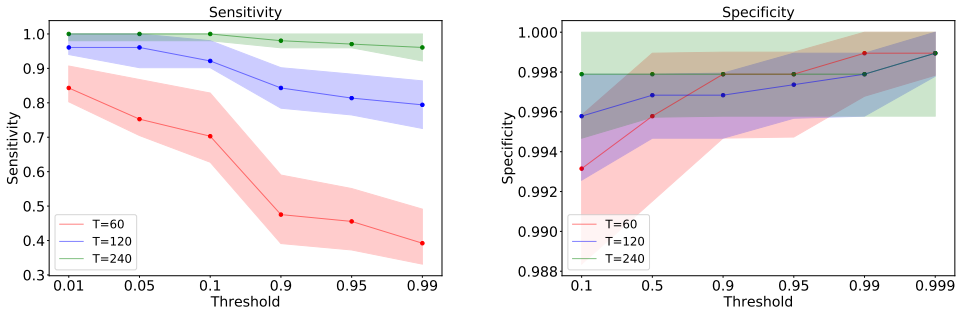

FIG 3. *Medians and* 90% *confidence intervals of sensitivity and specificity (y-axis), as a function of the threshold level* δ *(x-axis), obtained for* 20 *synthetic examples using our model for* $T = 60, 120$ *and* 240.

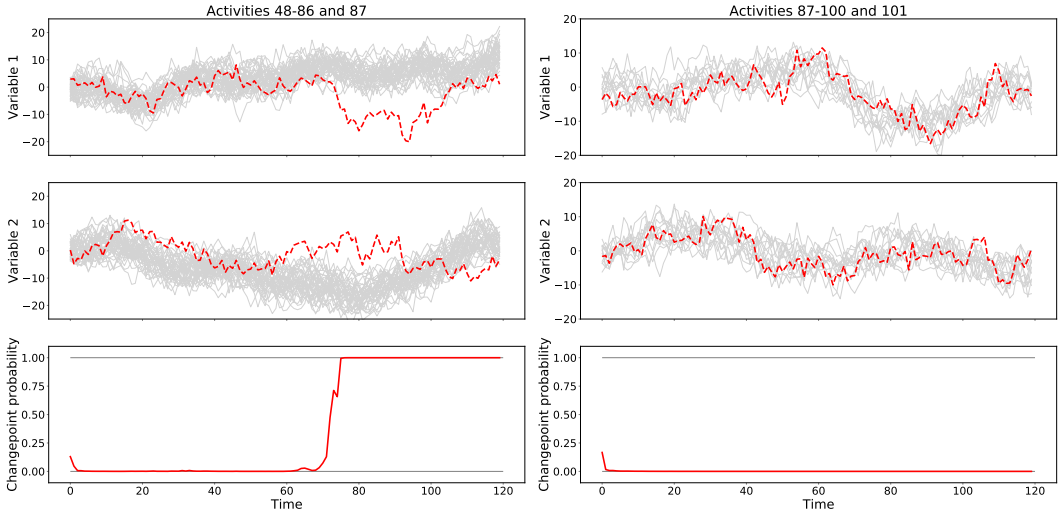

FIG 4. *Two instantiations of our simulation with different activities and the respective filtered probabilities of changepoints. Red dashed line: activity that is being monitored; gray lines: previous activities since the last changepoint.*

that our algorithm maintains high levels of specificity as δ changes, regardless of the length of the time series. In contrast, the sensitivity seems to decrease significantly as δ increases, particularly for $T = 60$ and $T = 120$, although it remains stable for $T = 240$.

The within-online setting allows to monitor online the probability of an activity changepoint, providing information on the athlete's behavior with respect to the past. Figure 4 shows two instantiations of this: the filtered probability $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(\mathbf{y}_{n,1:t}|D_n, \mathbf{y}_{1:(n-1),1:T})$, depicted in the bottom row, is estimated online as new observations are collected for the two simulated activities (top two rows). Each panel shows the current (dashed red line) and previous (solid gray) activities since the last changepoint. In the within-online setting, the changepoint detection is performed by estimating changepoint probabilities for various values of δ and t in 20 replications of the experiment. Figure 5 depicts the results of the simulation study in terms of sensitivity and specificity for different values of δ and t. As expected, the sensitivity drops as δ increases and as t decreases. Since all time series were simulated with initial values around zero, it is hard to achieve an early (at $t = 40$) changepoint detection, although the detection after having observed $2/3$ of the time series (at $t = 80$) seems to be satisfactory.

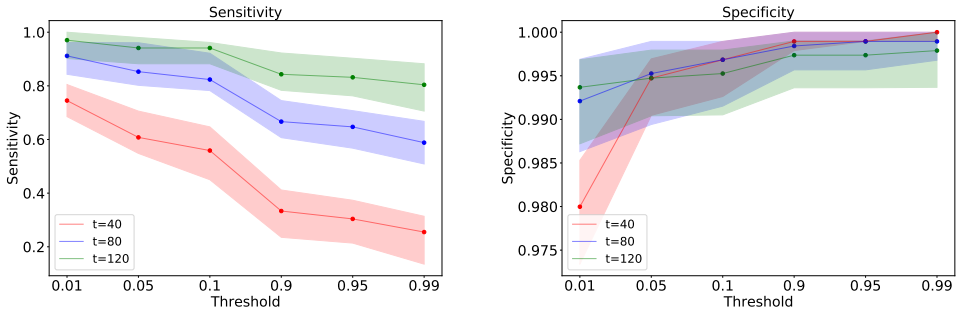

FIG 5. *Medians and* 90% *confidence intervals of sensitivity and specificity (y-axis), as a function of the threshold level* δ *(x-axis), evaluated for* 20 *synthetic examples using our model in the within-online setting.*

5. Application. We consider a set of 85 warm-up running activities on flat routes consisting of the first 10 minutes of running of a well-trained athlete. The difference between the maximum and minimum altitude reached during each activity was less than 10 meters, and the activities were measured every second by a Polar v800 smart watch and a Polar H10 heart rate monitor. Warm-up activities are extremely relevant in several sports because they prepare athletes for specific training sessions, influence sports performance, and reduce the risk of injury. Moreover, they inform on the training status of an athlete just before the training session so early decisions can be made. In the sports science literature, the choice of the relevant indicators for monitoring the health status and training loads with emphasis on the importance of pre-training analysis is well documented, see, for example, Buchheit (2014). In general, heart rate is the most evaluated variable, as it provides insights into oxygen consumption and the physical response to the external stimuli of the exercise (Dong, 2016; Schneider et al., 2018). Heart rate levels during exercise are also influenced by the intensity at which the exercise is performed, represented by the speed of running, which is why we have collected data for both heart rate and speed. Let $y_{\text{hr},n,t}$ be the heart rate in beats per minute and $y_{\text{sp},n,t}$ be the speed, equal to the difference between the cumulative distances at time t and $t-1$ for activity n. We specify a state space model with the measurement equation

$$\begin{bmatrix} y_{\text{hr},n,t} \\ y_{\text{sp},n,t} \end{bmatrix} = \begin{bmatrix} 1\,0\,0 \\ 0\,0\,1 \end{bmatrix} \begin{bmatrix} \alpha^{(s)}_{\text{hr},1,t} \\ \alpha^{(s)}_{\text{hr},2,t} \\ \alpha^{(s)}_{\text{sp},t} \end{bmatrix} + \begin{bmatrix} 1\,0 \\ 0\,1 \end{bmatrix} \begin{bmatrix} \alpha_{\text{hr},n,t} \\ \alpha_{\text{sp},n,t} \end{bmatrix} + \begin{bmatrix} \upsilon_{\text{hr},n,t} \\ \upsilon_{\text{sp},n,t} \end{bmatrix}, \quad \begin{bmatrix} \upsilon_{\text{hr},n,t} \\ \upsilon_{\text{sp},n,t} \end{bmatrix} \sim N_2(0, \boldsymbol{\Sigma}),$$

segment-specific state equations

$$\begin{bmatrix} \alpha^{(s)}_{\text{hr},1,t+1} \\ \alpha^{(s)}_{\text{hr},2,t+1} \\ \alpha^{(s)}_{\text{sp},t+1} \end{bmatrix} = \begin{bmatrix} 1\,1\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{bmatrix} \begin{bmatrix} \alpha^{(s)}_{\text{hr},1,t} \\ \alpha^{(s)}_{\text{hr},2,t} \\ \alpha^{(s)}_{\text{sp},t} \end{bmatrix} + \begin{bmatrix} \xi^{(s)}_{\text{hr},1,t} \\ \xi^{(s)}_{\text{hr},2,t} \\ \xi^{(s)}_{\text{sp},t} \end{bmatrix}, \quad \begin{bmatrix} \xi^{(s)}_{\text{hr},1,t} \\ \xi^{(s)}_{\text{hr},2,t} \\ \xi^{(s)}_{\text{sp},t} \end{bmatrix} \sim N_3(0, \boldsymbol{\Psi}),$$

and activity-specific state equations

$$\begin{bmatrix} \alpha_{\text{hr},n,t+1} \\ \alpha_{\text{sp},n,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \rho_{\text{sp}} \end{bmatrix} \begin{bmatrix} \alpha_{\text{hr},n,t} \\ \alpha_{\text{sp},n,t} \end{bmatrix} + \begin{bmatrix} \xi_{\text{hr},n,t} \\ \xi_{\text{sp},n,t} \end{bmatrix}, \quad \begin{bmatrix} \xi_{\text{hr},n,t} \\ \xi_{\text{sp},n,t} \end{bmatrix} \sim N_2(0, \boldsymbol{\Delta}),$$

with $\boldsymbol{\alpha}^{(s)}_1 = (\alpha^{(s)}_{\text{hr},1,1}, \alpha^{(s)}_{\text{hr},2,1}, \alpha^{(s)}_{\text{sp},1})' \sim N_3((75,0,0)', \text{diag}(100,10,30))$, $\boldsymbol{\alpha}_{n,1} = (\alpha_{\text{hr},n,1}, \alpha_{\text{sp},n,t})' \sim N_2(\mathbf{0}, \text{diag}(50,10))$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Delta}$ are full covariance matrices, and ρ_{sp} is an autoregressive coefficient.

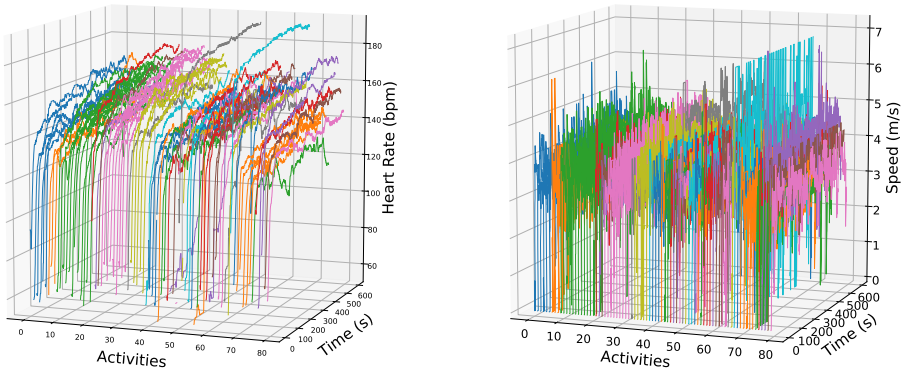

FIG 6. *Segmentation of warm-up activities in the between-online setting for an athlete. The segmentation of the activities was obtained defining the changepoint as those activities for which the filtered distribution at the end of the activity is* $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n = 1|\mathbf{y}_{1:n,1:T}) > 0.5$.

The segment-specific latent states that describe the physical condition and skills of the athlete were chosen to be modeled by a linear trend model that captures the segment-specific global trends for the heart rate and by a local level model for the speed. The local linear trend for the heart rate variable is composed by two latent states: the trend $\alpha^{(s)}_{\text{hr},1,t}$ and the drift $\alpha^{(s)}_{\text{hr},2,t}$. The trend describes how heart rate evolves over time if no disturbing aspects are present during the activities (activity-specific latent states). The drift $\alpha^{(s)}_{\text{hr},2,t}$ determines how trend direction evolves over time. Indeed, if $\alpha^{(s)}_{\text{hr},2,t}$ is constant over time, then the trend is exactly linear (see, e.g., Durbin and Koopman, 2012). The presence of positive drift is in the heart rate variable is typically observed and discussed in sports science, particularly for prolonged efforts (see, e.g., Souissi et al., 2021). In this work, we assume the drift to be stochastic to capture changes in directions of the trend due to different phases of the activities. The local level model for the speed variable allows to consider behaviors that locally evolve over time, a typical characteristic of warm-up activities that involve continuous running. The activity-specific states are modeled by a random walk process for the heart rate and using an AR(1) process for the speed. Once the variables are de-trended, the heart rate moves slowly over time, as it does not vary abruptly in healthy conditions, although speed may do so due to, for example, street obstacles. The random walk to describe the heart rate variable allows to capture variations from the segment-specific trend that may persist over time. The AR(1) process for the speed, on the contrary, describes variation from the speed level that are then reabsorbed over time. Considering AR process with different lags to describe the activity-specific behaviors is possible, at the cost of increasing the number of parameters and the complexity of the state space model.

We set $\lambda = 0.5$ and, following the guidelines of Yildirim, Singh and Doucet (2013), we estimated the parameters of the model $\boldsymbol{\theta} = \{\boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{\Delta}, \rho_{\text{sp}}\}$ by repeating the EM algorithm 50 times in order to reach convergence. Figure 6 provides an instantiation of our results. We depict segments in the between-online setting, obtained according to the rule $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n = 1|\mathbf{y}_{1:n,1:T}) > 0.50$. The estimated number of changepoints is 39, of which 24 involve activities with a single activity segment. This finding highlights the large variability

between successive activities. Of these 24 changepoints with a single activity segment, 18 are located in the last 42 activities and should be attributed not only to changes in the state of the athlete but also to the presence of systematic measurement errors, probably due to a device problem. The obtained results are robust with respect to different choices of λ and $k^\star$, that sligthly modify the changepoint probabilities computed in the within online setting. With a threshold of $\delta = 0.5$, however, the obtained segmentation at the end of the activities (i.e., in the between online setting) is robust with respect to different choices of these hyper-parameters. Varying the threshold δ has a slight effect on the number of estimated changepoints, ranging from 39 ($\delta = 0.01$) to 38 ($\delta = 0.99$). These additional aspects are reported in greater detail in the supplementary material. Note that although the threshold δ has an effect on the results obtained, in terms of retrospective changepoint identification, it is not involved in the sequential monte Carlo approach used to calculate the changepoint probabilities in our methodology, and thus has no effect in the process leading to the estimation of changepoint probabilities. This is particularly relevant to the purpose of the proposed approach, where the primary interest is in quantifying the uncertainty one has about changes in the behavior of variables, as opposed to a retrospective analysis of performance.

Figure 7 shows four instantiations of the within-online setting by presenting heart rate, speed, and changepoint probability $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n = 1|\mathbf{y}_{1:n,1:T})$ for the monitored activity (dashed red line) and for all activities subsequent to the previous changepoint (solid gray lines) for $\delta = 0.5$. In particular, activity 6 was identified as a changepoint because a different behavior was detected due to a lower heart rate during most of the activity, accompanied with a speed behavior that seems to be slightly lower with respect to the previous activities. Activity 21 was identified as a changepoint because a sub-optimal behavior was detected due to a higher heart rate (with similar speed behavior) compared with the previous activities. The bottom left panel shows activity 32, for which the changepoint probability is nearly zero during the entire activity, highlighting that the athlete is behaving as in the previous activities and that no changes are identified. Finally, the bottom right panel shows activity 39 for which heart rate is lower with respect to previous activities (with a similar speed behaviour), corresponding to the athlete which is putting in less effort. Note the ability of the algorithm to provide information about changepoint probabilities much before the end of the warm-ups, allowing an early identification of changes. A general discussion on the changepoints A general discussion on use of the model, its strengths and weaknesses is reported in the next section.

6. Justification of the multi-variables changepoint model. Our model development has emerged by the sports science perspective in which it is typically of interest jointly consider the so-called internal load variables that measure the physiological responses of the body to stimuli, and the external load variables which are physical measures of the intensity of the exercise (see, e.g. Bourdon et al., 2017). From a statistical point of view, a relevant question is to inquire whether our model (i) performs better than independent single-variable change point models and (ii) performs well in situations in which changepoints affect only one variable. Clearly, our model can be immediately applied to single variable change point scenarios so we have considered a simulation exercise to investigate these questions.

First, we simulated random changepoints that randomly affect one or two variables and we compared the sensitivity and specificity of our bivariate method with those of a "pragmatic" approach, in which we consider two separate single variable procedures that use the maximum between the respective filtered probabilities as a surrogate of the changepoint probability. We report the results of the proposed approach, for $T = 240$, in terms of sensitivity and specificity in Figure 8, in which we show how our joint approach might lead to a loss of sensitivity when changes do not affect both the variables as well. This is expected since

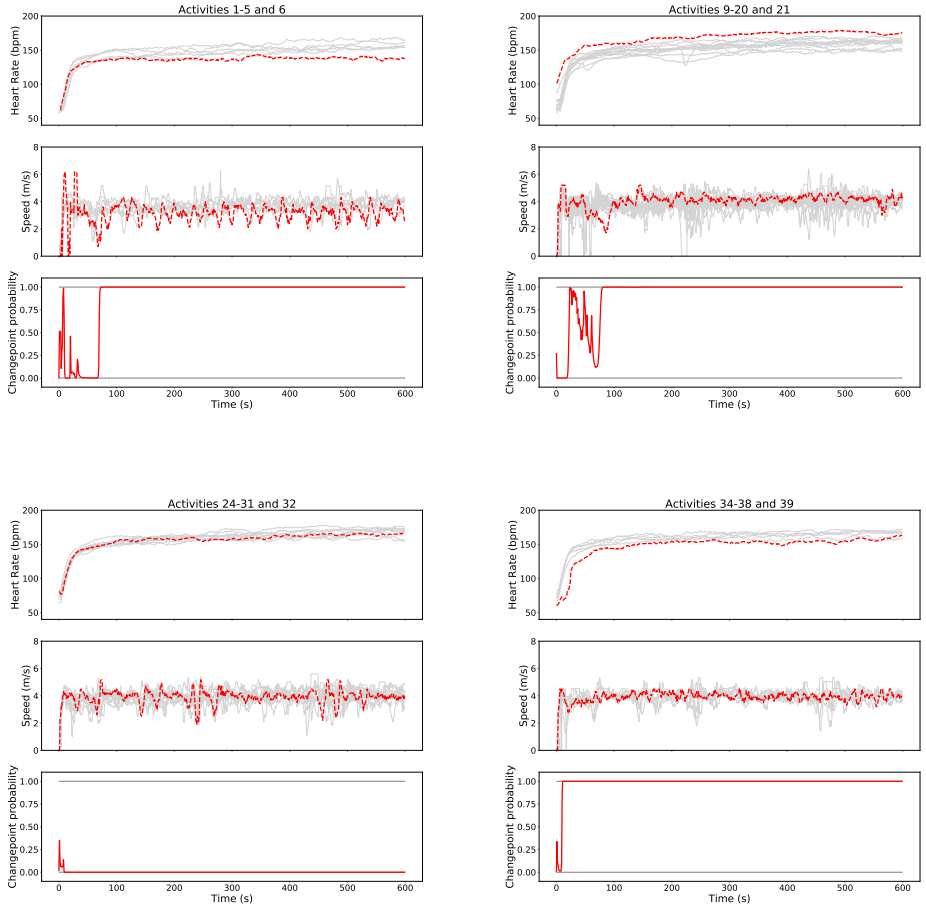

FIG 7. *Selected activities in which the changepoint probability is being monitored. The gray lines in the background represent activities since the last changepoint, obtained according to the rule* $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n = 1|\mathbf{y}_{1:n,1:T}) > 0.5$ *and* $k^\star = 0$.

our model assumes a different data generation mechanism. In a similar scenario in which the changepoints truly affect both the variables the corresponding results, depicted in Figure 9, our multiple-variables model dramatically outperforms the single variables model in terms of specificity. These results suggest how our model can be useful also for obtaining interesting insights regarding the activities. For a given threshold ($\delta = 0.5$ in our explanation), activities which are recognized as changepoints by both univariate and bivariate routines can be interpreted as activities that manifests strong changes in both the variables; in the real data application, just 2 activities were recognized as such. Similarly, when a changepoint is identified by just one univariate algorithm and by the bivariate one, the activity can be interpreted as one activity in which changes are mainly related to one variable, but are strong enough to be recognized also by the bivariate model; in the application, 10 activities registered this kind of behavior with the variable heart rate, and 15 with the variable speed. Activities that are classified as changepoints only by the bivariate model are those activities that do not show strong deviations if single variable models are considered, but that actually are characterized by large deviations when these are considered jointly. We recognize this aspect as one of the main benefit of having developed a multivariate model, as also the pragmatic approach

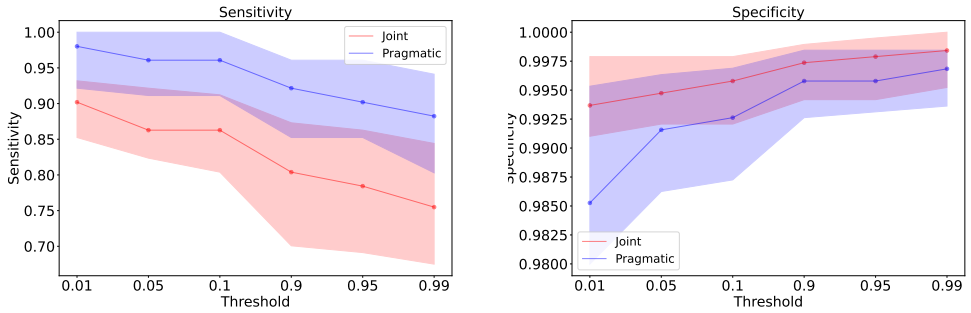

FIG 8. *Medians and* 90% *confidence intervals of sensitivity and specificity (y-axis), as a function of the threshold level* δ *(x-axis), evaluated for* 20 *synthetic examples using the bivariate method and the pragmatic approach, for* $T = 240$. *In the simulation studies, the changepoints can affect one or two variables, randomly chosen.*

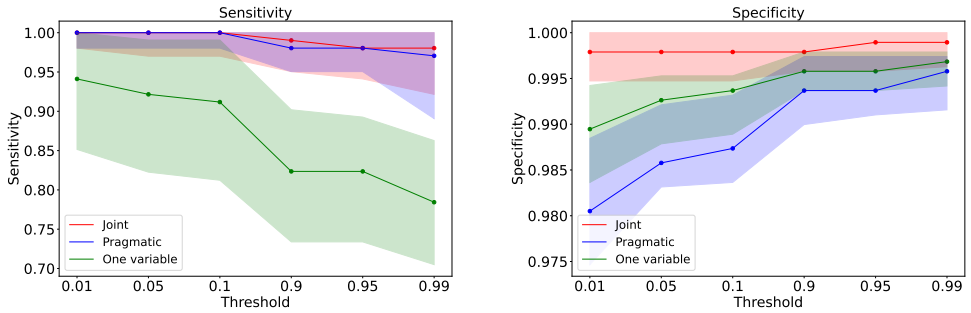

FIG 9. *Medians and* 90% *confidence intervals of sensitivity and specificity evaluated for* 20 *synthetic examples using our procedure with the two variables considered jointly, the pragmatic approach, or using our algorithm with just one variable, for* $T = 240$. *In the simulation studies, the changepoints affect both the variables.*

would fail in recognizing this behavior. In the real data, 12 activities were recognized as such. Finally, activities recognized as changepoint by only one univariate algorithm can be interpreted as activities in which there are present changes in one variable, but they are not so strong to be recognized as such also by the bivariate model. In the data, 6 activities are affected by this behavior for the variable speed.

To conclude, it is worth asking how real-time identification of changes can be used in real monitoring environments. Such an answer is not simple at all, given the high variability shown by the activities and the high number of factors coaches and athletes should take into account to make decisions. For example, extreme deviations in both the variables can suggest that the athlete is not behaving as expected, and knowing it before the activity ends would permit to adjust such behaviors. In the practical example, an extreme fatigue (the heart rate is too high) accompanied with an higher speed would imply that the warm-up is too demanding, an aspect that would lead to under-performance during the subsequent training. In a similar way, a warm-up activity which is too soft (e.g. speed and heart rate are too slow) would imply an inadequate preparation for training, an aspect considered relevant both for obtaining better performances and for avoiding the risk of injuries. Softer changepoints reveal aspects of the activities which are harder to identify, and that can be related to different health status of the athlete. Recognizing that the heart rate is too high despite the speed level can be assumed similar to previous activities, allows to identify a sub-optimal status of the athlete, and the gained time in knowing it before the activity ends would permit, on the one

side, to try to adopt running strategies to reduce the fatigue, on the other side to investigate the athlete's physical conditions and to adapt the training according to them. Finally, this aspect highlights also the importance of choosing a threshold for determining which activities are changepoints, a choice which is strongly related to the use of this tools by runners and coaches. As first, it is relevant to highlight, that the choice of such a threshold does not affect the SMC approximation that allows for obtaining changepoint probabilities, but simply permits to obtain classification of the activities based on their behavior. In general, an higher threshold helps in recognizing less changepoints, in such a way only extreme changes are recognized as such. This is useful in cases in which it is important to detect only strong deviations, and has the advantage that the interpretation of the changepoint is easier to derive. In contrast, lower thresholds allow to better capture the variability of activities, but has the drawback that interpretation of the results becomes more challenging.

7. Conclusion and future developments. Motivated by the need to develop an online probabilistic inference framework for runners who collect data using smart devices, we have proposed a new model for changepoint detection in a doubly-online framework. Our focus lies on the early detection of distributional changes between a set of repeated running activities. The proposed model combines and leverages tools from the classical changepoint model by Yildirim, Singh and Doucet (2013) and the linear and Gaussian state space model (Durbin and Koopman, 2012; Shumway and Stoffer, 2017). The former allows the use of an SMC approach with constant complexity in a between-online framework, while the latter provides the user with updated information on the activity as new data are observed by means of Kalman filter routines. We adopted a linear and Gaussian state space model, which is a general family of models that allows to include many standard modeling specifications used in time series analysis.

We considered design matrices that are fixed with respect to both t and n and potentially preclude time-dependent and activity-specific covariates. It is probably reasonable to assume that covariates such as different types of terrain or changes in elevations could affect heart rate or speed. This limitation can be easily overcome by modifying the Kalman recursions appropriately without a substantial change in the remaining methodology.

We also assumed that activity-specific elements do not interact with segment-specific latent states by imposing a block-diagonal structure on both the transition matrix and the covariance matrix of disturbances in Equation (5). One possible generalization could assume that the block transition matrix in Equation (5) is a block matrix in which the elements outside the diagonal of the first column block are non-zero. This generalization also requires a modification of the Kalman recursions without a substantial change in the methodology, also allowing for activity-specific states determined by some segment-specific states, such as, the autoregressive process with segment-specific coefficients. Details are provided in the supplementary material (Section S5).

To simplify the exposition, we have only considered time series with the same length T. The case of different length activities $T_n < T = \max(T_1, \ldots, T_N)$ for some n can be readily considered, see the supplementary material (Section S2). Activities with shorter lengths are simply treated as activities in which the last $T - T_n$ observations are missing. In a real-world wearable data collection settings, activities will clearly have different lengths but, from a sport science perspective, input data analyzed by the algorithm should be of the same type of exercise, and typically of similar time length, so the physical and physiological responses to stimuli are expected to be comparable.

The changepoint prior probability λ can be modeled as $\lambda_{\boldsymbol{\theta}} = \lambda_{\boldsymbol{\theta},n} = \lambda_{\boldsymbol{\theta}}(\mathcal{X}_n)$ depending on a set $\mathcal{X}_n$ of time-invariant activity-specific covariates. The set $\mathcal{X}_n$ may represent the meteorological condition during the activity or health-related measures taken prior to the activity,

such as heart rate variability in the morning or the number of hours of sleep. This generalization requires the computation of sufficient statistics and a maximization step that is dependent on the specification of the link function. Both developments are generally related to the standard methods used for the binomial model; see Yildirim, Singh and Doucet (2013) for details.

A particularly appealing possible future development that requires additional methodological effort is to consider nonlinear and non-Gaussian state space models. This might be of interest in contexts in which the use of smart devices allows for the collection of varied data (Bourdon et al., 2017), violating the common Gaussian assumptions.

Acknowledgments. The authors gratefully acknowledge the cardiologist Dr Costas Thomopoulos for valuable discussions.

Funding. This research was supported by funding from the University of Padova Research Grant BIRD203991.

SUPPLEMENTARY MATERIAL

Supplementary 1: `Supplementary1.zip`
Code and data are provided in `Supplementary1.zip` file with instructions.

Supplementary 2
Details on computations and application, including different choices of λ, δ and $k^\star$.

REFERENCES

ADAMS, R. P. and MACKAY, D. J. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*.

AMINIKHANGHAHI, S. and COOK, D. J. (2017). A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **51** 339–367.

BOURDON, P. C., CARDINALE, M., MURRAY, A., GASTIN, P., KELLMANN, M., VARLEY, M. C., GABBETT, T. J., COUTTS, A. J., BURGESS, D. J., GREGSON, W. et al. (2017). Monitoring athlete training loads: consensus statement. *Int. J. Sports Physiol.* **12** 161–170.

BUCHHEIT, M. (2014). Monitoring training status with HR measures: do all roads lead to Rome? *Front. physiol* **5** 73.

CAPPÉ, O. and MOULINES, E. (2009). On-line expectation-maximization algorithm for latent data models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 593–613. MR2749909

CARON, F., DOUCET, A. and GOTTARDO, R. (2012). On-line changepoint detection and parameter estimation with application to genomic data. *Stat. Comput.* **22** 579–595. MR2865037

CHIB, S. (1998). Estimation and comparison of multiple change-point models. *J. Econometrics* **86** 221–241. MR1649222

DE CHAUMARAY, M. D. R., MARBAC, M. and NAVARRO, F. (2020). Mixture of hidden Markov models for accelerometer data. *Ann. Appl. Stat.* **14** 1834–1855.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. With discussion. MR501537

DENEVI, G., STAMOS, D., CILIBERTO, C. and PONTIL, M. (2019). Online-Within-Online Meta-Learning. In *Advances in Neural Information Processing Systems* (H. WALLACH, H. LAROCHELLE, A. BEYGELZIMER, F. D' ALCHÉ-BUC, E. FOX and R. GARNETT, eds.) **32**. Curran Associates, Inc.

DONG, J.-G. (2016). The role of heart rate variability in sports physiology. *Exp. Ther. Med.* **11** 1531–1536.

DURBIN, J. and KOOPMAN, S. J. (2012). *Time series analysis by state space methods*, second ed. *Oxford Statistical Science Series* **38**. Oxford University Press, Oxford. MR3014996

ELLIOTT, R. J., FORD, J. J. and MOORE, J. B. (2002). On-line almost-sure parameter estimation for partially observed discrete-time linear systems with known noise characteristics. *International Journal of Adaptive Control and Signal Processing* **16** 435-453.

FEARNHEAD, P. and LIU, Z. (2007). On-line inference for multiple changepoint problems. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 589-605.

FREE, C., PHILLIPS, G., GALLI, L., WATSON, L., FELIX, L., EDWARDS, P., PATEL, V. and HAINES, A. (2013). The Effectiveness of Mobile-Health Technology-Based Health Behaviour Change or Disease Management Interventions for Health Care Consumers: A Systematic Review. *PLoS Med.* **10** 1-45.

FRICK, H. and KOSMIDIS, I. (2017). trackeR: Infrastructure for Running and Cycling Data from GPS-Enabled Tracking Devices in R. *J. Stat. Softw.* **82** 1–29.

GARCÍA-GUZMÁN, J. J., PÉREZ-RÀFOLS, C., CUARTERO, M. and CRESPO, G. A. (2021). Microneedle based electrochemical (Bio)Sensing: Towards decentralized and continuous health status monitoring. *Trends Anal. Chem.* **135** 116148.

GRUNDY, T., KILLICK, R. and MIHAYLOV, G. (2020). High-dimensional changepoint detection via a geometrically inspired mapping. *Stat. Comput.* **30** 1155–1166. MR4108696

HAYNES, K., FEARNHEAD, P. and ECKLEY, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Stat. Comput.* **27** 1293–1305. MR3647098

HUANG, L., BAI, J., IVANESCU, A., HARRIS, T., MAURER, M., GREEN, P. and ZIPUNNIKOV, V. (2019). Multilevel Matrix-Variate Analysis and its Application to Accelerometry-Measured Physical Activity in Clinical Populations. *J. Am. Stat. Assoc.* **114** 553-564.

JUNGBACKER, B. and KOOPMAN, S. J. (2008). Likelihood-based analysis for dynamic factor models Technical Report, Tinbergen Institute Discussion Paper.

KANTAS, N., DOUCET, A., SINGH, S. S., MACIEJOWSKI, J. and CHOPIN, N. (2015). On particle methods for parameter estimation in state-space models. *Statist. Sci.* **30** 328–351. MR3383884

KITAGAWA, G. (1987). Non-gaussian state—space modeling of nonstationary time series. *Journal of the American statistical association* **82** 1032–1041.

LUO, L. and SONG, P. X. K. (2020). Renewable estimation and incremental inference in generalized linear models with streaming data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 69–97. MR4060977

PELLICCIA, A., SHARMA, S., GATI, S., BÄCK, M., BÖRJESSON, M., CASELLI, S., COLLET, J.-P., CORRADO, D., DREZNER, J. A., HALLE, M. et al. (2021). 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease: The Task Force on sports cardiology and exercise in patients with cardiovascular disease of the European Society of Cardiology (ESC). *Eur. Heart J.* **42** 17–96.

PKVITALITY (2020). *PKvitality faqs*. https://www.pkvitality.com/.

QIAN, T., YOO, H., KLASNJA, P., ALMIRALL, D. and MURPHY, S. A. (2020). Estimating time-varying causal excursion effects in mobile health with binary outcomes. *Biometrika* **108** 507-527.

SAMÉ, A. and GOVAERT, G. (2017). Segmental dynamic factor analysis for time series of curves. *Stat. Comput.* **27** 1617–1637. MR3687329

SCHNEIDER, C., HANAKAM, F., WIEWELHOVE, T., DÖWELING, A., KELLMANN, M., MEYER, T., PFEIFFER, M. and FERRAUTI, A. (2018). Heart rate monitoring in team sports—a conceptual framework for contextualizing heart rate measures for training and recovery prescription. *Front. physiol* **9** 639.

SHUMWAY, R. H. and STOFFER, D. S. (2017). *Time series analysis and its applications*, fourth ed. *Springer Texts in Statistics*. Springer, Cham With R examples. MR3642322

SINGH, N., MONEGHETTI, K. J., CHRISTLE, J. W., HADLEY, D., FROELICHER, V. and PLEWS, D. (2018). Heart rate variability: an old metric with new meaning in the era of using mhealth technologies for health and exercise training guidance. Part two: prognosis and training. *AER* **7** 247.

SIQUEIRA DO PRADO, L., CARPENTIER, C., PREAU, M., SCHOTT, A.-M. and DIMA, A. L. (2019). Behavior Change Content, Understandability, and Actionability of Chronic Condition Self-Management Apps Available in France: Systematic Search and Evaluation. *JMIR Mhealth Uhealth* **7** e13494.

SONG, P. X.-K. (2007). *Correlated data analysis: modeling, analytics, and applications* **1**. Springer.

SOUISSI, A., HADDAD, M., DERGAA, I., BEN SAAD, H. and CHAMARI, K. (2021). A new perspective on cardiovascular drift during prolonged exercise. *Life Sciences* **287** 120109.

STATISTA (2020a). *Running and Jogging - Statistics and Facts*. https://www.statista.com/topics/1743/running-and-jogging/#dossierSummary.

STATISTA (2020b). *Wearables dossier*. https://www.statista.com/study/15607/wearables-statista-dossier/.

STATISTA (2020c). *eServices Report 2020 - Fitness*. https://www.statista.com/study/36674/fitness-report/.

R CORE TEAM (2020). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

TITSIAS, M. K., SYGNOWSKI, J. and CHEN, Y. (2020). Sequential Changepoint Detection in Neural Networks with Checkpoints. *arXiv preprint arXiv:2010.03053*.

VITABILE, S., MARKS, M., STOJANOVIC, D., PLLANA, S., MOLINA, J. M., KRZYSZTON, M., SIKORA, A., JARYNOWSKI, A., HOSSEINPOUR, F., JAKOBIK, A., STOJNEV ILIC, A., RESPICIO, A., MOLDOVAN, D., POP, C. and SALOMIE, I. (2019). *Medical Data Processing and Analysis for Remote Health and Activities*

Monitoring In *High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet* 186–220. Springer International Publishing, Cham.

XIE, L., ZOU, S., XIE, Y. and VEERAVALLI, V. V. (2021). Sequential (Quickest) Change Detection: Classical Results and New Directions. *JSAIT* **2** 494-514.

YILDIRIM, S., SINGH, S. S. and DOUCET, A. (2013). An online expectation-maximization algorithm for change-point models. *J. Comput. Graph. Statist.* **22** 906–926. MR3173749