UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXIV

Sports performance analysis with state space models

Coordinatore del Corso: Prof. Nicola Sartori

Supervisore: Prof. Mauro Bernardi

Co-supervisore: Prof. Petros Dellaportas

Dottorando: Mattia Stival

12 January 2022

Abstract

The study of sports performances is a topic of paramount importance in sports sciences, in which the role of data has been always fundamental. The evaluation of athletes' competition, for example, can be done on the basis of quantitative measurements of their performances, useful for obtaining the subsequent rankings. If in principle, according to this interest, various methods and approaches have been developed by whom was directly involved in the field, the progress in technology has attracted researches from other domains to this topic. Mathematicians, engineers, computer scientists, and statisticians are involved in different aspects of sports science, both in developing technological tools useful in collecting and using data and in answering to research questions of various levels of complexity.

The aim of this thesis is to provide statistical tools that can be used in analyzing sports performances, with a particular reference to the employment of state space models and time series analysis. The present thesis is composed of four chapters: the first two provide an overview of the treated topics; the remaining chapters present the main contributions of this work. In particular, the first chapter includes a general discussion of sports performances analysis. The second chapter presents selected tools and models useful in the time series analysis. In the third chapter, a Bayesian clustering model is presented in order to describe the personal best performances of Italian middle distance athletes. In more detail, the chapter provides a state space matrix model in which several multivariate trajectories of different athletes have been grouped on the basis of the trend of their performance and the pattern of missing data observed in the sample, this last considered as indicator of personal history and attitudes of athletes. The inference is conducted through a Markov Chain Monte Carlo simulation algorithm. The application on real data shows benefits and limitations of the proposed approach and it provides indications on which factors are relevant in order to obtain better sports

performances. The fourth chapter describes a model for monitoring the health status during sports activities. The inference has been conducted using an online Expectation-Maximization algorithm involving a sequential Monte Carlo approximation of changepoint predicted probabilities. As a byproduct of our model assumptions, the proposed algorithm processes sequence of time series in a doubly-online framework. While changepoint models identify changes between subsequent activities, the state space formulation of the model, together with the proposed algorithm, provides the additional benefit of estimating changepoint probability in real-time.

Sommario

Lo studio delle performance sportive è un argomento di notevole importanza nelle scienze motorie, in cui il ruolo dell'utilizzo dei dati è sempre stato fondamentale. La stessa valutazione di una gara di uno sportivo, per esempio, viene svolta a partire da misurazioni quantitative delle sue performance, sulla base delle quali vengono stilate poi le classifiche. Se all'inizio, a fronte di questo interesse, vari metodi ed approcci sono stati sviluppati negli anni da chi era direttamente coinvolto nell'ambito, il progresso della tecnologia ha avvicinato a questo campo studiosi e ricercatori di altri ambiti di ricerca. Matematici, ingegneri, informatici e statistici sono coinvolti in vari aspetti di questa disciplina, che li vede partecipi sia nello sviluppo di strumenti tecnologici utili alla raccolta stessa dei dati e al loro utilizzo, che nel rispondere a domande di ricerca con vari livelli di complessità.

Lo scopo di questa tesi è quello di fornire strumenti statistici utili per le analisi delle performance sportive, con particolare riferimento all'utilizzo dei modelli state space e all'analisi di serie storiche. La tesi è composta da quattro capitoli: i primi introducono in maniera complessiva gli argomenti trattati; i rimanenti, invece, presentano i principali contributi di questo lavoro. In particolare, il primo capitolo offre una visione generale delle analisi delle performance sportive, ne discute gli obiettivi e gli strumenti utilizzati, e delinea alcune opportunità di ricerca in campo statistico. Il secondo capitolo, invece, presenta una selezione di strumenti e modelli per le analisi di serie storiche. Nel terzo capitolo viene presentato un modello di clustering Bayesiano utile per descrivere le migliori performance annuali di atleti mezzofondisti italiani. Più nel dettaglio, il capitolo propone un modello state space matriciale in cui varie traiettorie multivariate di diversi atleti vengono raggruppate sulla base del trend delle performance e dei pattern di dati mancanti osservati nel campione, come indici della storia e delle attitudini personali degli atleti. L'inferenza è condotta mediante un algoritmo di simulazione nella classe dei metodi Markov Chain Monte Carlo. L'applicazione con dati reali mostra benefici e limitazioni dell'approccio proposto, fornendo indicazioni di quali siano i fattori rilevanti per ottenere performance sportive migliori. Il quarto capitolo descrive un modello per il monitoraggio dello stato di salute durante l'attività sportiva. Il modello proposto unisce la modellazione state space con i modelli per l'identificazione di changepoint al fine di individuare cambi distribuzionali in una sequenza di attività sportive. L'inferenza avviene tramite un algoritmo online di Expectation-Maximization che richiede un'approssimazione delle probabilità di changepoint predette, ottenuta tramite un metodo di approssimazione Monte Carlo sequenziale. Come conseguenza delle assunzioni fatte sul modello, l'algoritmo proposto processa sequenze di serie storiche in un contesto doppiamente online. Mentre i modelli di changepoint identificano cambi tra diverse attività successive, la formulazione state space del modello, unita all'algoritmo proposto, fornisce il beneficio aggiuntivo di stimare la probabilità di changepoint in tempo reale.

Dedicato a Faouzi perché il mio dottorato è iniziato in campo.

Acknowledgements

I would like to thank all those who allowed and pushed me to do this journey of PhD, and those who accompanied me in these recent years.

The first thanks goes to those who have worked with me. To Mauro, for believing in me since my master degree and for allowing me to study what I am most passionate about. To Petros, for our last year of meetings on Zoom, that, even if virtual, have been highly formative for me. I hope to be able to come to Greece soon.

I have to thank all my cycle's mates, including Fabio, for these years together, also for the various breaks at the food machines, appetizers and dinners. And Brasilena, of course! I also thank the guys from the other cycles, from the 32nd cycle onwards, for the exchange of advices, shared lunches and chats. In particular, I thank Cristian and Nicolas for the statistical lunches at La Roccia, seasosend with so many dilemmas.

Other important thanks go to all my athletics teammates and Faouzi. I miss the field and those who lived and trained with me for so many years. Leo, Jacopo, Phil, Marco F., Marco B., Riccardino, Fava, Tobia, Simone, Francesco, Francesca, Ele, Condo, Vale, and all the remaining, young and old, that shared all the moments of trainings with. This thesis is also dedicated to you. I also thank AlePhoks, Ricky Gara, and who has spent good times with me over the last years. I thank also my family and my grandmothers.

Finally, thank you Giulia for these last two years together. Thank you for being there for me every step of the way, for the trips to Puglia to meet Carla, for puntatine, for journeys around Padova with your Panda, and for helping so much in all aspects of my life! You actively partecipated throughout the process that led to the writing of this work. This thesis is dedicated to you, too!

Contents

List of Figures

List of Tables

Introduction

Overview

Sports performance analysis is a topic with a long history in sports science, but has received a further boost in recent years thanks to the availability of new technologies capable of collecting a large amount of data and to the increasing interests of other fields, such as statistics, mathematics, and computer science. This branch of research is involved in several aspects of the sports sector. On one side, the interest of the athletes and their equippes is to have information as accurate as possible on the activities carried out, on the other side, the interest of the companies that develop these new technologies is providing new analytical tools —hardware and software— which try to meet the needs of those who perform the sport professionally, but also try to involve amateurs, fans, and people that have an economic return from this sector. Sources of data are multiple: not only data regarding the competitions of athletes are systematically collected by the official federations, but, increasingly, athletes are monitored by smart devices, cameras, and other tools that allow to collect and share data regarding their activities, with the aim, not only to track progress, but also as a valid alternative to expensive and inaccessible laboratory tests. However, there is a dark side to the high availability and easy accessibility: data are typically collected in an unstructured manner, without a declared sampling design, and their volume and complexity require tools that are not accessible to most. The role of the statistician is clear: to extract knowledge from the enormous amount of available data using appropriate tools, which are able to respond to the questions and needs of athletes —and those interested in sports— without neglecting the intrinsic complexity of the phenomena under consideration. Among the many tools that can be used in this context, in this research project we explore the use of state space models, as time dimension is a central aspect in sports data: athletes are monitored during the time of an activity, from day to day, month to month, and so on. State space models are indeed a broad class of models used for time series analysis, that include ARMA models, linear regression models, and structural models involving trend, periodicity, and cycle components (Durbin and Koopman, 2012). In addition to the natural advantage of being time series models, they allow for the construction of complex models where multiple sources of variation are present, including missing values, together with a unified methodology for offline and online inference through standard filtering routines and sequential Monte Carlo approaches (Kantas *et al.*, 2015). We use these models for two distinct purposes: to identify athletes' careers through a longitudinal study and to monitor health status during sports activities using wearable and smart devices.

On the first side, the identification of athletes careers is an important aspect in performance evaluation, as it allows coaches to plan the future of young athletes with consistent goals, based on personal attitudes and experiences, to avoid overloads and early drop-outs. Good planning, along with support during the period of injury, have been identified among the relevant factors in avoiding drop-out in young athletes (Bussmann, 1999; Larsen and Alfermann, 2017). Despite the large availability of data, made possible by federations systematically collecting athletes' competitions in public repositories, identifying possible careers in middle distance athletes is a difficult task: drop-out, as the main source of missing observations, is potentially correlated to performance (Enoksen, 2011). Moreover, drop-out is not the only source of missing data that correlates with athletic performance: among these, we include injuries, late entry into competitions, and personal attitude, which implies a different propensity for the type of race performed (Sandford and Stellingwerff, 2019). We address career identification as a clustering problem, where clusters, that identify existing careers, depend on athletes' performances over the years, but also on the patterns of presence or absence of observations, as well as the drop-in and drop-out phenomena.

On the second side, the use of wearable technologies and sensor data for medical problems is gaining increasing interest from the statistical community (Huang *et al.*, 2019; de Chaumaray *et al.*, 2020; Qian *et al.*, 2020). Apps and wearables are indeed driving the next digital health and fitness revolution, in which intelligent and automatic real-time control and monitoring tools will become extremely relevant (Statista, 2020a). It is expected that in the near future, 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). We focus on identifying variations in the behavior of one or more measurements, i.e. changepoints, caused by changes in physical condition such as physical discomfort, periods of prolonged de-training, or even the malfunction of measuring devices during running activities. 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, together with 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 using a state space model that considers the data as a sequence of activities, where each activity is a multivariate time series representing a part of the training session. Changepoints are identified in a doubly-online framework: in the between-online setting activities are processed sequentially, and changepoint identification occurs when a new one is fully observed; in the within-online setting, data are processed while activities are performed. During a run, information on the behavior difference between the current and previous activities is translated into motivational feedback or a potential alert before the end of the activity.

Main contributions of the thesis

The purpose of this research project is to develop innovative methods and models for the analysis of data collected in sports contexts, with particular reference to individual sports and athletics. The thesis consists of four chapters. The first chapter aims to introduce the reader to performance analysis by discussing some existing contributions in the literature, briefly presenting some datasets (including a new one), and outlining some research opportunities. The second chapter introduces state space models for the analysis of vector and matrix time series and the main tools utilized in later chapters, along with some alternative models present in the literature. The main contributions of the thesis can be found in the third and fourth chapters. In particular:

Contribution 1: Time series clustering of athletes' careers under informative missing data patterns

In the third chapter, we propose a model-based clustering approach with the aim of identifying the careers of Italian middle-distance runners, born in 1988, represented by multivariate time series of their performance in the 800, 1500, 5000 meters races in the period 2006–2019. We address the clustering problem through a hierarchical specification. First, we define a matrix state space model whose purpose is to describe the time evolution of the observed races and at the same time to capture the cross-sectional dependence present among the variables. At this stage of the model specification, clustering is achieved through a selection matrix involved in the measurement equation, the purpose of which is to associate the different athletes with the states that describe the group dynamics over the years. Second, we consider the presence and absence of variables as an informative aspect for clustering. We include this information through two variables: the first one describes the drop-in and drop-out phenomena in the sample; the second one describes the actual participation in the competitions by the athletes. In this case, clustering is determined both by the drop-in and drop-out probabilities, governed by a Markov chain with conditionally constant probabilities, but also by the different probabilities of the athletes participating in different races. Inference is obtained through a Gibbs sampling algorithm, which is easily derived through the use of conjugate prior distributions. The identification of the number of groups is discussed, together with prior specification. Application with real data shows benefits and limitations of the model, which are discussed along with other possible developments.

Contribution 2: Doubly-online changepoint detection for monitoring health status during sports activities

In the fourth chapter, 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 de-training, or even the malfunction of measuring devices. Our framework considers data as a sequence of running activities represented by a 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.

Chapter 1

Sports performance analysis

1.1 Today's sports performance analysis

For as long as sports have existed, athletes and coaches have been asking themselves: *How can we improve*? This question, which goes beyond the simple competitive aspect, is a key aspect in sports science, so much so that, over the years, coaches and experts have collected material and experiences to answer this question and disseminate the information gathered through years of observation and practice in performance analysis. In this thesis, the terms *sports performance analysis* refer to the analysis of data that have been collected in sports contexts and are related the activities carried out by athletes, without any particular reference to the type of data, the specific research question, or the methodologies used. The data typically collected by athletes include: training diaries, training schedules, videos and data related to training and competitions, and results of tests and scientific protocols carried out in the fields and laboratories (see, among others, Sargent *et al.*, 2014; Müller and Glad, 2014; Coh *et al.*, 2019; Alvero-Cruz *et al.*, 2020). While interest in analyzing these data has always existed among athletes, coaches, and federations, with the advent of new technologies, other fields of research have also become extremely interested in this world, not only because of the associated challenging scientific problems but also out of pure economic interest. Demonstrations of these increasing interests are the recent contributions and books in the fields of mathematics (Karlis *et al.*, 2021), engineering (Allen and Goff, 2018), computer science (Tuyls *et al.*, 2021; Richter *et al.*, 2021; Baca, 2014), and statistics (Santos-Fernandez *et al.*, 2019; Severini, 2020) as well as those by companies in the private sector, such as Xsens, Polar and PKvitality, among others (Paulich *et al.*, 2018; Emig and Peltonen, 2020). This increased interest is supported by a growing availability of data. While athletes are monitored using the classic methods mentioned above, they are also tracked using technologically advanced and intelligent devices, which are becoming increasingly more present in their lives. Athletes are monitored using non-invasive, or minimally invasive, GPS tracking devices, heart rate monitors, accelerometers, gyroscopes, wearable sleep and lactate monitoring devices, and so on (Cardinale and Varley, 2017; Villena Gonzales *et al.*, 2019; Bourdon *et al.*, 2017).

Although it is possible to collect data and extract increasingly precise information on athletes and, on the basis of the same, improve knowledge, training and racing choices, there are still open questions discussed in the literature regarding the limitations and problems with such data, and doubts with respect to their reliability (Halson, 2014; Bourdon *et al.*, 2017; Vermeulen and Venkata, 2018). We have to make a distinction between ethical and practical problems of this topic. From the ethical point of view, Halson *et al.* (2016) highlight the primary need to not harm athletes and to evaluate the implications of the choices made on the basis of poor scientific evidence for their safety and health, which includes their psychological wellbeing (too much information can raise stress levels). We add to this the need to guarantee the confidentiality of data, not only to preserve the privacy of athletes but also avoid offering, on the basis of data, unfair competitive advantages. From the practical perspective, on the contrary, we mention the difficulties related to both the collection and the use of these data. In the first place, data are typically collected in an unstructured manner, without a declared sampling design, and their volume and complexity require tools that are not accessible to most. The collected data are characterized by different formats, which vary based on the different companies that provide tools their collection (Mackie, 2016; Frick and Kosmidis, 2017) and also on the types of data collected (video, GPS route, IMU based motion tracking, etc.). In addition, nowadays there is no one specific method that is used to collect data, with athletes using different strategies to do the same, depending on their needs, uses, and devices, which creates general difficulties with regard to aggregating them. In conclusion, we highlight an important aspect in sports: athletes interact with a highly dynamic environment (Pol *et al.*, 2020) composed of coaches, teammates, competitors as well as training and technological devices. Adequate performance analysis tools can manage such complexity, and statistics can highly improve research in this field.

1.2 Selected works in statistics

1.2.1 The trackeR package

GPS-enabled tracking devices and heart rate monitors are widely used in several individual sports, such as running, swimming, and cycling, and during athletic training in other field and team sports. The `trackeR` package by Frick and Kosmidis (2017) aims to fill the gap between their data collection and their analysis using the `R` statistical software (R Core Team, 2020). The package offers several routines for both basic and advance retrospective statistical analysis. These include importing utilities for files in different formats (`.tcx`, `.db3`, and Golden Cheetah's `.JSON` files); handling units of measurements; session-specific summary statistics (time, distance covered, duration, average speed, average heart rate, etc.); missing values correction and distance correction using altitude; visualization tools (both for single and multiple sessions); work capacity quantification (Skiba *et al.*, 2015) and distribution and concentration profiles and smoothing (Kosmidis and Passfield, 2015).

We concentrate on the *distribution profile*, which is defined by Kosmidis and Passfield (2015) as the curve $\{v, \Pi(v) | v \geq 0\}$ such that

$$\Pi(v) = \int_0^T \mathrm{I}(v(t) > v)\mathrm{d}t,$$

where T is the total duration of the training session. The distribution profile is a monotone decreasing function that describes the time spent exercising above a threshold for a given variable V. The negative derivative of the distribution profile is defined by Kosmidis and Passfield (2015) as the *concentration profile*, which is useful for determining the concentrations of time around certain values of the variable under consideration. The use of the described two functions is motivated in a predictive framework, in which they are used in place of the standard summary statistics, obtained after preliminary smoothing operations. Among the advantages of using these functions, there is the possibility of comparing through a simple expedient (integrating out the time), many different sessions using a single time-free domain, even when the sessions have different durations. However, the distribution of the energies over the duration of the session is, in some cases, important in performance analysis, and it is possible to obtain artificial examples in which two different types of effort have similar distribution profiles.

1.2.2 Monitoring ultrarunners' behavior in a 24-hours race

24-hour races are complex competitions and are different from the typical regular races held in athletics. In these races, the athletes follow a predefined route, and victory is awarded to those who, within the same day, manage to cover the greatest distance. A 24-hour race requires a strategy that is personal and allows athletes to finish the race to the best of their ability. During the race, the athletes are allowed not only to decide how fast they should run but also to slow down or stop when needed in order to finish the race in the best possible way. Significant work on pacing strategies has been done by Abbiss and Laursen (2008), who propose a review of the various strategies that are employed during competitions. In the same vein, Bartolucci and Murphy (2015) propose a finite mixture of linear and multinomial logit regressions models to cluster the observed speed trajectories of athletes over the number of laps completed in a 24-hour race. In particular, for athlete q and lap l, they consider the following dependence structure:

$$p_{\boldsymbol{\theta}}(S_q = g|\mathbf{z}_q)p_{\boldsymbol{\theta}}(b_{q,l}|S_q = g, \mathbf{x}_l)p_{\boldsymbol{\theta}}(y_{q,l}|S_q = g, b_{q,l}, \mathbf{x}_l).$$

Here, $\mathbf{x}_l$ and $\mathbf{z}_q$ are vectors of covariates, $\boldsymbol{\theta} \in \mathbb{R}^d$ is a vector with unknown parameters, $S_q \in \{1, \dots, G\}$ is the latent cluster allocation, $b_{q,l} \in \{0, 1, 2\}$ is the observed variable that denotes if the athlete is running, walking (or resting), or has left the race during lap l (before the end), respectively, and $y_{q,l}$ is the observed speed during lap l, for $l = 1, \dots, L_q$, and $q = 1, \dots, Q$. In the equation, $p_{\boldsymbol{\theta}}(S_q = g|\mathbf{z}_q)$ and $p_{\boldsymbol{\theta}}(b_{q,l}|S_q = g, \mathbf{x}_l)$ are generalized Bernoulli densities with probabilities parametrized on the basis of multinomial logits. If $b_{q,l} = 0$, the density $p_{\boldsymbol{\theta}}(y_{q,l}|S_q = g, b_{q,l}, \mathbf{x}_l)$ is the density of a normal distribution with a cluster-dependent mean and variance shared across individuals; if $b_{q,l} = 1$, this density is left unspecified (athlete is resting), or it is degenerate in 0 if $b_{q,l} = 2$ (athlete completed the race). The athletes and the laps completed by them are assumed to be conditionally independent, which allows to obtain a likelihood that is easily maximized using the EM algorithm. A desirable property of the model is that it allows to obtain clusters of trajectories, not only based on the speed observed during various laps, but also based on the different states of the athletes, which provide information regarding the phenomenon of drop-out, where the athletes leave the competition before it ends. For more details on both the model and the application, we refer to the original article by Bartolucci and Murphy (2015).

1.2.3 Other papers

In addition to the works explained in the previous sections, several contributions in statistics that deal with problems in sports performance analysis have been proposed in recent years. Other than the work of Kosmidis and Passfield (2015) and Jacques and Samardžić (2021), which are briefly discussed in Section 1.3.2, we mention the works by Haynes *et al.* (2017), Pradier *et al.* (2016), Leroy *et al.* (2018), and Dolmeta *et al.* (2021). More specifically, Haynes *et al.* (2017) develop a computationally efficient nonparametric approach for changepoint detection, and apply their method to detect changes in heart rate during one physical activity. Once they have segmented the activity using the changepoints identified analyzing the behavior of the heart rate variable, they have compared the segmentation obtained with the variables speed and altitude, in order to visually understand how these changes could be driven by changes in these other two variables. By using a Bayesian non-parametric approach, Pradier *et al.* (2016) study the impact of age, gender and environment on the runners' performances in marathons with two different aims. First, they derive a grading method that allows for direct comparison of runners regardless of their age and gender, based not only on top world records, but also on the performances of all runners. Second, they analyze the running patterns of the marathoners in time, obtaining information that can be useful for training purposes and for predicting the finishing time, applying their methods with data of different marathons. Leroy *et al.* (2018) solve the problem of early detection of promising young swimmers by clustering longitudinal data describing their race performances over the years, using a functional data analysis approach. Finally, Dolmeta *et al.* (2021) develop a hierarchical Bayesian GARCH model for functional observations, useful for describing and predicting the evolution of performances of shot put athletes over time. Their model provides an accurate description of athletes' performance trajectories over the years, by considering both the intra- and inter-seasonal variability present in the measurements.

1.3 Datasets

While a growing number of statistical tools are available for the analysis of sports data, it is relevant to determine where and how one should collect data that can be used for the analysis and development of new techniques in the context of sports performance analysis. Data are collected on the field during training and competitions by athletes and coaches, official federations, sports science research teams, people interested in the topic, private or public institutes as well as companies that produce and sell the devices and services used for athletic monitoring (e.g. Strava, PolarFlow, StatsPerform, etc.). In this section, we review some datasets and data sources that are freely available for research.

1.3.1 Official data sources

One of the main resources of data for the performance analysis of various athletes are official federations. In athletics, for example, different data are collected at the regional, national, and international levels by the respective federations. Table 1.1 lists the official websites of athletic federations from Veneto, Italy, Europe, the World, and some other data providers. On the national site, for instance, it is possible to access athletes' results

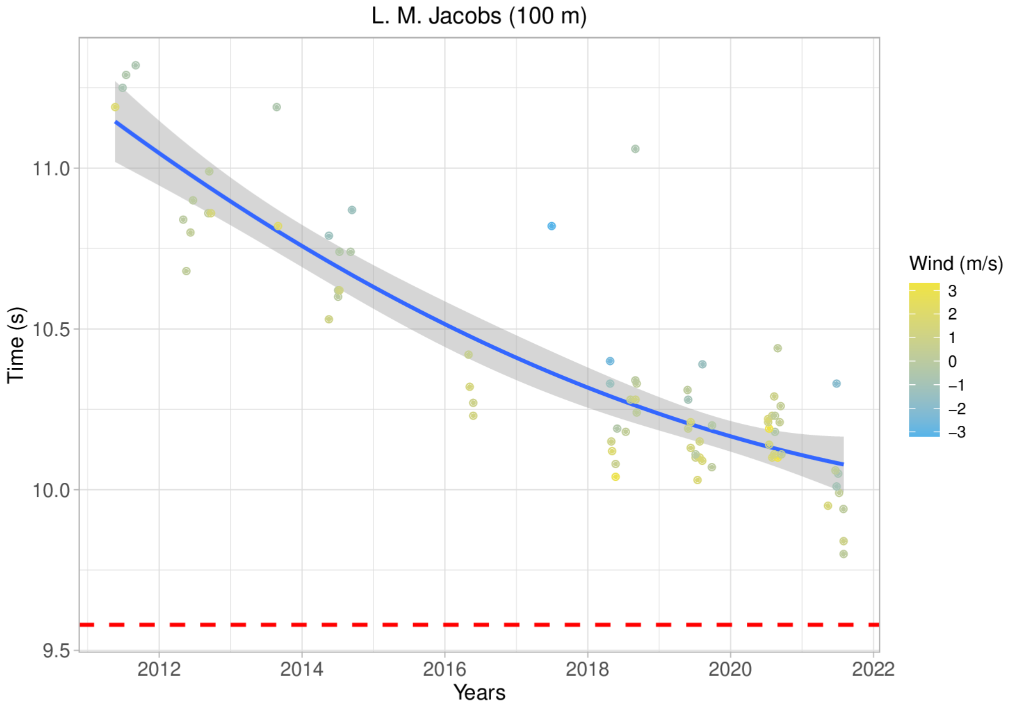

Figure 1.1: Performances in the 100 meters by L. M. Jacobs. The red dashed line represents the world record held by U. Bolt (2009). The solid blue line indicates the regression line with linear and quadratic terms. The gray band is the `ggplot2`'s default confidence band (Wickham, 2016).

over time and certain personal information (sex, age, team membership history, etc.) or important news related to their career as well as obtain information regarding races and competitions. The Italian national rankings have been updated annually since 2005 and, therefore, allow to track the evolution of the best annual performances of each athlete over time. Through web scraping, it is possible to obtain useful data for performance analysis both for single athletes and for comparisons between multiple athletes over time (see Chapter 3).

Figure 1.1 shows an example of a graphical visualization based on the data collected

Table 1.1: Example of data sources and useful links in athletics.

Name	Level (role)	Links and description
Fidal Veneto	Regional (federation)	Material and experiences
Fidal	National (federation)	Results, rankings, and material
European Athletics	International (federation)	Results and rankings
World Athletics	International (federation)	Results and rankings
TDS	National (private company)	Road and cross country races
New York Marathon	International (race)	Results

by official federations for a single athlete. More specifically, the figure shows the performance progression of L. M. Jacobs, the 2021 Tokyo Olympic champion in the 100 meters race. His performances are plotted over time and colored according to the wind speed recorded during competitions, where positive wind speed means that the wind was blowing from behind the athletes. Since parabolic shapes are typically used to describe an athlete's career (see, e.g., Haugen *et al.*, 2018), the regression line was added using a linear regression model with both linear and quadratic terms. In the figure, we can grasp the possible presence of a relationship between wind and performance, with a headwind typically slowing the athlete down. However, other relevant information is present in the official databases and not used in this analysis, such as the lane of the races, the type of races (final, semifinal, heat, etc.), the place and time of the races as well as the knowledge of the experts (rules for regular races, training period, athlete's history, etc.) and other information (injuries, coach, etc.). Although these analyses are of interest in the field of sports science, further in-depth analysis of Jacobs' career is beyond the scope of this section, the purpose of which is to simply show and discuss some typical data and data sources in sports performance analysis.

1.3.2 Other datasets

Other data sources come from other scientific research in the field if researchers make the datasets available. In statistics, `R` packages are often accompanied by datasets. The dataset `runs` in the work of Frick and Kosmidis (2017) is composed of 27 training sessions of a single male runner during June 2013, with distances ranging between 2.79 km and 22.35 km. The relevant variables included in the dataset are latitude, longitude, altitude, heart rate, running cadence, distance, and speed and are similar in characteristics to those presented in the next section. Kosmidis and Passfield (2015) develop a multiplicative effects model for the identification of the factors that influence the performance of highly-trained endurance runners and analyze a one-year dataset composed of the training, laboratory, and field tests carried our by 14 competitive endurance male athletes (3469 distinct training sessions). Jacques and Samardžić (2021) develop an ordinal logistic regression model with functional covariates to analyze 216 one-hour bike sessions recorded by an amateur cyclist during 2019. The relevant variables included in the study are power, speed, heart rate, altitude, slope, cycling cadence, and temperature. The package by Mackie (2016) is composed of only a few examples of cycling data. Similar datasets were collected by Rauter and Fister (2015) and have been updated the years following the publication of their work. Their latest version is composed of 15 cyclists (14 male, 1 female), which include 4 cyclists who compete at a professional level (Fister *et al.*, 2017). The collection of these datasets typically occurs through direct contact or through requests from researchers to the users of platforms and social networks in the field, such as Strava, Garmin Connect, or Polar Flow.

Puchowicz *et al.* (2018) discuss the use of tracking technologies and portable sensors for doping detection and identify the critical power (CP) model (Hill, 1993) as a potential tool for doing this for cyclists. Their proposal uses data collected by one professional cyclist for a period of 6 years (Pinot and Grappe, 2015). However, their approach is based on "an added simulated doping effect for selected years", and no real observations are available for athletes who doped. Alternative anti-doping methods rely on monitoring athletes' hematological measurements, such as the use the biological passport (see, e.g., Sottas *et al.*, 2010; Schütz and Zollinger, 2018; Faiss *et al.*, 2020); but there is no interest in relating such blood measurements to athletes' performances.

An additional source of data comes from the companies that are directly involved in the field. With the goals of creating new hardware and improving user experience, Strava makes their dataset, which is made up of an ever-growing number of user activities, available to developers. In April 2021, Xsens organized the Xsens Biomechanics Challenge, in which teams of engineers, sports scientists, and computer scientists from around the world challenged each other to solve a case study that is related to the daily life of a physiotherapist. In the competition, partecipants were required to calculate the power expressed during 4 different repetitive athletic actions (squat, bench press, ball kick and ball throw) and evaluate the presence of asymmetries, as well as the variability and repeatability of the exercises. In brief, data that described the movements of the body (angles, speeds, etc.) during the performance of the exercises were collected by an Xsens suit that consisted of 17 wireless sensors placed all over the body. We direct readers to the company's website for an overview of their products, as well as to specialized literature for an overview of their uses in biomechanics (see, e.g., Rana and Mittal, 2021; Camomilla *et al.*, 2018).

1.3.3 A new dataset with sports activities

The dataset collected for this thesis is currently composed of 5875 sports activities but is constantly updated with new ones. In total, 24 subjects are currently included in this dataset, who practice running and athletics at both amateur and competitive levels (the highest level in Italy is national). Of these, 21 athletes are male, and 3 are female. Data collection was conducted for scientific purposes, by a social contact after a request was received by the author through his Strava account, and started on September, 2017. However, not all the participants consented to the publication of sensitive data

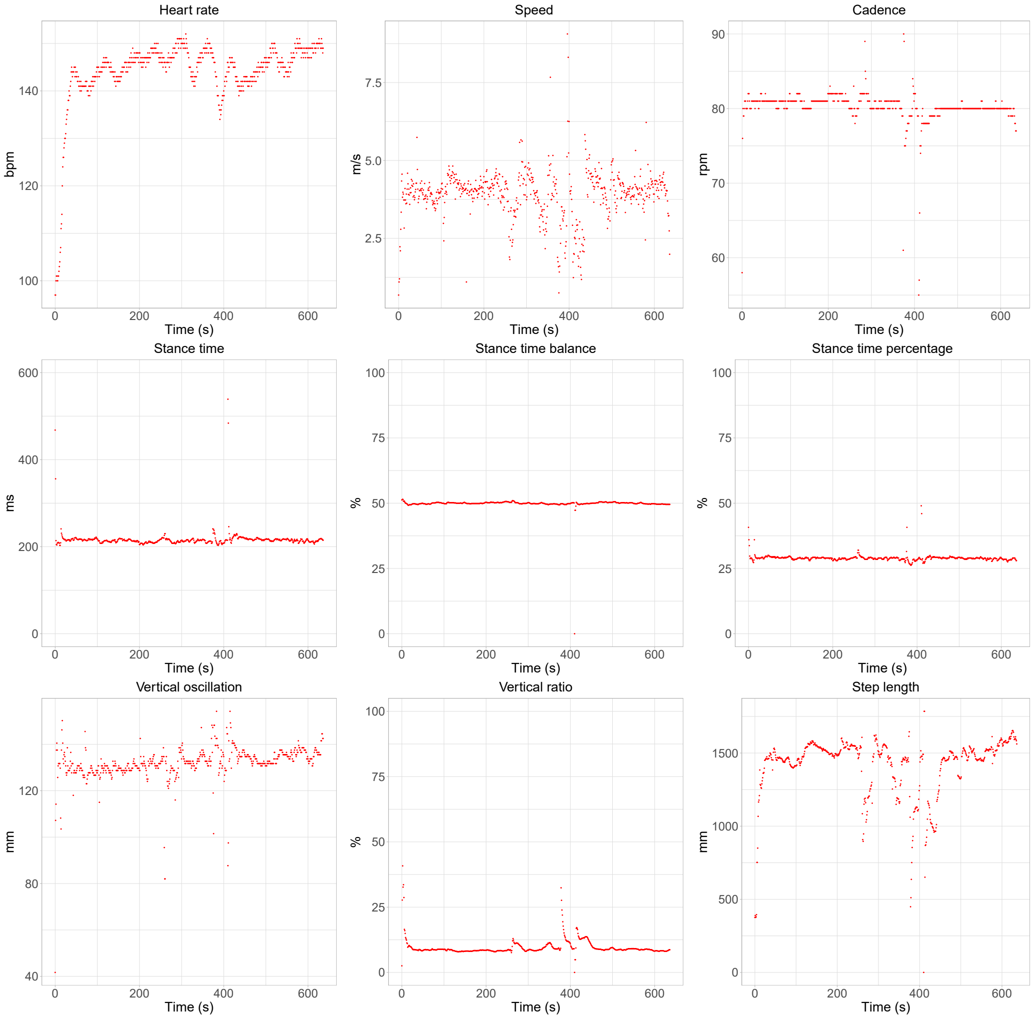

Figure 1.2: Example of one activity carried out by an athlete and some of its measured variables. Speed is obtained as the first difference in the cumulative distance, since observations for the latter were recored every second. The activity represents approximately 10 minutes of continuous running. The variables latitude, longitude, and altitude have been excluded.

in a disaggregated form, which can trace back to their habits and personal details. The minimum number of activities per subject is 1, while the maximum is 1866, and these refer to more than 1167 hours and 14100 km of activities. The activities include running, swimming, biking, walking, and hiking. The devices used are from various brands and include Garmin, Polar, and Suunto. The activities were downloaded in various formats, including `.gpx`, `.tcx`, `.json`, and `.fit`. The routine `FitCSVTool.jar` available on Garmin's site can be used to convert `.fit` files to `.csv`. The most common variables found are UTC time during the activity, latitude, longitude, altitude, heart rate, cadence, speed, cumulative distance, and temperature. More recent files have measurements related to vertical oscillation, vertical oscillation ratio, stance contact time, stance time percentage, and balance during stance contact.

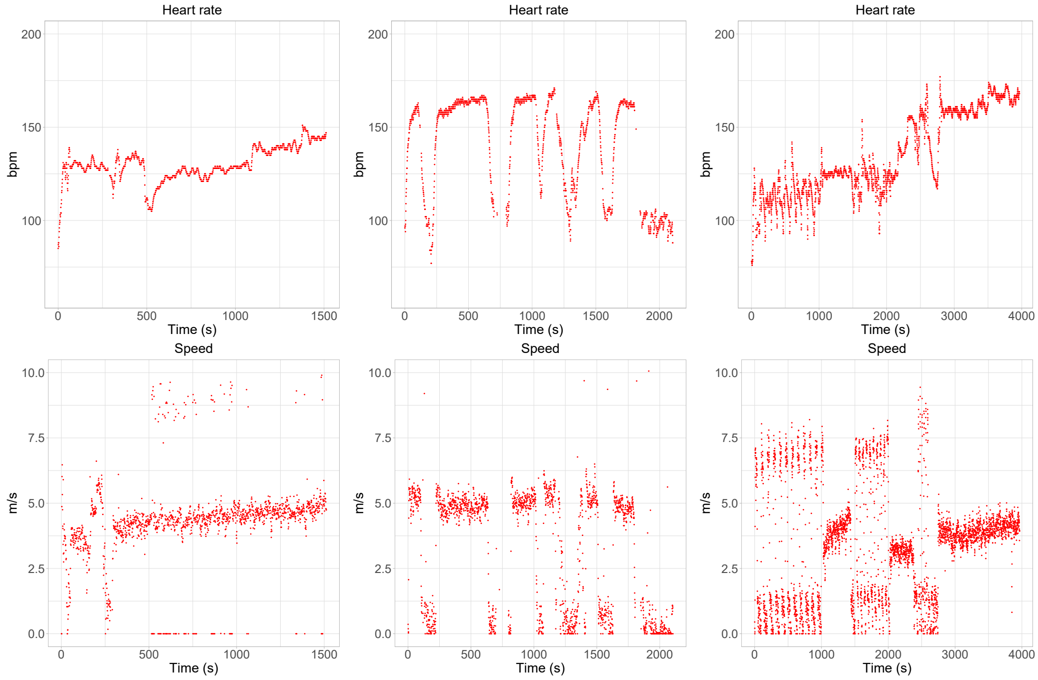

FIGURE 1.3: Example of one athlete's activities that vary in intensity, effort distribution, and duration. Different activities are placed in different columns.

Figure 1.2 shows an example of an approximately 10-minute continuous running activity performed by an athlete, for which the location variables (latitude, longitude and altitude) were excluded. In each graph, the x-axis represents time (in seconds), while the y-axis represents the variable measured in the original unit of measurement in the file. As we can see, the variables are characterized by different behaviors: some variables have an approximately constant behavior (stance time, stance time balance, etc.), while others change in level (speed, cadence), tend to grow (vertical oscillation and heart rate), or are characterized by moments during which there is greater variability in the measurements (speed). The possible presence of outlying observations can also be noted. A clear example of this is observation with value 0 around 400 seconds for the stance time balance variable. The same anomalous behavior is also present to a more or less severe degree in the other variables.

However, the activity shown in Figure 1.2 is only one of many performed by a single athlete. They vary in intensity, effort distribution (there are not only continuous running activities), duration, and mode of measurement (the measured variables change if different devices are used, and the interest in monitoring activities may vary based on the type of activity being performed). To illustrate how the different activities of a single athlete may differ from each other, Figure 1.3 shows the heart rate and the speed for 3 distinct activities performed by the same athlete, in which the behavior of the variables is irregular. Irregularity in observations is a typical feature in certain types of training, in which athletes perform activities that alternate between high-intensity and low-intensity phases. Such irregularities sometimes follow predefined rules (e.g. the runner has to alternate 30s at high intensity and 30s at low intensity); other times, they occur randomly. Further, at times workouts can combine some of these components (Kenneally *et al.*, 2018). A comprehensive view of the entire dataset would require significant effort and is outside the scope of this thesis. However, a subset of the collected activities is used in Chapter 4.

1.4 Research opportunities

Given the increased availability of data collected in a sporting context and the fact that they vary in regard to form, format, and the way they are collected and used, it is worth asking what is important in the research in this field and which questions need to be answered. To address such questions, it is necessary to consider three aspects together: (a) the needs of those working in the field and the current research questions; (b) the tools that are available and the goals of the companies involved in the field; (c) the tools that are available in statistics (and mathematics, engineering, and computer science) that can help answer these questions. One way to partially respond these needs is to observe the current trends in scientific research, by either considering the reviews and perspectives reported in the literature or examining the topics that are proposed and most discussed at conferences. A complete view of the topic, however, is difficult to obtain, especially for the aspects related to companies and sports people, who rarely make their goals and questions explicit to the public.

From the sports science point of view, Bourdon *et al.* (2017) provide an interesting “consensus statement”, in which actual (up to 2017; hence, not even actual) training load monitoring was discussed by experts of the field. They first distinguish between *internal* and *external* load monitoring (Bourdon *et al.*, 2017; Cardinale and Varley, 2017). Internal monitoring is related to the physiological and psychological response to training: the typical variables monitored are heart rate, blood lactate, oxygen consumption, and rate of perceived exertion (Bourdon *et al.*, 2017). External monitoring relates to the work performed by the athlete and is described by physical variables such as speed, running cadence, developed power, etc. Among the challenges, Bourdon *et al.* (2017) highlight the need for valid and reliable criteria to evaluate athletes’ loads. In addition to the simpler methods that use the measured variables in their units of measurement, the use of composite and derived methods is intriguing (Bourdon *et al.*, 2017). This context justifies our work proposed in Chapter 4, where we use both internal and external load variables within the same framework, in a completely online fashion. Real-time monitoring is indeed one the challanges discussed by Bourdon *et al.* (2017) in the section titled “What is the future of athlete load monitoring?”. Another aspect to be studied which we include among future research opportunities, involves the definition of a link between external and internal load monitoring variables based on the data that are observed or between external and internal load monitoring variables and the prescribed trainings. This would lead to new insights into training prescriptions.

How to prescribe training is an important aspect in sports science (Kasper, 2019). Training prescription is based on the experience and knowledge of coaches, which are developed over the years both through studies in the field and through trial and error. By training prescription, however, we do not only refer to the actual decision regarding which training to execute and include, but also the evaluation of the athlete’s capabilities as well, based on their short-, medium-, and long-term goals. Official databases are growing in number and completeness. New visions can arise through the exploration of the collected material at any level and for any kind of sport. This is in line with Chapter 3, in which we try to extract different profiles of athletes on the basis of the races held over the years, in an attempt to outline the best strategies to follow based on our results.

Another aspect to consider is related to conferences and workshops. Table 1.2 summarizes the topics of the calls for papers at the IJCAI-AISA-2021 conference. Among the several conferences and workshops on the topic, this one was reported because its calls for papers was detailed in reporting several topics in the world of sports analytics (and sports performance analysis). Other conferences worth mentioning are the MIT SLOAN analytics conference, AUEB Sports Analytics Workshops, NESSIS, MathSport International, ISEA, and icSPORTS. The last two are better aligned with a more industrial and engineering vision of sports, rather than a purely statistical one, which allows an alternative perspective on the topic.

In the next chapter, the focus of our work will be on state space models (both vector- and matrix-variate). Compared to many other tools, state space models offer certain advantages: they are general and easily generalizable; they include many models that are used in time series analysis; they allow the use of a unified strategy for inference (offline and online); finally, they are interpretable. Among these reasons, we recognize the relevance of the time component: athletes are monitored during an activity, day after day, month after month, and year after year. A vector time series is observed if, for an athlete, multiple variables are monitored over time. Matrix time series, on the

Table 1.2: Topics included in the IJCAI-AISA-2021 conference as examples of research opportunities. The symbol $\star\star\star$ indicates the existence of a relationship with the problems addressed in this thesis.

Main topic	Sub-topic
Representation learning and aggregate statistics $\star\star\star$	Player- and team-level statistics, vectors, and/or learned embeddings for analysis of in-game situations
	Modeling and learning of player/team rankings, strengths, and weaknesses
	Data processing, filtering, and visualization techniques/demos
Evaluation of actions, trajectories and strategies, and learning of optimal policies $\star\star\star$	Value estimation during in-game situations (e.g., action values for actions and players)
	Detection and optimization of in-game tactics
	Reinforcement learning for sports analytics
Game-theoretic and multi-agent aspects	Predictive and prescriptive analysis of set pieces and in-game play
	Learning of coordination of player and team behaviors
	Transfer and imitation learning of human play
	Synergy or "chemistry" of groups of players
Physical and human factors $\star\star\star$	Physics-simulation of real play
	Human factors such as injury and fatigue predictions
Video-based modeling	Event detection and activity recognition
	Pose detection
	Generative modeling of video data

other hand, are found in various contexts regardless of whether the observations are naturally in the matrix form or the matrix form is a technical expedient for analysis. Multiple sensors measuring the same multiple variables can result in a matrix time series if the observations from each sensor are placed in different columns. Similarly, when considering multiple athletes (columns) participating in the same races (rows), placing athletes in columns allows for a compact notation that can facilitate the analysis. Chapter 2 provides a selected overview of this topic. In Chapters 3 and 4,more specific contributions related to performance analysis have been presented.

Chapter 2

Selected results in state space modeling

The purpose of this chapter is to introduce some of the tools, which are mostly known in the literature, that are used for analyzing multivariate time series using state space methods and that have been used in subsequent chapters. Notations introduced follow the classical treatment employed by Durbin and Koopman (2012), which is mainly likelihood-oriented. It is important to note, however, that some of these tools are also used in Bayesian analysis of dynamic linear models, in which our main reference for their treatment is the book by West and Harrison (1997). Here, we introduce the concept of the matrix state space and review some of the main proposals for the analysis of matrix-variate time series. Although this chapter does not provide innovative contributions in the field of state space modeling, it is useful to present certain key elements that can be used as building blocks for more advanced techniques in sports performance analysis.

2.1 Vector state space model

In general, the term *state space models* refers to a general class of models that allows an unified treatment of a wide range of problems that occur in time series analysis. From hereafter, the set $\mathcal{Y}_{1:T} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$ denotes the time series of T ordered observations, where T is the total duration of the time series, and $\mathbf{y}_t \in \mathbb{R}^L$ corresponds to the L–dimensional vector of observations at time t, for $t = 1, \ldots, T$. The linear Gaussian state space model defined by Durbin and Koopman (2012) assumes that the observations over time are driven by the model equations

$$\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathrm{N}_L(\mathbf{0}, \mathbf{H}_t), \tag{2.1}$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{R}_t\boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \mathrm{N}_M(\mathbf{0}, \mathbf{Q}_t), \tag{2.2}$$

with $\boldsymbol{\alpha}_1 \sim N_K(\widehat{\boldsymbol{\alpha}}_{1|0}, \mathbf{P}_{1|0})$. In the above specification, the first equation is called *measurement equation*, that connects the observation vector $\mathbf{y}_t$ to the vector of the latent states $\boldsymbol{\alpha}_t \in \mathbb{R}^K$. The second equation is called *state transition equation*, that represents a first-order autoregressive process, which determines the behavior of the latent states over time. The elements $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\xi}_t$ are assumed to be serially independent and independent of each other for all time points. At time t, the matrices $\mathbf{Z}_t$, $\mathbf{H}_t$, $\mathbf{T}_t$, $\mathbf{R}_t$, and $\mathbf{Q}_t$ are fixed and have a known structure, the dimensions of which are implicitly determined by the other elements involved in the equations. Further, these matrices eventually depend on an unknown finite dimensional parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$, where $d < \infty$. Thus, the measurement equation has the structure of a linear regression model where the coefficient vector $\boldsymbol{\alpha}_t$ varies over time, and the error terms $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T$ are independent of each other but are heteroskedastic, as they depend on the covariance matrices $\mathbf{H}_1, \ldots, \mathbf{H}_T$, respectively. The matrix $\mathbf{Z}_t$ plays the role of the matrix of covariates in the regression framework; it is a design matrix that depends on the specification of the model being considered and can also depend on any of the values in $\mathcal{Y}_{1:(t-1)} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}\}$. The state transition equation accounts for the temporal dependence between the different states and, by extension, between the observations recorded at different time points. The initial state $\boldsymbol{\alpha}_1$ is assumed to be independent of both $\boldsymbol{\xi}_t$ and $\boldsymbol{\varepsilon}_t$, for any $t = 1, \ldots, T$, and it depends on the mean $\widehat{\boldsymbol{\alpha}}_{1|0}$ and the covariance matrix $\mathbf{P}_{1|0}$. These elements are generally assumed to be known; however, if they are not, many alternative solutions to the issue have been proposed in the literature (see, e.g., Durbin and Koopman, 2012; Shumway and Stoffer, 2017). The matrix $\mathbf{T}_{t-1}$ is a square design matrix, called *state transition matrix*, that, like $\mathbf{Z}_t$, can depend on $\mathcal{Y}_{1:(t-1)} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}\}$ and is determined by the model being considered. Finally, the matrix $\mathbf{R}_t$ is a selection matrix with the role of selecting specific elements of $\boldsymbol{\xi}_t$. For most of the models, $K = M$, and $\mathbf{R}_t$ is the identity matrix, such that $\mathbf{R}_t\boldsymbol{\xi}_t$ is substituted with $\boldsymbol{\xi}_t$, which is an M-dimensional vector that follows a zero-mean Gaussian distribution with covariance $\mathbf{Q}_t$.

The independence of errors $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T$ and disturbances $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T$, and their Markovian structure, allows for an easy resolution of many of the major problems in time series analysis. At time t, given the past $\mathcal{Y}_{1:(t-1)}$, all elements involved in the model are assumed to be non-stochastic, except for states, errors, and disturbances. Various generalizations have been proposed in the literature, such as the assumption of normality being dropped or the use of nonlinear relationships between observations and states. These generalizations are not of interest in this chapter, and we refer to specialist books for further details (see, e.g., Tsay and Chen, 2018). However, in this form, the model is general enough to include within this class a wide range of models present in the literature, including ARMA, local level and local linear trend models, models with periodic components, linear regression, VAR and dynamic factor models (see, e.g., Durbin and Koopman, 2012; Shumway and Stoffer, 2017).

2.2 Matrix state space model

Let $\mathbf{Y}_t \in \mathbb{R}^{P\times Q}$ be the t-th matrix of observations of the matrix time series $\mathcal{Y}_{1:T} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_T\}$. Similar to the previously defined vector state space, the matrix state space model is composed of a measurement and a state transition equation

$$\mathbf{Y}_t = \sum_{j=1}^{J_1} \mathbf{Z}_{j,t}\mathbf{A}_t\mathbf{S}_{j,t}^\top + \mathbf{E}_t, \quad \mathbf{E}_t \sim \mathrm{MN}_{P\times Q}(\mathbf{0}, \mathbf{H}_t), \tag{2.3}$$

$$\mathbf{A}_{t+1} = \sum_{j=1}^{J_2} \mathbf{T}_{j,t}\mathbf{A}_t\mathbf{U}_{j,t}^\top + \sum_{j=1}^{J_3} \mathbf{R}_{j,t}\mathbf{\Xi}_t\mathbf{B}_{j,t}^\top, \quad \mathbf{\Xi}_t \sim \mathrm{MN}_{R\times S}(\mathbf{0}, \mathbf{Q}_t) \tag{2.4}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{F\times G}(\widehat{\mathbf{A}}_{1|0}, \mathbf{P}_{1|0})$ independent of both $\mathbf{E}_t$ and $\mathbf{\Xi}_t$, for any $t = 1, \dots, T$. The error $\mathbf{E}_t$ and the disturbance $\mathbf{\Xi}_t$ are assumed to be independent of each other and serially independent, with non-singular covariance matrices $\mathbf{H}_t$ and $\mathbf{Q}_t$, respectively. In the equations, notation $\mathbf{X} \sim \mathrm{MN}_{B\times C}(\mathbf{M}, \mathbf{\Sigma})$ means that the matrix $\mathbf{X}$, of dimensions $B \times C$, follows a matrix-variate normal distribution with mean $\mathbf{M}$ and covariance $\mathbf{\Sigma}$; $\mathbf{X}$ is defined here as the random variable such that $\mathrm{vec}(\mathbf{X})$, that is obtained by stacking its columns one underneath the other is $\mathrm{vec}(\mathbf{X}) \sim \mathrm{N}_{BC}(\mathrm{vec}(\mathbf{M}), \mathbf{\Sigma})$. For the sake of generality, note that, under this specification the matrix-variate normal distribution is not meant to adhere to the definition proposed by Gupta and Nagar (2000), in which the covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}^C \otimes \mathbf{\Sigma}^R$ is decomposed by a Kronecker product into a row $\mathbf{\Sigma}^R$ and a column $\mathbf{\Sigma}^C$ covariance matrices.

The model used in Equations (2.3) and (2.4) has been proposed in engineering literature, and particularly by Choukroun *et al.* (2006), who pointed out that any matrix-variate linear discrete-time stochastic system can be described by the state transition equation in Equation (2.4). This aspect is motivated by the fact that $\mathbf{A}_{t+1}$ can be written as the sum of J_2 elements, involving a left and right multiplication of $\mathbf{A}_t$ by the state transition matrices $\mathbf{T}_{j,t}$ and $\mathbf{U}_{j,t}^{\top}$, respectively, for $j = 1, \dots, J_2$, and an additive disturbance term, which is represented by $\sum_{j=1}^{J_3} \mathbf{R}_{j,t}\Xi_t\mathbf{B}_{j,t}^{\top}$. In this way, any scalar element of $\mathbf{A}_{t+1}$ can be written as a linear combination of $\mathbf{A}_t$ plus a scalar disturbance term (see, Choukroun *et al.*, 2006, for detailed proofs). Differently from Choukroun *et al.* (2006), however, the disturbance term $\sum_{j=1}^{J_3} \mathbf{R}_{j,t}\Xi_t\mathbf{B}_{j,t}^{\top}$ is a sum over J_3 elements, involving a left and a right multiplication of Ξ_t, by $\mathbf{R}_{j,t}$ and $\mathbf{B}_{j,t}^{\top}$, respectively, for $j = 1, \dots, J_3$. This representation is preferred to the one used by Choukroun *et al.* (2006) because it allows the matrix model in Equations (2.3) and (2.4) to be more easily linked to the vector model in Equations (2.1) and (2.2) and prevents the necessity of considering matrix-variate normal random variables with a singular covariance matrix (Gupta and Nagar, 2000). For what concern the measurement equation, the matrix of observations $\mathbf{Y}_t$ is linked to the matrix of the states $\mathbf{A}_t$ through a sum of J_1 terms, involving a left and right multiplication by $\mathbf{Z}_{j,t}$ and $\mathbf{S}_{j,t}^{\top}$, respectively, for any $j = 1, \dots, J_1$. In this way, any scalar element of $\mathbf{Y}_t$ is a linear combination of elements of $\mathbf{A}_t$ plus a scalar error term. At time t, the elements $\mathbf{Z}_{j,t}$, $\mathbf{S}_{j,t}$, $\mathbf{H}_t$, $\mathbf{T}_{j,t}$, $\mathbf{U}_{j,t}$, $\mathbf{R}_{j,t}$ $\mathbf{Q}_t$, and $\mathbf{B}_{j,t}$ are fixed and have known structures but may depend on an unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$, with $d < \infty$. Moreover, $\mathbf{Z}_{j,t}$ and $\mathbf{S}_{j,t}$ can depend on the past elements $\mathcal{Y}_{1:(t-1)}$, and $\mathbf{T}_{j,t}$ and $\mathbf{U}_{j,t}$ can depend on $\mathcal{Y}_{1:t}$. The matrices $\mathbf{R}_{j,t}$ and $\mathbf{B}_{j,t}$ are selection matrices used to select specific entries of Ξ_t. In general, the number of distinct elements involved in the terms in which a summation is present can be obtained by considering the product of the number of rows and the number of columns of the left and right matrices involved in the respective terms, such that $J_1 = PQFG$, $J_2 = (FG)^2$, and $J_3 = FRSG$. However, in many models, these numbers are smaller because they involve simple structures with fewer elements. Table 2.1 summarizes the dimensions of the elements involved in the model along with the respective dimensions of the model in the vectorized form, which will be discussed in the next subsection.

The model that was proposed trivially includes the vector model in Equations (2.1) and (2.2) as a special case, in which both the observations and the states are vectors. In addition, it includes, as special cases, some of the latest models proposed in the literature to analyze time series of matrices, thus providing a unified framework to treat them. Compared to any other model that considers a vectorized form of $\mathbf{Y}_t$, it permits the construction of interpretable models that preserve the original matrix form of the observations and, in some cases, a drastic reduction in the size d of the unknown parameter $\boldsymbol{\theta}$. Section 2.3 reviews the current proposals found in the literature for analyzing matrix-variate time series.

Table 2.1: Dimensions of the matrix state space model and its vectorized form.

Equation	Matrix form	Dimensions	Vectorized form	Dimensions
Measurement	$\mathbf{Y}_t$	$P \times Q$	$\boldsymbol{y}_t$	$PQ \times 1$
	$\mathbf{E}_t$	$P \times Q$	$\boldsymbol{\varepsilon}_t$	$PQ \times 1$
	$\mathbf{Z}_{j,t}$	$P \times F$	$\mathbf{S}_{j,t} \otimes \mathbf{Z}_{j,t}$	$PQ \times FG$
	$\mathbf{S}_{j,t}$	$Q \times G$	$\mathbf{H}_t$	$PQ \times PQ$
	$\mathbf{H}_t$	$PQ \times PQ$	$J_1 = PQFG$	1×1
State	$\mathbf{A}_t$	$F \times G$	$\boldsymbol{\alpha}_t$	$FG \times 1$
	$\boldsymbol{\Xi}_t$	$R \times S$	$\boldsymbol{\xi}_t$	$RS \times 1$
	$\mathbf{T}_{j,t}$	$F \times F$	$\mathbf{U}_{j,t} \otimes \mathbf{T}_{j,t}$	$FG \times FG$
	$\mathbf{U}_{j,t}$	$G \times G$	$\mathbf{B}_{j,t} \otimes \mathbf{R}_{j,t}$	$FG \times RS$
	$\mathbf{R}_{j,t}$	$F \times R$	$\mathbf{Q}_t$	$RS \times RS$
	$\mathbf{B}_{j,t}$	$G \times S$	$J_2 = (FG)^2$	1×1
	$\mathbf{Q}_t$	$RS \times RS$	$J_3 = FRGS$	1×1

2.2.1 Vectorized form

Let $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$, $\boldsymbol{\alpha}_t = \text{vec}(\mathbf{A}_t)$, $\boldsymbol{\varepsilon}_t = \text{vec}(\mathbf{E}_t)$, and $\boldsymbol{\xi}_t = \text{vec}(\Xi_t)$ be the vectors obtained by stacking the columns of $\mathbf{Y}_t$, $\mathbf{A}_t$, $\mathbf{E}_t$, and Ξ_t, respectively, one underneath the others. The matrix-variate state space model in Equations (2.3) and (2.4) has an alternative vectorized representation

$$\begin{aligned} \mathbf{y}_t &= \widetilde{\mathbf{Z}}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{N}_{PQ}(\mathbf{0}, \mathbf{H}_t), \\ \boldsymbol{\alpha}_{t+1} &= \widetilde{\mathbf{T}}_t \boldsymbol{\alpha}_t + \widetilde{\mathbf{R}}_t \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \text{N}_{RS}(\mathbf{0}, \mathbf{Q}_t) \end{aligned}$$

with $\boldsymbol{\alpha}_1 \sim \text{N}_{FG}(\widehat{\boldsymbol{\alpha}}_{1|0}, \mathbf{P}_{1|0})$. The elements $\widetilde{\mathbf{Z}}_t = \sum_{j=1}^{J_1}(\mathbf{S}_{j,t} \otimes \mathbf{Z}_{j,t})$, $\widetilde{\mathbf{T}}_t = \sum_{j=1}^{J_2}(\mathbf{U}_{j,t} \otimes \mathbf{T}_{j,t})$, and $\widetilde{\mathbf{R}}_t = \sum_{j}^{J_3}(\mathbf{B}_{j,t} \otimes \mathbf{R}_{j,t})$ are obtained from Equations (2.3) and (2.4) by simply using the distributive property over matrix additions and the property that states that, for compatible matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}^\top$, $\text{vec}(\mathbf{A}\mathbf{B}\mathbf{C}^\top) = (\mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B})$ (Magnus and Neudecker, 2019).

It is important to note that, through this representation, it is possible to write the matrix state space model in the vector form presented in Equations (2.1) and (2.2). The matrix representation allows the construction of interpretable models that preserve the original matrix form of the data. Deriving the vector representation enables to leverage a wide range of tools that have already been developed for the latter form and, thus, available and established for different types of time series analysis, some of which are reviewed in Section 2.4.

2.3 Other models for matrix-variate observations

2.3.1 Alternative matrix state space models

Matrix-variate time series analysis has drawn the attention of various research fields over the years. Some early models in the literature date back to the work by Quintana (1987), elements of which were later discussed in the book by West and Harrison (1997) and in more recent work by Carvalho and West (2007) and Wang and West (2009). In addition to the previously mentioned work of Choukroun *et al.* (2006), matrix state space models can be considered a special case of dynamic models for tensor responses, discussed by Rogers *et al.* (2013) and Chen *et al.* (2021b), among others. Of these, we describe the model of Wang and West (2009), as it is general enough to simultaneously include the models described in Quintana (1987), West and Harrison (1997), and Carvalho and West (2007). The description of tensor models is beyond the scope of this section.

Let $\mathcal{Y}_{1:T} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_T\}$ be a matrix-variate time series, with $\mathbf{Y}_t \in \mathbb{R}^{P\times Q}$. The dynamic linear model developed by Wang and West (2009) is characterized by the following equations

$$\mathbf{Y}_t = \mathbf{Z}_t\mathbf{A}_t + \mathbf{E}_t, \quad \mathbf{E}_t \sim \mathrm{MN}_{P,Q}(\mathbf{0}, \mathbf{H}_t), \tag{2.5}$$

$$\mathbf{A}_{t+1} = \mathbf{T}_t\mathbf{A}_t + \Xi_t, \quad \Xi_t \sim \mathrm{MN}_{F,Q}(\mathbf{0}, \mathbf{Q}_t), \tag{2.6}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{F,Q}(\widehat{\mathbf{A}}_{1|0}, \mathbf{P}_{1|0})$. In the equations, $\mathbf{Z}_t = \mathbf{I}_P \otimes \mathbf{Z}^0_t$ where $\mathbf{Z}^0_t$ is a known row vector of dimension F_0, and $\mathbf{T}_t = \mathbf{I}_P \otimes \mathbf{T}^0_t$ where $\mathbf{T}^0_t$ is a known state transition matrix of dimensions $F_0 \times F_0$. The covariance matrices are $\mathbf{H}_t = \mathbf{G}^C \otimes \mathbf{G}^R$, $\mathbf{Q}_t = \mathbf{G}^C \otimes (\mathbf{G}^R \otimes \mathbf{Q}^0_t)$, and $\mathbf{P}_{1|0} = \mathbf{G}^C \otimes (\mathbf{G}^R \otimes \mathbf{Q}^0_0)$, where both $\mathbf{G}^R$ and $\mathbf{G}^C$ are covariance matrices whose structures are associated with two underlying graphs $\mathcal{G}^R$, for rows, and $\mathcal{G}^C$, for columns, respectively. $\mathbf{Q}^0_t$ is an $F_0 \times F_0$ discount factor. The dimensions of $\mathbf{A}_t$ are $F \times Q$, where $F = PF_0$ and Q corresponds to the number of columns in $\mathbf{Y}_t$. Under this specification, each (p, q)–th entry of $\mathbf{Y}_t$ is described by the same model

$$\begin{aligned} y_{pq,t} &= \mathbf{Z}^0_t\boldsymbol{\alpha}_{pq,t} + \varepsilon_{pq,t}, \quad \varepsilon_{pq,t} \sim \mathrm{N}_1(0, g^R_{pp}g^C_{qq}), \\ \boldsymbol{\alpha}_{pq,t+1} &= \mathbf{T}^0_t\boldsymbol{\alpha}_{pq,t} + \boldsymbol{\xi}_{pq,t}, \quad \boldsymbol{\xi}_{pq,t} \sim \mathrm{N}_{F_0}(\mathbf{0}, g^R_{pp}g^C_{qq}\mathbf{Q}^0_t), \end{aligned}$$

for $\boldsymbol{\alpha}_{pq,1} \sim \mathrm{N}_{F_0}(\widehat{\boldsymbol{\alpha}}_{pq,1|0}, g^R_{pp}g^C_{qq}\mathbf{Q}^0_0)$, in which the scalar observation $y_{pq,t}$ is linked to the dynamic vector of the states through the linear relation $\mathbf{Z}^0_t\boldsymbol{\alpha}_{pq,t}$, and the model's matrices $\mathbf{Z}^0_t$ and $\mathbf{T}^0_t$ are shared across all elements of $\mathbf{Y}_t$, finding justification in the context of exchangeable time series.

The model is easily ascribable as a special case of the one presented in Equations (2.3) and (2.4), in which $J_1 = J_2 = J_3 = 1$. The right matrix $\mathbf{S}_{1,t} = \mathbf{I}_Q$ is the identity matrix, meaning that each column of $\mathbf{Y}_t$ has its own column in $\mathbf{A}_t$ and that different columns of $\mathbf{Y}_t$ interact with each other only through $\mathbf{E}_t$, by means of the column covariance matrix $\mathbf{G}^C$. Similarly, $\mathbf{U}_{1,t} = \mathbf{I}_Q$. This implies that interactions among columns of $\mathbf{A}_t$ are determined by the disturbance $\boldsymbol{\Xi}_{t-1}$, by means of the column covariance matrix $\mathbf{G}^C$. The fact that both the errors' and state disturbances' matrices are characterized by the same covariance matrices is a practical solution that finds justification in the context of the graphical models considered by Wang and West (2009). In general, such strong assumptions are not required by the model described in Equations (2.3) and (2.4). In particular, the right matrices of the model are not required to be identities, and the covariance matrices may differe in principle. In this way, the rows or columns of $\mathbf{Y}_t$ may share the rows or columns of $\mathbf{A}_t$, or, conversely, they may be characterized by multiple rows or columns of states, respectively.

2.3.2 Matrix-variate regression model

The matrix-variate regression model for the observation $\mathbf{Y}_t \in \mathbb{R}^{P\times Q}$ with predictor $\mathbf{X}_t \in \mathbb{R}^{F\times G}$, introduced by Ding and Cook (2018), takes the following form:

$$\mathbf{Y}_t = \mathbf{M} + \boldsymbol{\Lambda}\mathbf{X}_t\boldsymbol{\Gamma}^\top + \mathbf{E}_t. \tag{2.7}$$

In this equation, $\mathbf{M}$ is a $P \times Q$ dimensional matrix that represents the overall mean, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are the matrices of the coefficients with dimensions $P\times F$ and $Q\times G$, respectively, and $\mathbf{E}_t$ is an error term assumed to have mean equal to zero and covariance $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R$. Further error $\mathbf{E}_t$ is assumed to be independent of both the observed variable $\mathbf{X}_t$ and any other $\mathbf{E}_s$ for any $t = 1, \dots, T$ and $s \neq t$. Multiple viewpoints are available to interpret this model. From now on, for simplicity, we consider centered observations in which $\mathbf{M} = \mathbf{0}$. At first, the model is said to be *bi-linear*, since it involves a left and right linear transformation of the predictors matrix through the matrices of the coefficients. For fixed parameters, the expected value of $\mathbf{Y}_t$, given $\mathbf{X}_t$, is $\boldsymbol{\Gamma}\mathbf{X}_t\boldsymbol{\Lambda}^\top$, highlighting how $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}^\top$ reflect the row-wise and column-wise interactions between $\mathbf{Y}_t$ and $\mathbf{X}_t$, respectively. The generic (p, q)–th element of $\mathbf{Y}_t$ is indeed expressed as

$$y_{pq,t} = \sum_{j=1}^{F}\sum_{k=1}^{G} \lambda_{pj} x_{jk,t} \gamma_{qk} + \varepsilon_{pq,t} = \boldsymbol{\lambda}_{p\cdot}\mathbf{X}_t\boldsymbol{\gamma}_{q\cdot}^\top + \varepsilon_{pq,t},$$

where $\boldsymbol{\lambda}_{p\cdot}$ and $\boldsymbol{\gamma}_{q\cdot}^{\top}$ denote the p-th row of $\boldsymbol{\Lambda}$ and the q-th column of $\boldsymbol{\Gamma}^{\top}$. Hence, the observation $y_{pq,t}$ is determined by the p-th row of $\boldsymbol{\Lambda}$ and the q-th column $\boldsymbol{\Gamma}^{\top}$, which correspond to the vectors of the coefficients that collect the effects of $\mathbf{X}_t$ on the p-th row and the q-th column of $\mathbf{Y}_t$, respectively. Another way to interpret the model is in terms of a representation that involves multiple multivariate regression models of rows and columns of $\mathbf{Y}_t$. If $\boldsymbol{\Gamma} = \mathbf{I}_Q$, and $Q = G$, for example, each column $\mathbf{y}_{\cdot q,t}$ of $\mathbf{Y}_t$ can be alternatively represented by the same multivariate regression of the form

$$\mathbf{y}_{\cdot q,t} = \boldsymbol{\Lambda}\mathbf{x}_{\cdot q,t} + \boldsymbol{\varepsilon}_{\cdot q,t},$$

for each $q = 1, \ldots, Q$. In this way, $\boldsymbol{\Lambda}$ reflects how the different rows of $\mathbf{X}_t$ interact to determine different $\mathbf{Y}_t$. With $\boldsymbol{\Gamma} = \mathbf{I}_Q$, interactions between columns of $\mathbf{X}_t$ are not present. Similar reasoning applies to $\boldsymbol{\Lambda}$ when $\boldsymbol{\Gamma} = \mathbf{I}_P$. The matrix $\boldsymbol{\Gamma}^{\top}$ reflects how the different columns of $\mathbf{X}_t$ interact to determine $\mathbf{Y}_t$, with rows not interacting between each other. Under the general specification, however, neither matrix is the identity. This aspect introduces a third way of interpreting the model in terms of a two-step adjustment procedure, in which, first, a column adjustment $\widetilde{\mathbf{X}}_t = \mathbf{X}_t\boldsymbol{\Gamma}^{\top}$ is applied, and then the same multivariate regression model,

$$\mathbf{y}_{\cdot q,t} = \boldsymbol{\Lambda}\widetilde{\mathbf{x}}_{\cdot q,t} + \boldsymbol{\varepsilon}_{\cdot q,t}, \tag{2.8}$$

is applied to the columns of $\mathbf{Y}_t$. In this way, $\boldsymbol{\Lambda}$ reflects the interactions between the rows of the column-adjusted variable $\widetilde{\mathbf{X}}_t$. Interchanging the steps would trivially lead to a similar interpretation, where the roles of the rows and columns are swapped.

Let $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$ and $\boldsymbol{\varepsilon}_t = \text{vec}(\mathbf{E}_t)$. One interesting characteristic of the model is that its vectorized form,

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t,$$

is a multivariate regression model, in which both the matrix of coefficients $\mathbf{B} = \boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}$ and the error's covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R$ have a Kronecker decomposable structure. Both $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are identifiable. This is not the case with the single components $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Sigma}^R$, and $\boldsymbol{\Sigma}^C$, which are uniquely defined only up to a proportionality constant. For any $c \neq 0$, $\mathbf{B} = \boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda} = (\boldsymbol{\Gamma}/c) \otimes (c\boldsymbol{\Lambda})$, which highlights that infinite possible combinations of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ lead to the same $\mathbf{B}$. The same can be derived for $\boldsymbol{\Sigma}$ and any $c > 0$. To achieve identifiability, one typical assumption requires $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}^C$ to have unit Frobenius norm and positive 1–1 entry (see, Ding and Cook, 2018, for details).

Another aspect that is interesting to note is related to the number of parameters. The vectorized form of the model with unstructured coefficients and covariance matrix is characterized by $PFQG + PQ(PQ+1)/2$ free parameters, corresponding to the sum of the number of different entries in a $PQ \times FG$–dimensional matrix, the matrix of coefficients, and the number of different entries in a $PQ \times PQ$–dimensional covariance matrix. The same model, under the Kronecker product restrictions, is composed of $PF + QG + P(P+1)/2 + Q(Q+1)/2$ unknown parameters, which are subject to 3 other restrictions related to the identifiability constraints (2 on the norms, 1 on the sign). The reduction of the number of parameters and the easier interpretation of the coefficients' matrices are achieved at the cost of having a rigid structure imposed between the rows and columns. In fact, not only does the model impose the same dependency structure between the columns of $\mathbf{Y}_t$ and the adjusted columns of $\widetilde{\mathbf{X}}_t$ (or viceversa), it also requires that each column of the error matrix $\mathbf{E}_t$ is characterized by a covariance proportional to $\mathbf{\Sigma}^C$ (or viceversa–each row has a covariance proportional to $\mathbf{\Sigma}^R$). Other than in specific cases, such as in Hsu *et al.* (2021), where the matrix configuration of the observations is related to the place where the data are collected, this assumption is often found to be realistic when dealing with matrices of observations (see, e.g., Viroli, 2012; Huang *et al.*, 2019; Chen *et al.*, 2021a, among others). To further reduce the number of parameters, Ding and Cook (2018) combine typical sparsity principles with an envelope strategy—a relatively new approach introduced by Cook *et al.* (2010)—in which sufficient dimension reduction techniques are used to gain efficiency in the estimation. Since the purpose of this section is to describe the existing tools avaiable for the analysis of matrix-variate observations, we refer to the specialized literature for more details on envelope models (Cook *et al.*, 2010; Su *et al.*, 2016).

2.3.3 Matrix-variate autoregressive process

The matrix-variate autoregressive process developed by Chen *et al.* (2021a) for the time series $\mathcal{Y}_{1:T} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_T\}$, with $\mathbf{Y}_t \in \mathbb{R}^{P \times Q}$, is characterized by the following equation

$$\mathbf{Y}_t = \mathbf{\Lambda}\mathbf{Y}_{t-1}\mathbf{\Gamma}^\top + \mathbf{E}_t. \tag{2.9}$$

The model is called MAR(1), i.e. matrix autoregressive process of order 1, as the observations matrix $\mathbf{Y}_t$ is explained by the lagged variable $\mathbf{Y}_{t-1}$ by means of a bilinear form, that is determined by pre- and post-multiplying $\mathbf{Y}_{t-1}$ by the coefficients matrices $\mathbf{\Lambda}$ and $\mathbf{\Gamma}^\top$, respectively, plus the error term $\mathbf{E}_t$. The errors terms are assumed to follow a white noise process with covariance $\mathbf{\Sigma}$. The typical assumption is that $\mathbf{\Sigma}$ has a structure, that is decomposable by a Kronecker product of the form $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R$, where $\boldsymbol{\Sigma}^R$ and $\boldsymbol{\Sigma}^C$ are full covariance matrices with dimensions $P \times P$ and $Q \times Q$, respectively. Adopting a MAR(1) process instead of a VAR(1) process constructed with the vectorized observations has two main advantages: it maintains the original matrix structure of the observations, ensuring greater interpretability and reduces the number of parameters. As regards the number of parameters, under the MAR(1) process, the dimensions of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are $P \times P$ and $Q \times Q$, respectively; hence, if an unstructured version of these matrices is considered, the number of free parameters in the matrices of the coefficients is $(P^2 + Q^2)$, which is much lower than $(PQ)^2$ for the generic VAR(O) model under the vectorized observations. These parameters must be added to those related to the covariance matrix $\boldsymbol{\Sigma}$, whose number of different entries is $P(P+1)/2$ and $Q(Q+1)/2$, respectively, if a Kronecker decomposable structure is considered. As regards interpretability, Chen *et al.* (2021a) provide multiple viewpoints to understand the role of the parameter matrices. It is worth noting that the MAR(1) process with a variance that is decomposed by a Kronecker product can be thought as a special case of the matrix-variate regression model (previously defined), in which the matrix of the predictors is $\mathbf{X}_t = \mathbf{Y}_{t-1}$. This implies that it can be interpreted in the same way as the model in Equation (2.7). We refer to the original article by Chen *et al.* (2021a) for a detailed discussion on the stationarity and causality of the MAR(1) process.

From the initial specification of the model, certain generalizations have been proposed to obtain greater flexibility and interpretability and consider particular dependency structures between the rows and columns. First, the model is easily generalized to the generic order O by requiring $\mathbf{Y}_t$ to be explained as a sum of the lagged variables $\mathbf{Y}_{t-1}, \dots,$ $\mathbf{Y}_{t-O}$ that are pre- and post-multiplied by the respective matrices of the coefficients $\boldsymbol{\Gamma}_o$ and $\boldsymbol{\Lambda}_o^\top$ for $o = 1, \dots, O$. Second, Chen *et al.* (2021a) extend their model to propose the multiple lag-one autoregressive model, in which $\mathbf{Y}_t$ is explained by many terms that involve bilinear transformations of $\mathbf{Y}_{t-1}$, such as the model

$$\mathbf{Y}_t = \boldsymbol{\Lambda}_1 \mathbf{Y}_{t-1} + \mathbf{Y}_{t-1} \boldsymbol{\Gamma}_1^\top + \boldsymbol{\Lambda}_2 \mathbf{Y}_{t-1} \boldsymbol{\Gamma}_2^\top + \mathbf{E}_t,$$

where $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Gamma}_1^\top$ capture the rows and columns main effects, while $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Gamma}_2$ capture two-way interactions. Hsu *et al.* (2021) consider a MAR(O) process in which the neighborhood structure present in the observations matrices is exploited in both the coefficient matrices and the variance of error terms. With regard to the matrices of the coefficients, the local spatial dependence among the rows and columns is obtained by requiring a certain banding structure in which entries far away from the diagonal are imposed to zero. On the other hand, the error terms are characterized by a fixed-rank kriging model in which the covariance is $\mathbf{\Sigma} = \mathbf{F}\mathbf{\Omega}\mathbf{F}^\top + \sigma_\varepsilon^2 \mathbf{I}_{PQ}$, where $\mathbf{F}$ is a matrix of basis functions, with rank $J \leq PQ$ and dimensions $PQ \times J$, $\mathbf{\Omega}$ is a $J \times J$ non-negative definite matrix, and σ_ε^2 is an unknown coefficient. If the error terms are assumed to be Gaussian, the presented models can be written in a matrix-variate state space form in several ways, by means of many linear algebra tricks. A similar result holds for ARIMA model in the scalar case (Durbin and Koopman, 2012), but are outside the scope of this thesis.

2.3.4 Matrix-variate dynamic factor models

The matrix-variate dynamic factor model developed by Wang *et al.* (2019) for the time series $\mathcal{Y}_{1:T} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_T\}$, with $\mathbf{Y}_t \in \mathbb{R}^{P \times Q}$, is characterized by the following equation

$$\mathbf{Y}_t = \mathbf{\Lambda}\mathbf{X}_t\mathbf{\Gamma}^\top + \mathbf{E}_t. \tag{2.10}$$

In this equation, $\mathbf{X}_t$ is the $F \times G$–dimensional matrix with common fundamental factors, which are unobserved and drive the dynamics and co-movements of $\mathbf{Y}_t$, whose dimensions are much higher than those of $\mathbf{X}_t$ (i.e., $F \ll P$ and $G \ll Q$). In general, the number of rows in the matrix $\mathbf{X}_t$ corresponds to the rank of $\mathbf{\Lambda}$, and its number of columns corresponds to the rank of $\mathbf{\Gamma}$ (i.e., $\text{rank}(\mathbf{\Lambda}) = F$ and $\text{rank}(\mathbf{\Gamma}) = G$). The elements $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ are the left and the right factor loading matrices of dimensions $P \times F$ and $Q \times G$, respectively, and reflects the importance of common fundamental factors. Other ways to interpret the model, which are similar to those discussed for the MAR(1) model, are reported in the work of Wang *et al.* (2019).

One common assumption regarding the factor process is the one proposed by Lam *et al.* (2011), where its vectorized representation is assumed to be a weak stationary process with two finite first moments, independent of the the errors $\mathcal{E}_{1:T} = \{\mathbf{E}_1, \dots, \mathbf{E}_T\}$. The noise process $\mathcal{E}_{1:T}$ is generally assumed to be white noise, where the generic $\mathbf{E}_t$ is a zero-mean matrix of errors with fixed covariance matrix $\mathbf{\Sigma}$. A typical problem related to the assumption made regarding the erratic component is that it is identified as a factor when a weak form of correlation is present. Chen *et al.* (2020) try to mitigate this issue by introducing a constrained matrix factor model, in which the loading matrix $\mathbf{\Lambda} = \mathbf{C}_\Lambda \mathbf{\Lambda}_0$ and $\mathbf{\Gamma} = \mathbf{C}_\Gamma \mathbf{\Gamma}_0$ are subject to known constraints that are related to prior or domain knowledge on the variables and are incorporated in the model through the constraint matrices $\mathbf{C}_\Lambda$ and $\mathbf{C}_\Gamma$. Under this specification, $\mathbf{C}_\Lambda$ and $\mathbf{C}_\Gamma$ are full column rank matrices of dimensions $P \times F_0$ and $Q \times G_0$, while $\mathbf{\Gamma}_0$ and $\mathbf{\Lambda}_0$ are loading matrices whose dimensions are $F_0 \times F$ and $G_0 \times G$, respectively. To be meaningful, the dimensions of the involved matrices have to be such that $F \leq F_0 \ll P$ and $G \leq G_0 \ll Q$. In addition, the structure of the constraint matrices is determined by the type of constraints to be imposed. A simple case would be where the constraint matrices are made up of selection matrices whose rows are composed of the rows of the identity matrix and whose role is to select specific rows (or columns) of $\mathbf{\Lambda_0 X_t \Gamma_0^\top}$. The use of such selection matrices is discussed in more detail in the next chapter, and we refer to Chen *et al.* (2020) for other examples. Additional multifactorial models have been presented in the literature, in which the measurement equation is determined by an additive combination of multiple dynamic latent factors. Their treatment is no different from the standard case seen in Equation (2.10) and is covered in the works of Wang *et al.* (2019) and Chen *et al.* (2020).

If each matrix of errors $\mathbf{E}_t$ and of disturbances $\mathbf{\Xi}_t$ are assumed to be Gaussian, the link with the matrix-variate state space model presented in Equations (2.3) and (2.4) is easy to derive. In particular, Equation (2.10) represents the measurement equation, in which $J_1 = 1$, the loading matrices $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ correspond to $\mathbf{Z}_{1,t}$ and $\mathbf{S}_{1,t}$, respectively, and the matrix latent factor $\mathbf{X}_t$ represents the matrix of latent states $\mathbf{A}_t$. The assumption of weak stationarity can be obtained by first imposing some suitable conditions on the model in the vectorized form and then deriving the respective matrix form of the state transition equation.

2.4 Essential toolbox for state space modeling

One of the conveniences of state space models is related to the possibility of treating a wide range of problems in a unified way through routines known as Kalman recursions. The derivation of these last is covered in many textbooks (Durbin and Koopman, 2012; Shumway and Stoffer, 2017), and, here, we report some results that we use in the following chapters, providing only a brief explanation of their meaning. With regard to the Kalman filter, Kalman smoother, and lagged Kalman smoother, their derivation is possible regardless of whether the approach used is classical or Bayesian. The likelihoods are obtained under the normality assumptions. We further consider the "reduction by transformation" technique proposed by Jungbacker and Koopman (2008) for the quick estimation of dynamic factor models and discuss its use with matrix-variate dynamic factor models.

2.4.1 Kalman routines

Consider the vector state space in Equations (2.1) and (2.2), where mean $\widehat{\boldsymbol{\alpha}}_{1|0}$, and covariance $\mathbf{P}_{1|0}$ are assumed to be known. Let $\widehat{\boldsymbol{\alpha}}_{t|t} = \mathrm{E}(\boldsymbol{\alpha}_t|\mathcal{Y}_{1:t})$, $\mathbf{P}_{t|t} = \mathrm{Var}(\boldsymbol{\alpha}_t|\mathcal{Y}_{1:t})$, $\widehat{\boldsymbol{\alpha}}_{t+1|t} = \mathrm{E}(\boldsymbol{\alpha}_{t+1}|\mathcal{Y}_{1:t})$, and $\mathbf{P}_{t+1|t} = \mathrm{Var}(\boldsymbol{\alpha}_{t+1}|\mathcal{Y}_{1:t})$. The Kalman filter is an iterative procedure composed of the following steps

$$\begin{aligned} \boldsymbol{v}_t &= \mathbf{y}_t - \mathbf{Z}_t\widehat{\boldsymbol{\alpha}}_{t|t-1} & \mathbf{F}_t &= \mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t^\top + \mathbf{H}_t \\ \widehat{\boldsymbol{\alpha}}_{t|t} &= \widehat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}_t^\top\mathbf{F}_t^{-1}\boldsymbol{v}_t & \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}_t^\top\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}_{t|t-1} \\ \widehat{\boldsymbol{\alpha}}_{t+1|t} &= \mathbf{T}_t\widehat{\boldsymbol{\alpha}}_{t|t} & \mathbf{P}_{t+1|t} &= \mathbf{T}_t\mathbf{P}_{t|t}\mathbf{T}_t^\top + \mathbf{R}_t\mathbf{Q}_t\mathbf{R}_t^\top \end{aligned}$$

for $t = 1, \ldots, T$. The vector $\boldsymbol{v}_t$ collects the one-step ahead forecast error of $\mathbf{y}_t$ given $\mathcal{Y}_{1:(t-1)}$, and it is also called the vector of *innovations*. The matrix $\mathbf{F}_t$ represents its variance, i.e. $\mathbf{F}_t = \mathrm{Var}(\boldsymbol{v}_t|\mathcal{Y}_{1:(t-1)})$. The element $\widehat{\boldsymbol{\alpha}}_{t|t}$ is the vector of *filtered* states, where $\mathbf{P}_{t|t}$ is its respective variance. Similarly, $\widehat{\boldsymbol{\alpha}}_{t+1|t}$ is the vector of *predicted* states, where $\mathbf{P}_{t+1|t}$ represents respective variance. To summarize, the Kalman filter is a forward procedure that allows to update the knowledge about the states at time t to be updated once the observations for this time point are available. The filter has no retroactive action, which means that it does not update the knowledge about the states of previous time points once the observation at time t is available. One of the main bottlenecks of the algorithm is the need to invert the $L \times L$–dimensional matrix $\mathbf{F}_t$ at each time instant. Depending on the problem, various alternative approaches can be applied to improve performance, such as by using the Sherman-Morrison-Woodbury formula, exploiting the sparsity of some of the matrices involved, or reducing the size of the vector of observations (Durbin and Koopman, 2012). We discuss the use of the latter approach in Section 2.4.3.

Let $\widehat{\boldsymbol{\alpha}}_{t|T} = \mathrm{E}(\boldsymbol{\alpha}_t|\mathcal{Y}_{1:T})$ and $\mathbf{P}_{t|T} = \mathrm{Var}(\boldsymbol{\alpha}_t|\mathcal{Y}_{1:T})$. The classical Kalman *smoother*, originally developed by Rauch *et al.* (1965), is an iterative procedure such that

$$\begin{aligned} \mathbf{J}_{t-1} &= \mathbf{P}_{t-1|t-1}\mathbf{T}_t^\top\mathbf{P}_{t|t-1}^{-1} \\ \widehat{\boldsymbol{\alpha}}_{t-1|T} &= \widehat{\boldsymbol{\alpha}}_{t-1|t-1} + \mathbf{J}_{t-1}(\widehat{\boldsymbol{\alpha}}_{t|T} - \widehat{\boldsymbol{\alpha}}_{t|t-1}) \\ \mathbf{P}_{t-1|T} &= \mathbf{P}_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{P}_{t|T} - \mathbf{P}_{t|t-1})\mathbf{J}_{t-1}^\top \end{aligned}$$

for $t = T, T-1, \ldots, 2$. The algorithm is a backward recursion in which the previously calculated means and variances are adjusted based on all available information. It, therefore, finds its utility in retrospective analyses, where all available information is used to analyze the historical course of events. The algorithm requires the inversion of the predicted covariance matrices, which are $K \times K$–dimensional covariance matrices. We refer to the discussion of Durbin and Koopman (2012) on the alternative algorithms in which these inversions are avoided. Another algorithm that is used in our tractation is called *lagged-one* smoother, which allows to determine $\mathbf{P}_{t-1,t|T} = \text{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t-1}|\mathcal{Y}_{1:T})$ (Shumway and Stoffer, 2017). Its backward recursion simply computes

$$\mathbf{P}_{t-1,t-2|T} = \mathbf{P}_{t-1|t-1}\mathbf{J}_{t-2}^{\top} + \mathbf{J}_{t-1}(\mathbf{P}_{t,t-1|T} - \mathbf{T}\mathbf{P}_{t-1|t-1})\mathbf{J}_{t-2}^{\top}$$

for $t = T, \ldots, 2$.

2.4.2 Likelihoods

Regardless of whether they are classical or Bayesian, many statistical procedures use the joint density of the observations as the tool that is at the heart of inference. We first introduce the concept of *augmented likelihood*, which is the density of the observations $\mathcal{Y}_{1:T}$ jointly evaluated with the latent states $\mathcal{A}_{1:T} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T\}$, i.e.

$$\begin{aligned} p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}, \mathcal{A}_{1:T}) &= p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}|\mathcal{A}_{1:T})p_{\boldsymbol{\theta}}(\mathcal{A}_{1:T}) \\ &= \left[\prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\mathbf{y}_t|\boldsymbol{\alpha}_t)\right]\left[p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_1)\prod_{t=2}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1})\right]. \end{aligned}$$

The dependence on the parameter $\boldsymbol{\theta}$ is explicit through the subscript in $p_{\boldsymbol{\theta}}(\cdot)$ but has to be considered as a conditional dependence in the Bayesian setting. The quantity $p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}|\mathcal{A}_{1:T})$ is called *conditional likelihood*, and it represents the density of the observations provided that the states are given. This quantity is determined by the measurement equation in Equation (2.1) and can be written as a product of T independent Gaussian densities, evaluated in $\mathbf{y}_t$, with mean $\mathbf{Z}_t\boldsymbol{\alpha}_t$ and covariance $\mathbf{H}_t$, for $t = 1, \ldots, T$. The quantity $p_{\boldsymbol{\theta}}(\mathcal{A}_{1:T})$ is the joint density of the latent states and is determined by the state transition equation in Equation (2.2). It is a product of T densities with a Markovian structure and depends on how $\mathbf{T}_t$, $\mathbf{R}_t$, and $\mathbf{Q}_t$ are specified for $t = 1, \ldots, T$.

The likelihood of the observed process is obtained by marginalization and has a set of alternative representations. Specifically, we refer to the likelihood when considering

$$p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}) = \int p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}, \mathcal{A}_{1:T})\mathrm{d}\mathcal{A}_{1:T}.$$

It is interesting to note that it can be expressed as

$$p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}) = p_{\boldsymbol{\theta}}(\mathbf{y}_1)p_{\boldsymbol{\theta}}(\mathbf{y}_2|\mathcal{Y}_1)\cdots p_{\boldsymbol{\theta}}(\mathbf{y}_T|\mathcal{Y}_{1:(T-1)}),$$

where

$$
\begin{aligned}
p_{\boldsymbol{\theta}}(\mathbf{y}_t|\mathcal{Y}_{1:(t-1)}) &= \int p_{\boldsymbol{\theta}}(\mathbf{y}_t|\boldsymbol{\alpha}_t, \mathcal{Y}_{1:(t-1)})p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_t|\mathcal{Y}_{1:(t-1)})\mathrm{d}\boldsymbol{\alpha}_t \\
&= \int p_{\boldsymbol{\theta}}(\mathbf{y}_t - \mathbf{Z}_t\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_t, \mathcal{Y}_{1:(t-1)})p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_t|\mathcal{Y}_{1:(t-1)})\mathrm{d}\boldsymbol{\alpha}_t = p_{\boldsymbol{\theta}}(\boldsymbol{v}_t|\mathcal{Y}_{1:(t-1)}).
\end{aligned}
$$

It follows that the likelihood can be also expressed in terms of the innovations as

$$
p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}) = \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{v}_t|\mathcal{Y}_{1:(t-1)}) = \prod_{t=1}^{T}(2\pi)^{-L/2}\det(\mathbf{F}_t)^{-L/2}\exp\big[-\frac{1}{2}\boldsymbol{v}_t^{\top}\mathbf{F}_t^{-1}\boldsymbol{v}_t\big], \tag{2.11}
$$

where $\boldsymbol{v}_t$ and $\mathbf{F}_t$ are obtained by the Kalman filtering recursion, and its dependence on $\boldsymbol{\theta}$ is determined by the model configuration.

2.4.3 Reduction by transformation technique

Consider, for simplicity, that $\mathbf{Z}_t = \mathbf{Z}$ and $\mathbf{H}_t = \boldsymbol{\Sigma}$, so the vector state space model can be expressed as

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{Z}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathrm{N}_L(\mathbf{0}, \boldsymbol{\Sigma}), \\
\boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{R}_t\boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \mathrm{N}_M(\mathbf{0}, \mathbf{Q}_t),
\end{aligned} \tag{2.12}
$$

with $\boldsymbol{\alpha}_1 \sim N_K(\widehat{\boldsymbol{\alpha}}_{1|0}, \mathbf{P}_{1|0})$. The reduction by transformation technique, proposed by Jungbacker and Koopman (2008), aims to reduce the computational complexity of Kalman filtering when L, the dimension of the vector of observations, is large. This technique aims to find a non-singular matrix

$$
\mathbf{A} = \begin{bmatrix} \mathbf{A}' \\ \mathbf{A}'' \end{bmatrix},
$$

such that, if we consider the transformation

$$
\mathbf{y}_t^{\star} = \mathbf{A}\mathbf{y}_t = \begin{bmatrix} \mathbf{A}'\mathbf{y}_t \\ \mathbf{A}''\mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{y}_t' \\ \mathbf{y}_t'' \end{bmatrix},
$$

the resulting model for $\mathbf{y}_t^\star$ takes the form

$$\mathbf{y}_t' = \mathbf{A}'\mathbf{Z}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t', \tag{2.13}$$

$$\mathbf{y}_t'' = \boldsymbol{\varepsilon}_t'', \tag{2.14}$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{R}_t\boldsymbol{\xi}_t, \tag{2.15}$$

with $\boldsymbol{\varepsilon}_t' = \mathbf{A}'\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_t'' = \mathbf{A}''\boldsymbol{\varepsilon}_t$ satisfying the relations

$$\mathrm{E}(\boldsymbol{\varepsilon}_t') = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{\varepsilon}_t'') = \mathbf{0}, \quad \mathrm{Var}(\boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma}', \quad \mathrm{Var}(\boldsymbol{\varepsilon}_t'') = \boldsymbol{\Sigma}'', \quad \mathrm{Cov}(\boldsymbol{\varepsilon}_t', \boldsymbol{\varepsilon}_t'') = \mathbf{0},$$

for $t = 1, \ldots, T$, $\boldsymbol{\Sigma}' = \mathbf{A}'\boldsymbol{\Sigma}\mathbf{A}'^\top$, and $\boldsymbol{\Sigma}'' = \mathbf{A}''\boldsymbol{\Sigma}\mathbf{A}''^\top$. In the above equations, $\mathbf{A}$ is an $L \times L$ block matrix composed of the $l \times L$ matrix $\mathbf{A}'$ and the $(L - l) \times L$ matrix $\mathbf{A}''$, where l is exactly the rank of $\mathbf{Z}$. The relations in the equation of transitions have not been modified instead. Calculating the matrix $\mathbf{A}$ requires that only $\mathbf{y}_t'$ is material for state estimation, that $\mathbf{y}_t''$ is immaterial, and, thus, that Equation (2.14) can be neglected during state estimation routines.

Jungbacker and Koopman (2008) achieve this goal by providing conditions and lemmas for finding such matrices. In practice, suitable matrices $\mathbf{A}$ have $\mathbf{A}'$ such that $\mathbf{A}' = (\mathbf{Z}_\dagger)^\top\boldsymbol{\Sigma}^{-1}$, where $\mathbf{Z}_\dagger$ is an $L \times l$ with columns that form a basis for the column space of $\mathbf{Z}$. If $\mathbf{Z}$ is not full column rank, then $l < K$, and $\mathbf{Z}_\dagger$ can be obtained by the decomposition $\mathbf{Z} = \mathbf{Z}_\dagger\mathbf{C}$ for any full rank $l \times K$ matrix $\mathbf{C}$. If $\mathbf{Z}$ is a full column rank matrix instead, then $l = K$, and $\mathbf{Z}_\dagger = \mathbf{Z}\mathbf{C}^{-1}$ for any non-singular matrix $\mathbf{C}$ of dimension $l \times l$. Finding possible candidates for $\mathbf{Z}_\dagger$ may be computationally convenient when $l \ll L$, as the matrices of innovations would become $l \times l$–dimensional, and their inversions would become much more efficient. The authors do not focus on the form of $\mathbf{A}''$ and, hence, on that of $\boldsymbol{\Sigma}''$ because it is not used in practice and because its computation requires intensive computational routines. The only condition they require is that the determinant $\det(\boldsymbol{\Sigma}'') = 1$, which is a convenient condition from a practical point of view, albeit unnecessary, since it simplifies likelihood derivation. Thus, while this approach is used for estimating states efficiently, it is useful to understand how to compute the likelihood without any loss of information using the reduced vectors of observations. Jungbacker and Koopman (2008) show that it is possible to obtain the likelihood exactly without needing to know $\mathbf{A}''$. To obtain this result, they first consider the likelihood of the observed process, which is

$$\begin{aligned} p_{\boldsymbol{\theta}}(\mathcal{Y}_{1:T}) &= p_{\boldsymbol{\theta}}(\mathcal{Y}^{\star}_{1:T}) \det(\mathbf{A})^{T} \\ &= p_{\boldsymbol{\theta}}(\mathcal{Y}'_{1:T}) p_{\boldsymbol{\theta}}(\mathcal{Y}''_{1:T}) \det(\mathbf{A})^{T}, \end{aligned} \tag{2.16}$$

where $\mathcal{Y}'_{1:T} = \{\mathbf{A}'\mathbf{y}_1, \ldots, \mathbf{A}'\mathbf{y}_T\}$, and $\mathcal{Y}''_{1:T} = \{\mathbf{A}''\mathbf{y}_1, \ldots, \mathbf{A}''\mathbf{y}_T\}$. In the factorization of Equation (2.16), $p_{\boldsymbol{\theta}}(\mathcal{Y}'_{1:T})$ denotes the likelihood of the observed process in the reduced form, which is obtained by applying the Kalman routines to Equations (2.13) and (2.15). The quantity $p_{\boldsymbol{\theta}}(\mathcal{Y}''_{1:T})$ is the part of the likelihood that is related to Equation (2.14), which is immaterial for the state estimation and can be obtained as a product of T independent normal densities, with zero mean and variance $\boldsymbol{\Sigma}'' = \mathbf{A}''\boldsymbol{\Sigma}\mathbf{A}''^{\top}$, whose determinant is 1. The dependence of $p_{\boldsymbol{\theta}}(\mathcal{Y}''_{1:T})$ on $\mathbf{A}''$, which is directly involved in computing $\boldsymbol{\Sigma}''$, is dropped if $\mathbf{A}' = \mathbf{Z}_{\dagger}^{\top}\boldsymbol{\Sigma}^{-1}$ is considered for any decomposition $\mathbf{Z} = \mathbf{Z}_{\dagger}\mathbf{C}$ with $\mathbf{C}$ full column rank. Therefore, given the *generalized least square residuals*

$$\mathbf{e}_t = \Big[\mathbf{I}_L - \mathbf{Z}_{\dagger}(\mathbf{Z}_{\dagger}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{Z}_{\dagger})^{-1}\mathbf{Z}_{\dagger}^{\top}\Big]\mathbf{y}_t,$$

Jungbacker and Koopman (2008) show that

$$\mathbf{y}_t''^{\top}(\boldsymbol{\Sigma}'')^{-1}\mathbf{y}_t'' = \mathbf{e}_t{}^{\top}(\boldsymbol{\Sigma})^{-1}\mathbf{e}_t,$$

leading to

$$p_{\boldsymbol{\theta}}(\mathcal{Y}''_{1:T}) = \prod_{t=1}^{T}(2\pi)^{-(L-l)/2}\exp\big[-\frac{1}{2}\mathbf{e}_t{}^{\top}(\boldsymbol{\Sigma})^{-1}\mathbf{e}_t\big].$$

Finally, the identity

$$\det(\mathbf{A})^2 = \det(\boldsymbol{\Sigma})^{-1}\det(\boldsymbol{\Sigma}')$$

holds, which allows to recover all the elements of the likelihood in Equation (2.16) without knowing $\mathbf{A}''$.

As noted in Equation (2.12), the reduction by transformation technique requires the $\mathbf{Z}$ and $\boldsymbol{\Sigma}$ matrices to be time independent. In cases where these matrices are non-stochastic, as in Equations (2.1) and (2.2), this technique can be still applied even though these matrices are time dependent. However, it would require to obtain $\mathbf{Z}_{\dagger,t}$ for each time instant using a suitable decomposition $\mathbf{Z}_t = \mathbf{Z}_{\dagger,t}\mathbf{C}_t$ for any full column rank matrix $\mathbf{C}_t$. The convenience of using this technique is, therefore, subordinate to the ability of beeing able to find those suitable decompositions without the computational burden being excessive. There is no fixed rule for doing this, and, often, the decomposition is closely related to the model under consideration. Let $\mathbf{Z}_t$ be a generic $L \times K$ real matrix such that $0 < \text{rank}(\mathbf{Z}_t) = l < L$. The (thin) singular-value decomposition applies; hence, $\mathbf{Z}_t$ can be expressed as $\mathbf{Z}_t = \mathbf{Z}'_t \boldsymbol{\Lambda}_t \mathbf{Z}''^{\top}_t$, where $\mathbf{Z}'_t$ and $\mathbf{Z}''_t$ are matrices such that $\mathbf{Z}'^{\top}_t \mathbf{Z}'_t = \mathbf{I}_l$, and $\mathbf{Z}''^{\top}_t \mathbf{Z}''_t = \mathbf{I}_l$, and $\boldsymbol{\Lambda}_t$ is an $l \times l$ diagonal matrix with positive eigenvalues (Magnus and Neudecker, 2019). The decomposition $\mathbf{Z}_t = \mathbf{Z}_{\dagger,t}\mathbf{C}_t$ can be obtained by imposing $\mathbf{Z}_{\dagger,t} = \mathbf{Z}'_t$ and $\mathbf{C}_t = \boldsymbol{\Lambda}_t \mathbf{Z}''^{\top}_t$, where $\mathbf{C}_t$ is an $l \times K$ full rank matrix. Note, however, that Jungbacker and Koopman (2008) do not require $\mathbf{Z}'_{\dagger,t}$ to be orthogonal, in order to obtain a suitable decomposition of $\mathbf{Z}_t$. Hence, even simpler decompositions can be considered, as shown in the following example.

Example with matrix-variate dynamic factor models. Consider the model in Equation (2.10) under Gaussian assumptions, and assume that $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are full column rank matrices such that $\text{rank}(\boldsymbol{\Lambda}) = F$ and $\text{rank}(\boldsymbol{\Gamma}) = G$, respectively. The vectorized form of the measurement equation is given by

$$\mathbf{y}_t = (\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda})\mathbf{x}_t + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$, $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$, and $\boldsymbol{\varepsilon}_t = \text{vec}(\mathbf{E}_t)$. It is clear that

$$\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda} = (\boldsymbol{\Gamma}\mathbf{I}_G) \otimes (\boldsymbol{\Lambda}\mathbf{I}_F) = (\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda})(\mathbf{I}_G \otimes \mathbf{I}_F),$$

where $\mathbf{I}_G \otimes \mathbf{I}_F = \mathbf{I}_{FG}$ is the FG–dimensional identity matrix. Fix $\mathbf{Z}_\dagger = \boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}$ and $\mathbf{C} = \mathbf{I}_G \otimes \mathbf{I}_F$. The reduction by transformation technique can be obtained using

$$\mathbf{A}' = (\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda})^{\top}\boldsymbol{\Sigma}^{-1} = (\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda})^{\top}(\boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R)^{-1} = \left[\boldsymbol{\Gamma}^{\top}(\boldsymbol{\Sigma}^C)^{-1}\right] \otimes \left[\boldsymbol{\Lambda}^{\top}(\boldsymbol{\Sigma}^R)^{-1}\right],$$

where the last two relations hold if $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R$. In the reduced model, the Kalman routines are applied to $l = FG$–dimensional vectors instead of $L = NP$.

2.5 Discussion

In the present chapter, state space models have been reviewed, with a focus on vector and matrix state space models and related tools useful for the analysis. In sports performance analysis, the use of state space models can be varied, as the time component is present in several aspects of sports activities. For example, an athlete can monitor multiple variables (weight, weekly miles, minutes of activity, etc.) on a weekly basis: as a result, a multivariate time series can be observed. Similarly, athletes can monitor themselves during a single activity. The data collected during the activity can be represented as a multivariate time series as well. Matrix time series can also be observed in various contexts. For example, if the activity is monitored over time and different sensors placed in different parts of the body measure the same variables, we can obtain a matrix time series. In this case, variables (e.g., speeds along the 3 dimensions of space) can be stored in different columns, one for each sensor, which together build a matrix that can be monitored over time. In the same way, multiple multivariate time series obtained by the athletes' self-monitoring can be arranged in a matrix time series. While some phenomena fits well with a matrix time series, in other contexts their use is a technical expedient to use models that can represent the data structures in a more compact way. Models and tools presented here have been employed in Chapter 3 and Chapter 4 at various points.

In Chapter 3 the vector state space model is used as a building block for a matrix state space model, and it is helpful for monitoring the performances in multiple races of different athletes over the years. The derived model can be interpreted in a similar way of the dynamic factor model proposed by Wang *et al.* (2019) and Chen *et al.* (2020), where a reduced number of states determines the observed behavior of the variables. Different states are linked to the behavior of each athlete's observed race, by using a random selection matrix; this last has the role of selecting, among the reduced number of states, those that describe the observed behavior of each athlete over time. Selection matrices are used in Chen *et al.* (2020) with the aim to constrain the dynamic factor model based on prior knowledge of their application domain. However, in contrast to their approach, in our model the involved selection matrix is unknown, thus it is a direct object of inference.

In Chapter 4 the vector state space model is combined with classical changepoint models (Chib, 1998; Fearnhead and Liu, 2007; Yildirim *et al.*, 2013). In this case, a sequence of activities performed by one athlete is observed, and our proposal aims to detect changes in behavior due to health problems (and other issues) during the performances. More specifically, each activity is represented by a multivariate time series, that stores observations of different variables over time. Hence, state space models are used to describe the behavior of the observed variables within each activity in time. Changepoint models are used to detect changes between subsequent activities that modify the behavior of the observed variables. Throughout the chapter, inferential tools (e.g. Kalman recursions) here presented will be extensively used with the aim to provide information to the athlete in real-time.

Chapter 3

Time series clustering of athletes' careers under informative missing data patterns

3.1 Introduction

Planning the future career of young athletes is a relevant aspect of the work of coaches, whose role is to guide athletes during training so that they can perform at their best in competition, so that they can achieve their desired results. Identifying athletes' capabilities and future possibilities is important for multiple reasons. On one hand, it allows the training load to be allocated over the years, in a way that is appropriate for the athlete. Good planning, along with support during injuries, has been identified as one of the relevant factors that help avoiding drop-out in athletes (Bussmann, 1999; Larsen and Alfermann, 2017). Moreover, a well-distributed training load not only allows the athlete to improve the performance, but it also reduces the risk of injury. On the other hand, good planning is important also from a psychological and emotional point of view, as it allows athletes to strive for achievable goals and collect successes over the years. Pleasant emotions (including satisfaction) have been associated to positive outcomes in, e.g., mental health, performance and engagement (see, e.g., Cece *et al.*, 2019). In this context, the identification of possible careers for an athlete, in terms of observed personal performance trajectories over time, is of paramount importance. For example, identifying the period in which athletes reach their peaks can help prepare the athletes for the most important events in their career. Similarly, knowing the expected progress of different athletes over the years can give an indication of whether the training process has been carried out correctly. The increasing awareness of the impact that a well-distributed training can have on the development of athletes' careers had lead to several analysis of athletes' trajectories in various sports. Leroy *et al.* (2018) have studied young swimmers' progression using a functional clustering approach; in Boccia *et al.* (2017) study, they focused on individual careers of Italian long- and high-jumpers to figure out which characteristics of young athletes are predictive of high-level results during their careers. In our work, we focus on the analysis of trajectories collected by Italian male middle distance runners, born in 1988, in a period ranging from 2006 to 2019. Studies on middle distance runners are few or limited to samples with a small number of athletes (see, e.g., Weippert *et al.*, 2021).

Among many tools that can be used for this problem, clustering of trajectories allows the identification of different careers present in the data and, as a consequence, the various possible observable scenarios of athletes' careers. Clustering of longitudinal data has been extensively explored in literature (see, among others, Frühwirth-Schnatter, 2011; Maharaj *et al.*, 2019; Bartolucci and Murphy, 2015). Here we focus on the use of state space models because they allow the construction of flexible models for multivariate time series in an intuitive manner and offer a number of well-known tools for inference, including the treatment of missing data (Durbin and Koopman, 2012). Indeed, unlike other types of athletes and sports, middle distance runners can compete in different distances, i.e. in the 800, 1500 and 5000 meters races, as well as in other spurious races (i.e. the mile, 3000m, etc.). The choice of races in which to compete is subjective and typically associated with personal attitudes (Mooses *et al.*, 2013). An athlete capable of developing greater speed and power typically competes in shorter distances, with respect to those with greater endurance who compete in longer distances. In this way, not only do we observe different races for each athlete over time, but the absence of a particular race is informative on the athlete's attitude. Beyond the variability among subjects related to the type of races performed, there is also variability in the developing of athletes' careers related to both their abilities and histories. If an athlete starts the career late in life, it is less likely that will reach high levels; similarly, athletes with unsatisfactory careers are likely to end their careers earlier, with respect to those satisfied with their performances (Hernandez *et al.*, 2011). These aspects are related to drop-in and drop-out phenomena, defined as the events where athletes enter and exit the observed sample, respectively. Specifically, the presence and absence of data is potentially correlated with observed performances.

In this study we propose a model based clustering method for longitudinal data, in which missing data inform on the clustering structure. The clustering problem is addressed through the specification of a matrix state space model, in which multivariate time series are clustered on the basis of their observed trend. In this phase, clustering is achieved via a selection matrix which is involved in the measurement equation. Among the advantages of this specification, we take in consideration the ability to include with a compact notation complex dependencies, both temporal and cross-sectional. The inclusion of temporal dynamics that aim to describe missing data patterns is accomplished by two different processes. First, athlete's personal history is described by a three state process, which describe their entry (drop-in) and exit (drop-out) from the sample. Second, different propensities to participate in competitions are considered based on the athlete's personal attitude. The probabilities of both the processes are assumed to be group dependent. In this way, clustering is not only achieved on the basis of athletes' performances, but the presence and absence of data is considered informative as well. Prior information is included based on qualitative reasoning on the observed phenomenon, and inference is obtained using a Gibbs sampling algorithm. The practical example shows benefits and limitations of the proposed approach.

3.1.1 Data collection and missing data

Data refer to annual seasonal best performances of male Italian athletes, born in 1988, on 5000, 1500 and 800 meters races in a period between 2006 and 2019. Data were collected from the annual rankings accessible on the website of the Italian athletics federation (`www.fidal.it`). For each year considered, all athletes (male, born in 1988) who participated in at least one of the three competitions considered were selected. In Italy, at 18 years old, athletes transit from "Allievi" to "Junior" (18-19 years old) category, and compete in the national championships in the same distances (i.e. 800, 1500, and 5000 meters races) in the following categories ("Promesse", 20–22, and "Senior", 23+). Athletes that participated to less than 2 races were removed from the sample. This selection was made in order to not include athletes who participated just in one middle-distance race in their career for fortuitous reasons other than a true interest in the discipline. Collected data are shown in Figure 3.1. By considering each graph in marginally, it is possible to imagine a U-shaped curve that describes the distribution of sample trajectories, especially for 5000 and 1500 races. However, their shape can be biased by the presence of selection, missing data, and late entry in the sample.

The drop-ins were identified as the year in which the athletes participate in their first competition in the period of observation. Drop-out, on the other hand, was defined as the year following the last race observed. An alternative definition of drop-ins and drop-outs, different from the one used, can be derived by looking at club registrations over the years, which are available in the personal page of each athlete. The final number of

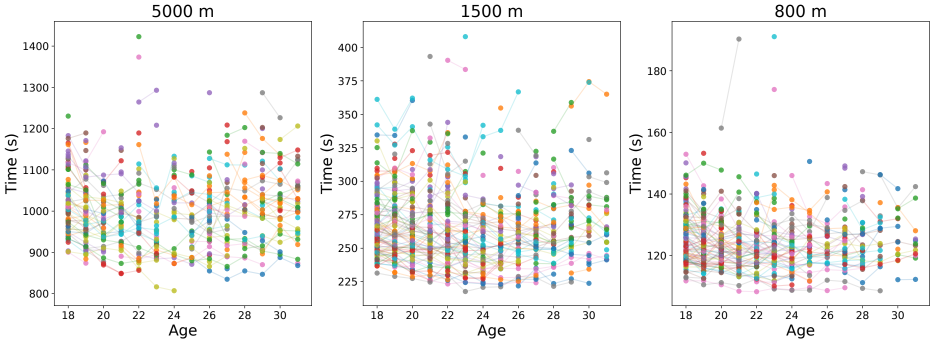

FIGURE 3.1: Trajectories observed in the 5000, 1500, 800 meters races of male Italian athletes born in the year 1988. Data were retrieved from the website of the Italian Athletics Federation (`www.fidal.it`).

athletes considered in the analysis is $Q = 369$ for $P = 3$ races and $T = 14$ years. The minimum number of observations per athlete is 2, the maximum 36, with mean 6.65. Minimum career length is 1 year, the maximum 14 years, with mean 5.04 years. In the sample, there is a strong presence of missing data, and there are differences in their nature. More specifically, a missing value occurs for athlete q in race p during year t if the athlete q does not end any official competition in race p during the considered year. The motivations for observing missing values can be multiple. Athletes can start and stop competing early in their careers, and thus data are observed only for the first years and are missing for the last ones. On the contrary, for athletes that start competing late in their life, data are observed only in the last years. Moreover, one athlete can compete only in a reduced number of races during one year (e.g. the athlete can compete only in 800 meters race and not compete in 1500 and 5000 meters races). This lack of data may be related, for example, to specific technical choices or personal attitudes of the athletes. Figure 3.2 shows 9 selected examples of missing data patterns that, potentially, represent 9 distinct classes considered in the analysis (see Section 3.3). As it can be seen, athletes compete in different races and in different years. In addition, career's history can have periods with no competitions held, although the athlete is still active.

3.2 The model

3.2.1 Clustering time series with matrix state space model

Imagine to observe the scalar element $y_{pq,t}$ that denotes the observation of race p for athlete q during the year t, for $p = 1, \dots, P$, $q = 1, \dots, Q$, and $t = 1, \dots, T$. Imagine also, in this phase of the explanation, that the complete set of observations is available in

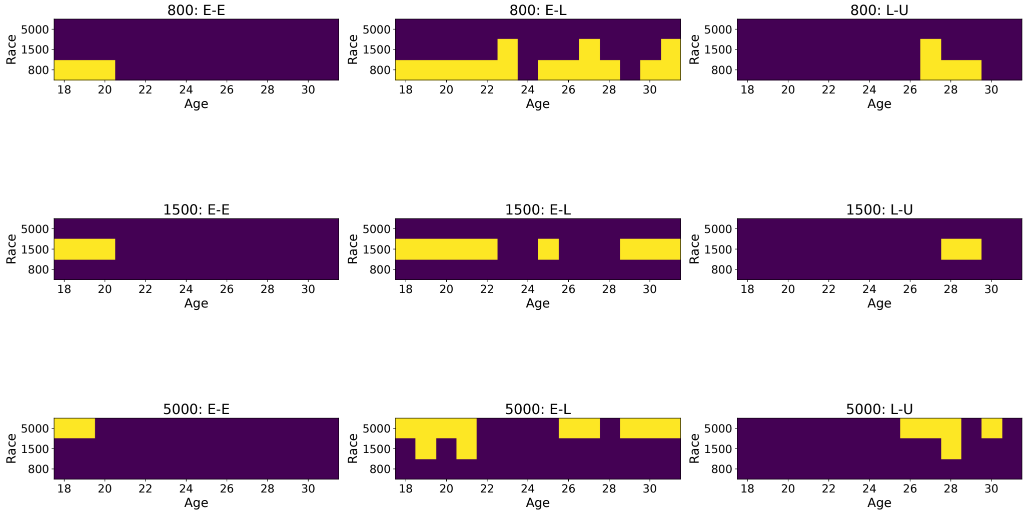

FIGURE 3.2: Examples of missing data patterns present in the data of 9 selected athletes. Yellow squares indicate presence of data. The acronyms "E-E", "E-L" and "L-U" describe different possible histories of the athletes. More specifically, they stand for "Early entry-Early exit", "Early entry-Late exit", "Early entry-Undefined exit" characteristics, respectively, and describe the entry and exit of the athletes in the sample. The numbers, i.e. 800, 1500, 5000, describe the reference race of each athlete. The matrices have been classified according to subjective knowledge on the phenomenon (see Section 3.3).

the sense that athletes participate in all P races during the years and that no drop-ins or drop-outs are observed. We assume that athletes are divided into G different unknown groups according to the evolutionary trajectories during their careers. Suppose now to know that athlete q belongs to group g. Its observations over time are then described by the following dynamic linear model

$$y_{pq,t} = \mathbf{z}_p^\top \boldsymbol{\alpha}_{p,t}^{(g)} + \varepsilon_{pq,t}, \tag{3.1}$$

$$\boldsymbol{\alpha}_{p,t+1}^{(g)} = \mathbf{T}_p \boldsymbol{\alpha}_{p,t}^{(g)} + \boldsymbol{\xi}_{p,t}^{(g)}, \tag{3.2}$$

in which $\boldsymbol{\alpha}_{p,1}^{(g)} \sim \mathrm{N}_{F_p}(\widehat{\boldsymbol{\alpha}}_{p,t}^{(g)}, \mathbf{P}_{p,1|0}^{(g)})$, for $p = 1, \ldots, P$, $t = 1, \ldots, T$. In the above specification, the row vector $\mathbf{z}_p^\top$, which is characterized by a known structure, links the observation $y_{pq,t}$ to the column vector $\boldsymbol{\alpha}_{p,t}^{(g)}$, which describes the group-specific dynamics of the p–race for all the athletes that belong to group g. These dynamics are determined by the state transition equation, that describes a first-order autoregressive process with transition matrix $\mathbf{T}_p$, which is race-specific, known, and shared across all the groups. In this way, for a generic race p, we require that the latent states of the different groups are different from each other, but are characterized by the same Markovian dependence induced by $\mathbf{T}_p$. Moreover, this dependence is not required to be common across different races, as $\mathbf{T}_p$ may differ from $\mathbf{T}_{p'}$ for any $p \neq p'$. The error terms $\varepsilon_{pq,1}, \ldots, \varepsilon_{pq,T}$ are not specified, but are assumed to be Gaussian with variance $\sigma^2_{pp,qq}$, assumed to be race- and subject-specific. They are assumed to be serially independent and independent of both the states $\boldsymbol{\alpha}^{(g)}_{p,1}, \dots, \boldsymbol{\alpha}^{(g)}_{p,T}$ and the disturbances $\boldsymbol{\xi}^{(g)}_{p,1}, \dots, \boldsymbol{\xi}^{(g)}_{p,T}$, for $p = 1, \dots, P$ and $g = 1, \dots, G$.

Let $\mathbf{y}_{\cdot q,t} = (y_{1q,t}, \dots, y_{Pq,t})^\top$, $\boldsymbol{\alpha}^{(g)}_t = (\boldsymbol{\alpha}^{(g)\top}_{1,t}, \dots, \boldsymbol{\alpha}^{(g)\top}_{P,t})^\top$, $\boldsymbol{\varepsilon}_{\cdot q,t} = (\varepsilon_{1q,t}, \dots, \varepsilon_{Pq,t})^\top$, $\boldsymbol{\xi}^{(g)}_t = (\boldsymbol{\xi}^{(g)\top}_{1,t}, \dots, \boldsymbol{\xi}^{(g)\top}_{P,t})^\top$, $\mathbf{Z} = \text{blkdiag}(\mathbf{z}_1^\top, \dots, \mathbf{z}_P^\top)$, and $\mathbf{T} = \text{blkdiag}(\mathbf{T}_1, \dots, \mathbf{T}_P)$. Equations (3.1) and (3.2) can alternatively be expressed in a vector form, which is

$$\mathbf{y}_{\cdot q,t} = \mathbf{Z}\boldsymbol{\alpha}^{(g)}_t + \boldsymbol{\varepsilon}_{\cdot q,t}, \tag{3.3}$$

$$\boldsymbol{\alpha}^{(g)}_{t+1} = \mathbf{T}\boldsymbol{\alpha}^{(g)}_t + \boldsymbol{\xi}^{(g)}_t, \tag{3.4}$$

for $\boldsymbol{\alpha}^{(g)}_1 \sim \text{N}_F(\widehat{\boldsymbol{\alpha}}^{(g)}_{1|0}, \mathbf{P}^{(g)}_{1|0})$, where $F = \sum_{p=1}^P F_p$ denotes the total number of states of each group. Although these equations represent nothing more than the vector formulation of the scalar one, they allow to introduce the covariance of the errors $\boldsymbol{\Sigma}_q$ and disturbances $\boldsymbol{\Psi}_g$. If full, they capture contemporaneous correlation between errors and the disturbances associated with different races. For example, it is possible to think that an improvement in the performance of one race can be reflected in the improvement of the performance of the others. However, this improvement can be temporary and associated with the error term, or related to the trajectory that describes the athlete's career. It is important to note, however, that $\boldsymbol{\Sigma}_q$ is subject- and $\boldsymbol{\Psi}_g$ is group-specific. One typical restriction sets $\boldsymbol{\Sigma}_q = \sigma^2_{qq}\boldsymbol{\Sigma}^R$ and $\boldsymbol{\Psi}_g = \psi^2_{gg}\boldsymbol{\Psi}^R$, where σ^2_{qq} and ψ^2_{gg} denote the q-th and the g-th diagonal element of $\boldsymbol{\Sigma}^C$ and $\boldsymbol{\Psi}^C$, respectively. In this way, different athletes and different groups would share the same correlation structure of the errors and disturbances up to subject- and group-specific proportionality constants.

To introduce a third viewpoint useful to interpret the model, we define the following matrices:

$$\mathbf{Y}_t = \begin{bmatrix}\mathbf{y}_{\cdot 1,t} & \dots & \mathbf{y}_{\cdot Q,t}\end{bmatrix}, \quad \mathbf{A}_t = \begin{bmatrix}\boldsymbol{\alpha}^{(1)}_t & \dots & \boldsymbol{\alpha}^{(G)}_t\end{bmatrix}, \quad \mathbf{S}^\top = \begin{bmatrix}\mathbf{s}^\top_{1\cdot} & \dots & \mathbf{s}^\top_{Q\cdot}\end{bmatrix},$$

$$\mathbf{E}_t = \begin{bmatrix}\boldsymbol{\varepsilon}_{\cdot 1,t} & \dots & \boldsymbol{\varepsilon}_{\cdot Q,t}\end{bmatrix}, \quad \boldsymbol{\Xi}_t = \begin{bmatrix}\boldsymbol{\xi}^{(1)}_t & \dots & \boldsymbol{\xi}^{(G)}_t\end{bmatrix},$$

where $\mathbf{s}^\top_{q\cdot} = (\text{I}(S_q = 1), \dots, \text{I}(S_q = G))^\top$ is an allocation vector, known in this specific phase of the model's specification, but object of inference in the later sections. This vector highlights the group to which the athlete q belongs, in such a way $\text{I}(S_q = g) = 1$ if athlete q belongs to group g, and 0 otherwise. The matrix-variate representation of the model

$$\mathbf{Y}_t = \mathbf{Z}\mathbf{A}_t\mathbf{S}^\top + \mathbf{E}_t, \tag{3.5}$$

$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t\mathbf{U}^\top + \boldsymbol{\Xi}_t, \tag{3.6}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{F,G}(\widehat{\mathbf{A}}_{1|0}, \mathbf{P}_{1|0})$, and $\mathbf{U} = \mathbf{I}_G$, is a special case of the matrix state space model described in the previous chapter. In this work, also $\widehat{\mathbf{A}}_{1|0}$ is assumed to be a matrix-variate normal random variable with known mean and diagonal covariance matrix $\mathbf{P}_{1|0}$. The matrix $\mathbf{S}$ is a selection matrix, with the role of selecting, for each athlete, the columns of states associated with the group the athlete belongs to, and silencing the others. The matrices of errors and disturbances are assumed to follow matrix-normal distribution with covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$, respectively. In matrix state space models, one typical assumption imposes $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}^C \otimes \boldsymbol{\Psi}^R$ (see, e.g., Wang and West, 2009; Chen *et al.*, 2020). Here, $\boldsymbol{\Sigma}^R$ and $\boldsymbol{\Psi}^R$ are row-covariance matrices with dimensions $P \times P$ and $F \times F$, and measure row-wise dependence of errors and disturbances, respectively. Conversely, the matrices $\boldsymbol{\Sigma}^C$ and $\boldsymbol{\Psi}^C$ are column-covariance matrices with dimensions $Q \times Q$ and $G \times G$ that measure instead column-wise dependence of errors and disturbances, respectively. Dependent rows or columns are characterized by full covariance matrices, while independent row or columns are characterized by diagonal matrices (Gupta and Nagar, 2000). Thus, the model is general enough that various forms of dependence can be considered on the basis of different possible specifications of the covariance matrices. However, with annual-based data describing the careers of different athletes, we require $\mathbf{Z} = \mathbf{I}_P$, $\mathbf{T} = \mathbf{I}_P$, $\boldsymbol{\Sigma} = \mathbf{I}_Q \otimes \boldsymbol{\Sigma}^R$ and $\boldsymbol{\Psi} = \mathrm{blkdiag}(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_G)$, since decomposing $\boldsymbol{\Psi}$ with a Kronecker product may be too restrictive. We note that this request solves the identifiability problem such that, for any $c > 0$ and $d > 0$, $\boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R = c\boldsymbol{\Sigma}^C \otimes \frac{1}{c}\boldsymbol{\Sigma}^R$ and $\boldsymbol{\Psi}^C \otimes \boldsymbol{\Psi}^R = d\boldsymbol{\Psi}^C \otimes \frac{1}{d}\boldsymbol{\Psi}^R$. Imposing $\boldsymbol{\Sigma}^C = \mathbf{I}_Q$ is a restriction even stronger than required, but it helps to stabilize the estimation of the other components of the model given the presence of many missing data. However, we also note that these restrictions do not solve another problem of non-identifiability present in the model, known in the literature as label switching. This problem happens when the posterior distribution presents multiple equal maxima, that correspond to different ways of swapping the columns of $\mathbf{S}$ in the presence of symmetry between the priors (see, e.g., Malsiner-Walli *et al.*, 2017). This issue is also related to the priors' specification and can be solved in the post processing phase, so we leave the related discussion to later sections.

Moreover, we make a distinction among $\mathcal{Y}$, which denotes the set of observations as if they were fully observed; $\mathcal{Y}^\star$, corresponding to the set of variables which are effectively observed; and $\widetilde{\mathcal{Y}}$ which corresponds to the completion of $\mathcal{Y}^\star$, i.e. such that $\mathcal{Y} = \mathcal{Y}^\star \cup \widetilde{\mathcal{Y}}$ and $\mathcal{Y}^\star \cap \widetilde{\mathcal{Y}} = \emptyset$. Beyond, we denote with $\mathcal{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_T\}$ the set storing the latent states of the state space model. These distinctions will be used in Section 3.4, since they are useful in the estimation of some components of the model.

3.2.2 Missing data inform on clustering structure

The previous section was developed conditional on all data being observed, i.e., that the athletes run all P races during the years and that drop-ins and drop-outs are not observed. However, this is not the case for data that describe the career trajectories of athletes, since the lack of data is part of the career itself. To include these factors as informative aspect of athletes' career, we consider two other variables in the model. As first, we consider

$$d_{pq,t} = \begin{cases} 1 & \text{if race } p \text{ for athlete } q \text{ is observed at time } t, \\ 0 & \text{otherwise,} \end{cases}$$

to describe the presence or absence of the observed races for the athletes. Then we consider the variable $d^\star_{q,t}$ that informs whether the athlete q is in career during year t, which is

$$d^\star_{q,t} = \begin{cases} 0 & \text{if athlete } q \text{ has never started the career before } t \text{ (included)}, \\ 1 & \text{if athlete } q \text{ is in career during } t, \\ 2 & \text{if athlete } q \text{ has finished the career in } t \text{ (included)}. \end{cases}$$

The variable $d^\star_{q,t}$ is not decreasing in t, and describes the three possible states of athlete's career. Moreover, if $d^\star_{q,t} \in \{0, 2\}$, then $d_{pq,t} = 0$ with probability 1, for $p = 1, \dots, P$, meaning that no races are observed since the athlete is not competing. On the contrary, there might be athletes such that $d_{pq,t} = 0$, for $p = 1, \dots, P$, even if $d^\star_{q,t} = 1$. This is typical of athletes who, despite being in a career, decide not to compete during one specific year, but compete in the following years.

The division into three non-concurrent states allows for the introduction of temporal dynamics within the model of missing data patterns in an easy way. In particular, let $\mathbf{d}^\star_q = (d^\star_{q,1}, \dots, d^\star_{q,T})$, $\mathbf{d}_{\cdot q,t} = (d_{1q,t}, \dots, d_{Pq,t})^\top$, and $\mathbf{D}_q = \begin{bmatrix} \mathbf{d}_{\cdot q,1} & \dots & \mathbf{d}_{\cdot q,T} \end{bmatrix}$, $\mathcal{D} = \{\mathbf{D}_1, \dots, \mathbf{D}_Q\}$, and $\mathcal{D}^\star = \{\mathbf{d}^\star_1, \dots, \mathbf{d}^\star_Q\}$. First, we make the following independence assumption among different subjects

$$p_{\boldsymbol{\theta}}(\mathcal{D}, \mathcal{D}^{\star}|\mathbf{S}) = \prod_{q=1}^{Q} p_{\boldsymbol{\theta}}(\mathbf{D}_q, \mathbf{d}_q^{\star}|S_q). \tag{3.7}$$

As a second step, we let $\mathbf{d}_q^{\star}$ and $\mathbf{D}_q$ be dependent on the group S_q to which the athlete q belongs, and make the following conditional independence assumption

$$\begin{aligned} p_{\boldsymbol{\theta}}(\mathbf{D}_q, \mathbf{d}_q^{\star}|S_q) &= p_{\boldsymbol{\theta}}(\mathbf{D}_q|\mathbf{d}_q^{\star}, S_q) p_{\boldsymbol{\theta}}(\mathbf{d}_q^{\star}|S_q) \\ &= \prod_{t=1}^{T} \Big[\prod_{p=1}^{P} p_{\boldsymbol{\theta}}(d_{pq,t}|d_{q,t}^{\star}, S_q) \Big] p_{\boldsymbol{\theta}}(d_{q,t}^{\star}|d_{q,t-1}^{\star}, S_q) \end{aligned} \tag{3.8}$$

where $p_{\boldsymbol{\theta}}(d_{q,1}^{\star} = 1|d_{q,0}^{\star}, S_q = g) = \pi_{1g}^{\star}$ and $p_{\boldsymbol{\theta}}(d_{q,1}^{\star} = 0|d_{q,0}^{\star}, S_q = g) = 1 - \pi_{1g}^{\star}$, with $d_{q,0}^{\star} = 0$ fixed for $q = 1, \ldots, Q$.

Note that, in the equations, the subscript $\boldsymbol{\theta}$ in $p_{\boldsymbol{\theta}}(A|B)$ denotes conditional dependence of the form $p(A|B, \boldsymbol{\theta})$, for slight abuse of notation. Table 3.1 provides details of the probabilities associated with the possible observable case histories. We note that, for athlete q, the conditional probabilities at time t of transition from state 0 to state 1—i.e. $p_{\boldsymbol{\theta}}(d_{q,t}^{\star} = 1|d_{q,t-1}^{\star} = 0, S_q = g) = \pi_{1g}^{\star}$— or from state 1 to state 2—i.e. $p_{\boldsymbol{\theta}}(d_{q,t}^{\star} = 2|d_{q,t-1}^{\star} = 1, S_q = g) = \pi_{2g}^{\star}$— are group dependent but constant over time. Similarly, for athlete q, the conditional probabilities at time t of observing a value for the generic race p—i.e. $p_{\boldsymbol{\theta}}(d_{pq,t} = 1|d_{q,t}^{\star} = 1, S_q = g) = \pi_{pg}$— are group-dependent, but fixed over time. We believe that both assumptions are plausible for the phenomenon under consideration. In particular, although transitions in the prevalence of the type of races done in a long career are possible for some athletes (for example, from shorter to longer races), these transitions are difficult to detect with annual based data—which are summaries of the entire years. This issue is due by the fact it is enough to compete in only one race in the distance to be included into the ranking lists. Similarly, the assumption of constant probabilities during years used to describe the presence of missing values does not contemplate the possibility that athletes would get seriously injured, and, thus, they would not compete in any race for more than a year. Although there is no clear indication in the literature about the average duration and severity of an injury in middle distance athletes (see, e.g., van Gent *et al.*, 2007), we assume here that severe injuries are present only in low proportions, leaving open possible investigations on this aspect in the future.

Table 3.1: Case histories' probabilities describing possible missing data patterns.

Quantity	Conditioning variable	Conditioned variable	Probability
$p_{\boldsymbol{\theta}}(d^\star_{q,t} \| d^\star_{q,t-1}, S_q = g)$	0	0	$1 - \pi^\star_{1g}$
		1	$\pi^\star_{1g}$
		2	0
	1	0	0
		1	$1 - \pi^\star_{2g}$
		2	$\pi^\star_{2g}$
	2	0	0
		1	0
		2	1
$p_{\boldsymbol{\theta}}(d_{pq,t} \| d^\star_{q,t}, S_q = g)$	0	0	1
		1	0
	1	0	$1 - \pi_{pg}$
		1	π_{pg}
	2	0	1
		1	0

3.3 Likelihood and prior specification

3.3.1 Likelihood

In order to derive the posterior distribution of the parameters, we present the likelihood of the observed process first, augmented for both the states $\mathcal{A}$, the missing observations $\widetilde{\mathcal{Y}}$, and $\mathbf{S}$. The augmented likelihood is characterized by the following conditional independence structure

$$p_{\boldsymbol{\theta}}(\mathcal{Y}, \mathcal{D}, \mathcal{D}^\star, \mathcal{A}, \mathbf{S}) = p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{A}, \mathbf{S}) p_{\boldsymbol{\theta}}(\mathcal{D}|\mathcal{D}^\star, \mathbf{S}) p_{\boldsymbol{\theta}}(\mathcal{D}^\star|\mathbf{S}) p_{\boldsymbol{\theta}}(\mathbf{S}) p_{\boldsymbol{\theta}}(\mathcal{A}). \tag{3.9}$$

In the equation, $p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{D}^\star, \mathcal{A}, \mathbf{S}) = p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{A}, \mathbf{S})$, and is determined by the measurement Equation (3.5), for which all observations are assumed to be available, and the prior on $\mathcal{A}$ is implicitly determined by the form of the state equation of the state space in Equation (3.6). However, only $\mathcal{Y}^\star = \{\mathcal{Y}^\star_1, \ldots, \mathcal{Y}^\star_T\}$ is observed, but $p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{A}, \mathbf{S})$ can be obtained by conditioning, noting that

$$p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{A}, \mathbf{S}) = p_{\boldsymbol{\theta}}(\mathcal{Y}^\star|\mathcal{D}, \mathcal{A}, \mathbf{S}) p_{\boldsymbol{\theta}}(\widetilde{\mathcal{E}}|\mathcal{Y}^\star, \mathcal{D}, \mathbf{S}),$$

where $\widetilde{\mathcal{E}}$ stores all those entries in $\mathcal{E} = \{\mathbf{E}_1, \ldots, \mathbf{E}_T\}$ associated with the observed missing values. To characterize $\mathbf{S}$, we make the following independence assumption

$$p_{\boldsymbol{\theta}}(\mathbf{S}) = \prod_{q=1}^{Q} p_{\boldsymbol{\theta}}(\mathbf{s}_{q\cdot}) = \prod_{q=1}^{Q}\prod_{g=1}^{G} \pi_g^{\mathrm{I}(S_q=g)}, \tag{3.10}$$

where $\boldsymbol{\pi} = (\pi_g, \ldots, \pi_G)$ is such that $\pi_g \in (0,1)$, for $g = 1, \ldots, G$, and $\sum_{g=1}^{G} \pi_g = 1$.

3.3.2 Prior specification

We assume that the parameter $\boldsymbol{\theta}$ factorizes as follow

$$p(\boldsymbol{\theta}) = p(\hat{\mathbf{A}}_{1|0})p(\boldsymbol{\Sigma}^R)p(\boldsymbol{\pi})p(\boldsymbol{\Psi})p(\boldsymbol{\pi}_1^\star)p(\boldsymbol{\pi}_2^\star)\prod_{p=1}^{P} p(\boldsymbol{\pi}_p), \tag{3.11}$$

where $\boldsymbol{\pi}_1^\star = (\pi_{11}^\star, \ldots, \pi_{1G}^\star)$, $\boldsymbol{\pi}_2^\star = (\pi_{21}^\star, \ldots, \pi_{2G}^\star)$, and $\boldsymbol{\pi}_p = (\pi_{p1}, \ldots, \pi_{pG})$ are vectors that store all the missing data probabilities. It is interesting to observe that their dimensions depend on the number of groups G, which is fixed. In this work, we fix $G = 9$, since the number of different classes is obtained by combining the factors personal attitude and history, as summarized by the following distinct classes

$$\text{Attitude} = \begin{cases} 800 \text{ m}, \\ 1500 \text{ m}, \\ 5000 \text{ m}, \end{cases} \qquad \text{History} = \begin{cases} \text{E-E: Early entry/Early exit}, \\ \text{E-L: Early entry/Late exit}, \\ \text{L-U: Late entry/Undefined exit}. \end{cases}$$

In order to include in the model prior information about these classes, the priors on π_{pg}, $\pi_{1g}^\star$, and $\pi_{2g}^\star$ are assumed to be informative Beta distributions, with the mean shifted toward one or zero depending on the meaning of the parameter, as shown in Figure 3.3. In substance, a group with athletes specialized in race is characterized by an high prior probability in the reference race, and lower prior probabilities in the others. A group with "Early entry" characteristic has high probability of dropping-in. On the contrary, a group with "Late entry" characteristic has low probability of dropping-in. Similarly, a group with "Early exit" characteristic has an high prior drop-out probability, differently from those with "Late exit" characteristic which have low prior probability of dropping-out, or "Undefined exit" for which we adopt a diffuse prior. Since no prior information is available on groups' proportions, $\pi \sim \mathrm{Dir}_G(1/G, \ldots, 1/G)$. Covariance matrices are assumed to be diffuse inverse Wishart, which are conjugate under Gaussian likelihood.

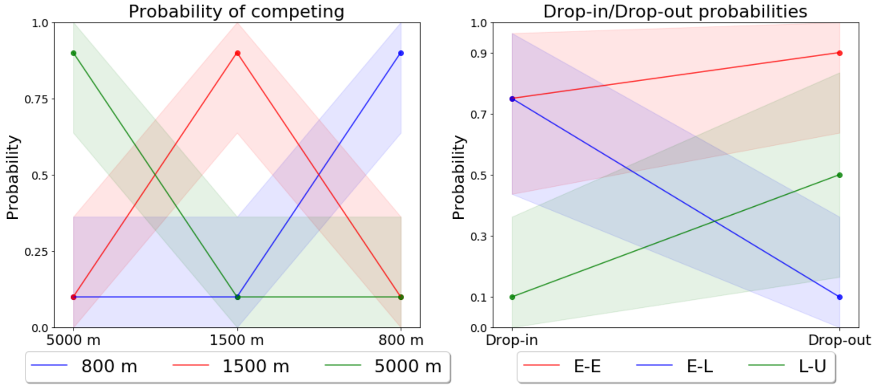

FIGURE 3.3: Prior probabilities describing missing data patterns. On the left, priors π_{pg} are represented on the basis of the athlete's attitude. On the right, priors $\pi^{\star}_{1g}$ (drop-in) and $\pi^{\star}_{2g}$ (drop-out) are represented considering the athlete's history. Points denote the means of the distributions. Colored bands describe 90% pointwise prior credible interval based on quantiles. The acronyms "E-E", "E-L" and "L-U" describe different possible histories of the athletes. More specifically, they stand for "Early entry-Early exit", "Early entry-Late exit", "Early entry-Undefined exit" characteristics, respectively.

More specifically, we assume $\mathbf{\Sigma}^R \sim \mathrm{IW}_3(\nu_\sigma^R, \mathbf{\Sigma}_0^R)$ for the errors and require

$$\mathbf{\Psi}_1 = \mathbf{\Psi}_2 = \mathbf{\Psi}_3 = \mathbf{\Psi}_{800}, \quad \mathbf{\Psi}_4 = \mathbf{\Psi}_5 = \mathbf{\Psi}_6 = \mathbf{\Psi}_{1500},$$
$$\mathbf{\Psi}_7 = \mathbf{\Psi}_8 = \mathbf{\Psi}_9 = \mathbf{\Psi}_{5000},$$

where $\mathbf{\Psi}_j \sim \mathrm{IW}_3(\nu_j, \mathbf{\Psi}_j^0)$, for $j \in \{800, 1500, 5000\}$. The constraints impose that athletes with the same attitude are characterized by the same covariance matrices for the disturbances, meaning that they share the same correlation structure regardless of their histories. This assumption allows for improved covariance estimates particularly for those groups characterized by a lot of missing data (E-E, L-U) by reducing the number of distinct parameters involved in these covariance matrices. If relaxed, in fact, the number of distinct parameters is $9 \cdot (3 \cdot 4/2) = 54$ with respect to $3 \cdot (3 \cdot 4/2) = 18$. Finally, the prior on $\hat{\mathbf{A}}_{1|0}$ is assumed matrix-normal with mean $\bar{\mathbf{y}}_1 \mathbf{1}_G^\top$ and variance $\mathbf{P}_{1|0} = \mathbf{I}_G \otimes \mathrm{diag}(\mathrm{p}_{1,1}^2, \ldots, \mathrm{p}_{1,P}^2)$, where $\bar{\mathbf{y}}_1$ is the vector storing sample average of observed races at first time instant, and $\mathrm{p}_{1,p}^2$ is twice the sample variance of the p–th observed race at the first time instant. Alternative specifications are possible, including diffuse and exact initialization (Durbin and Koopman, 2012).

3.3.3 Two step interpretation of clustering strategy

The goal of our inference procedure is to obtain a sample from the posterior distribution

$$p(\boldsymbol{\theta}, \mathcal{A}, \mathrm{S}, \widetilde{\mathcal{E}}|\mathcal{Y}^\star, \mathcal{D}, \mathcal{D}^\star) \propto p(\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{A}, \mathrm{S})p_{\boldsymbol{\theta}}(\mathcal{D}|\mathcal{D}^\star, \mathrm{S})p_{\boldsymbol{\theta}}(\mathcal{D}^\star|\mathrm{S})p_{\boldsymbol{\theta}}(\mathrm{S})p_{\boldsymbol{\theta}}(\mathcal{A}), \quad (3.12)$$

in order to derive a related posterior distribution of some quantity of interest. It is worth to mention that the posterior $\mathbb{Q}_{1,2}(\mathbf{S}) = p(\boldsymbol{\theta}, \mathcal{A}, \mathrm{S}, \widetilde{\mathcal{E}}|\mathcal{Y}^\star, \mathcal{D}, \mathcal{D}^\star)$ can be expressed as $\mathbb{Q}_{1,2}(\mathbf{S}) = \mathbb{Q}_{2|1}(\mathbf{S})\mathbb{Q}_1(\mathbf{S})$, where

$$\mathbb{Q}_1(\mathbf{S}) = p_{\boldsymbol{\theta}}(\mathcal{D}|\mathcal{D}^\star, \mathrm{S})p_{\boldsymbol{\theta}}(\mathcal{D}^\star|\mathrm{S})p_{\boldsymbol{\theta}}(\mathrm{S})p(\boldsymbol{\pi})p(\boldsymbol{\pi}_1^\star)p(\boldsymbol{\pi}_2^\star)\prod_{p=1}^{P} p(\boldsymbol{\pi}_p),$$

$$\mathbb{Q}_{2|1}(\mathbf{S}) = p(\hat{\mathbf{A}}_{1|0})p_{\boldsymbol{\theta}}(\mathcal{A})p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{A}, \mathrm{S})p(\boldsymbol{\Sigma}^R)p(\boldsymbol{\Psi}).$$

We can interpret our clustering strategy as a two-step procedure, in which different athletes are clustered by means of the posterior $\mathbb{Q}_1(\mathbf{S})$ first, and then $\mathbb{Q}_1(\mathbf{S})$ is used as a new prior on $\mathbf{S}$ for obtaining a new clustering of athletes, given by $\mathbb{Q}_{1,2}(\mathbf{S})$, in light of their performances. The posterior $\mathbb{Q}_1(\mathbf{S})$ reflects the grouping structure present in the data according to the missing data patterns. The prior $p(\boldsymbol{\pi}_1^\star)p(\boldsymbol{\pi}_2^\star)\prod_{p=1}^{P} p(\boldsymbol{\pi}_p)$ reflects prior beliefs about missing data patterns' dynamics for $G = 9$ distinct groups. The distinction into 9 groups derives from qualitative reasoning about the phenomenon under consideration, on the basis of a prior knowledge present in the application domain. In this study, the 9 groups were obtained by combining two distinct factors, personal attitude and personal history, which are considered relevant in the determination of performances. Differences across groups are accounted using $p(\boldsymbol{\pi}_1^\star)p(\boldsymbol{\pi}_2^\star)\prod_{p=1}^{P} p(\boldsymbol{\pi}_p)$, that introduces asymmetries across groups and leads to an asymmetrical posterior. This aspect helps in the identification of the groups and solves the problem of label switching—which is present when both likelihood and prior are symmetric (see Malsiner-Walli *et al.*, 2017, among others)—by weighing in different ways the multiple modes of the posterior.

The information derived from the first step is twofold. On one side, the clustering of athletes and uncertainty quantification can be obtained on the basis of their attitudes and histories only. For example, by considering $\mathbb{Q}_1(\mathbf{S})$, it is possible to obtain a posterior summary $\widetilde{\mathbf{S}}$ of $\mathbf{S}$ that can be used as an exploratory tool for computing $\mathbb{Q}_{2|1}(\widetilde{\mathbf{S}})$, where $\mathbf{S} = \widetilde{\mathbf{S}}$ is considered as known. On the other side, it is possible to update the knowledge about missing data probabilities and group proportions. We note that, while the prior specification reflects a belief about the missing data behaviors, it is possible that the posterior no longer reflects these prior beliefs about the groups.

However, the first step does not provide information about groups' performances over time described by the states $\mathcal{A}$, which is of primary interest to understand how athletes' careers develop over time. Obtaining a sample of $\mathbb{Q}_{1,2}(\mathbf{S})$ allows to obtain this information, accounting also for uncertainty present in $\mathbb{Q}_1(\mathbf{S})$. As a side benefit, we update knowledge on both $\boldsymbol{\theta}$ and $\mathbf{S}$, given the whole set of observations (and missing data patterns), that can be used for posterior analysis. For example, it is possible to obtain a posterior summary $\widehat{\mathbf{S}}$ of $\mathbf{S}$ and compare it with the $\widetilde{\mathbf{S}}$ previously obtained. Mismatches between $\widehat{\mathbf{S}}$ and $\widetilde{\mathbf{S}}$ identify athletes that belong to one group which is based on personal history and attitude only, but, in light of the observations of the races, have performances related to another one.

3.4 Inference via Gibbs sampling

Samples from the posterior distribution can be obtained with a Markov Chain Monte Carlo (MCMC) approach, a standard procedure used in Bayesian analysis (see, e.g., Gelman *et al.*, 2014; Robert and Casella, 2004). The use of conjugate priors has made the derivation straightforward. States estimation can be obtained with a reduced form of the model using a simulation smoothing technique (Durbin and Koopman, 2002). In particular, let $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$. We can apply the reduction by transformation technique described in Section 2.4.3 to the vector form of the model by considering the following decomposition

$$(\mathbf{S} \otimes \mathbf{Z}) = (\mathbf{S}\mathbf{I}_G) \otimes (\mathbf{I}_P\mathbf{Z}) = (\mathbf{S} \otimes \mathbf{I}_P)(\mathbf{I}_G \otimes \mathbf{Z}),$$

where $\mathbf{S}$ and $\mathbf{Z}$ are assumed to be full-column and full-row rank matrices, respectively. If we consider $\mathbf{Z}_\dagger = (\mathbf{S} \otimes \mathbf{I}_P)$, and $\mathbf{A}' = \mathbf{Z}_\dagger^\top \boldsymbol{\Sigma}^{-1}$, the state estimation can be applied to the reduced vector of observations $\mathbf{y}_t' = \mathbf{A}'\mathbf{y}_t$, which has dimensions $PG \times 1$, where PG is typically such that $PG \ll PQ$, leading to larger speed-ups when $G \ll Q$. Further details and other useful steps of the algorithm are reported in Appendix A.

3.4.1 Possible improvements

The algorithm is very sensitive to starting points, and the risk is that during its iterations: (a) the algorithm gets stuck in some local mode; (b) some groups remains without athletes. To avoid the problem (a), it is possible to include in the algorithm a Metropolis Hastings step in which the columns of $\mathbf{S}$ are exchanged. If case (b) happens, obtaining samples from the states is problematic, especially if the model involves transition matrices $\mathbf{T}_p$ that imply the use of non-stationary processes (e.g. random walk). States are indeed simulated conditional on no observed data. The resulting processes are generally highly variable and far away from the observed data, leading to groups that remain without athletes for many iterations. This problem, which depends on how the model is specified, can be solved by considering several approaches.

The first strategy is to start the iterations with good starting points. For example, it is possible to initialize the matrix $\mathbf{S}$ with a posterior summary $\widetilde{\mathbf{S}}$ obtained from $\mathbb{Q}_1(\mathbf{S})$, and then obtain other starting quantities using $\mathbb{Q}_{1,2}(\widetilde{\mathbf{S}})$, where $\mathbf{S} = \widetilde{\mathbf{S}}$ is considered fixed and known. However, this strategy does not guarantee that, during the iterations, all the groups are filled with athletes.

The second solution is to adopt a strategy inspired by anchoring (Kunkel and Peruggia, 2020). In this case, it is necessary to select at least one athlete per group to be allocated with probability one to a reference group. This strategy requires knowledge of the groups searched and ensures that the groups do not become empty during the estimation routine by modifying the prior $p_{\boldsymbol{\theta}}(\mathbf{S})$ on cluster allocations according to selected constraints. This is a viable strategy, if one deals with data regarding middle distance athletes, since there is knowledge about the group to search. The choice of athletes to anchor for each group can be made prior to the analysis, on the basis of the observed missing data patterns or after obtaining a sample $\mathbb{Q}_1(\mathbf{S})$ in a preliminary step. Kunkel and Peruggia (2020) interpret the anchoring strategy as a data-depending informative prior on the mixture components. This proposal can be used together with or as an alternative to the subjective priors described in Section 3.3.

A third strategy, more general and useful in contexts different from clustering of athletes' careers, requires modifying the Gibbs sampling algorithm substantially. More specifically, it is possible to derive a collapsed Gibbs sampler (see, e.g., Liu, 1994) in which states and missing data probabilities are integrated out during the allocation step. States can be marginalized out through the use of Kalman filter routines. Probabilities associated with missing data patterns can be marginalized out using standard calculations with Bernoulli likelihoods and beta priors. In this way, the impact of simulated states in groups without observations can be reduced. This is useful, for example, when over-parametrized mixtures are used, where some groups are expected to remain empty (see, e.g., Malsiner-Walli *et al.*, 2017; Frühwirth-Schnatter *et al.*, 2020). This strategy leads to more intensive routines, where efficient exploration of the discrete space of selection matrices is required. The matrix $\mathbf{S}$ can be updated one row at a time, conditional on the others. Alternative methods update the entire $\mathbf{S}$ or proceeding by blocks (Nobile and Fearnside, 2007; Titsias and Yau, 2017; Zanella, 2020; Grathwohl *et al.*, 2021). In Section 3.5 we adopt the first strategy, leaving the investigation of the others to future research.

3.5 Case study

The analysis were performed by obtaining two samples of size IT $= 2000$ from the posterior distributions $\mathbb{Q}_1(\mathbf{S})$ and $\mathbb{Q}_{1,2}(\mathbf{S})$. Burn-in periods of 8000 iterations were considered. $\mathbb{Q}_1(\mathbf{S})$ was first obtained, and $\widetilde{\mathbf{S}}$ was defined considering the MAP (maximum a posteriori) of cluster allocations, by considering each row of $\mathbf{S}$ separately. $\widetilde{\mathbf{S}}$ was considered fixed during the first 1000 iterations of the burn-in period of the algorithm for obtaining $\mathbb{Q}_{1,2}(\mathbf{S})$. During iterations, some groups remained without athletes, especially those with "Early entry/early exit" characteristic in all the attitudes. The algorithm demonstrated the ability to re-fill the groups despite the high variability present due to simulating states conditionally on no observations. However, this ability was lacking for the group with "Early entry/Early exit" characteristic and 5000 meters race attitude ("5000: E-E" group). This aspect may be related to the difficulty in identifying athletes with those characteristics, due to their actual low proportion in the sample. The actual elimination of this group typically requires a formal test, based on some information criteria (see, e.g., Spiegelhalter *et al.*, 2014) or alternative Bayesian methods for model selection (see, e.g., George, 2006). In this work, the number of groups is considered fixed as a result of qualitative reasoning deriving from domain knowledge of the application. So, the results obtained are valid conditionally on whether we believe the assumptions we have made. All aspects related to model validation, including the definition of alternative prior probabilities or the definition of alternative models in which the number of groups is unknown, are possible but left for future research (Frühwirth-Schnatter *et al.*, 2020).

Figure 3.4 shows the posterior distributions of the cluster allocation probabilities $\boldsymbol{\pi}$ derived from the obtained samples of $\mathbb{Q}_1(\mathbf{S})$ and $\mathbb{Q}_{1,2}(\mathbf{S})$. For both the samples, "5000: E-E" group's probability has a distribution shifted toward zero, in contrast to the posterior probability of "5000: E-L" and "5000: L-U" groups. This aspect may indicate a preference to not let young athletes compete in longer distances.

Figure 3.5 shows the posterior probabilities that describe athletes' attitudes and histories for the considered groups, obtained by sampling from the distribution $\mathbb{Q}_{1,2}(\mathbf{S})$. Identified groups are characterized by different posterior probabilities, which are updated with respect to the priors shown in Figure 3.3.

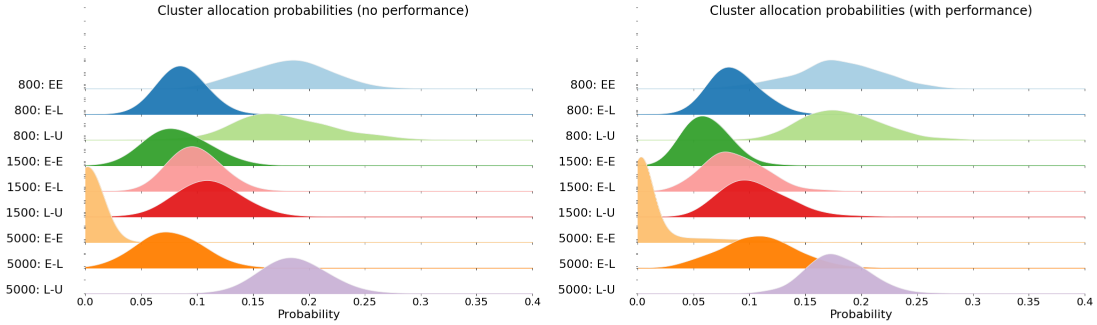

FIGURE 3.4: Posterior distributions of the cluster allocation probabilities $\boldsymbol{\pi}$. On the left, the distributions are derived from $\mathbb{Q}_1(\mathbf{S})$, without considering the performances of the athletes. On the right, the distributions are obtained from $\mathbb{Q}_{1,2}(\mathbf{S})$ considering athletes' performances over the years.

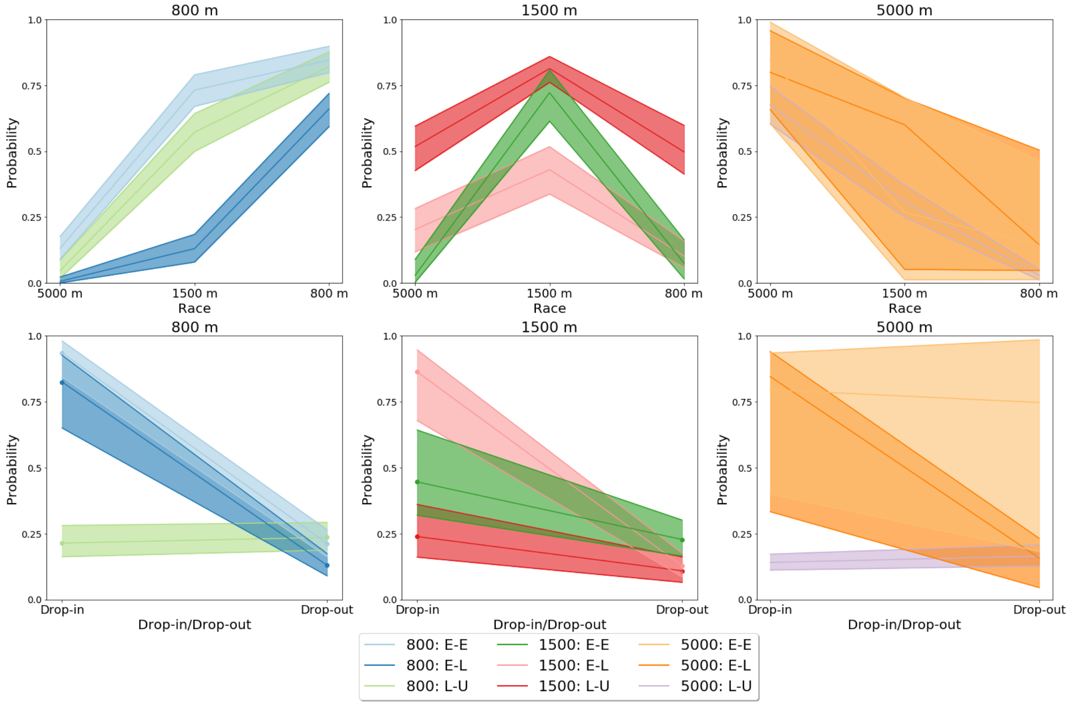

FIGURE 3.5: Posterior probabilities describing different athletes' attitude and history for the considered groups, obtained from the distribution $\mathbb{Q}_{1,2}(\mathbf{S})$. On the first row, the probabilities describing athletes' attitudes are shown. The second row shows the probabilities describing their histories. Groups were divided according race attitude. Colored bands denote 90% pointwise posterior credible intervals based on quantiles.

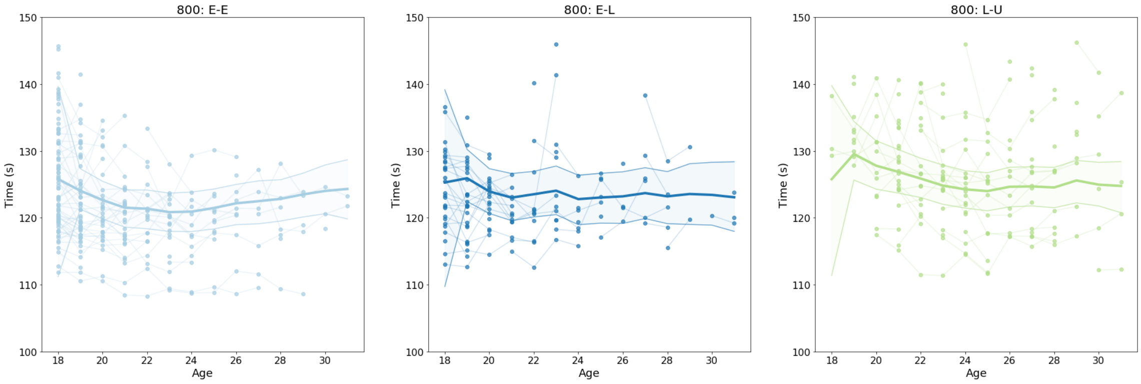

FIGURE 3.6: Performances on 800 meters race for groups “800: E-E”, “800: E-L”, and “800: L-U”. Thicker lines denote posterior medians of the states. Colored bands denote the respective 90% pointwise posterior credible intervals based on quantiles. Observed data are represented in the background, according to athletes’ MAP cluster allocations.

No groups with high drop-out probability were identified, except for group “5000: E-E” which is often empty, and whose posterior probabilities are not much different from the prior. In contrast, there are differences between groups in the probabilities describing athletes’ attitudes. For example, greater differences between “800: E-E” and “800: E-L” groups are present in the probabilities of their participation in the 1500 meters races, in comparison to the probabilities that describe their histories and other races. Indeed, “800: E-E” group (which is no longer characterized by an high drop-out probability) has a much higher probability of competing in 1500 races with respect to “800: E-L”. Similarly, “1500: E-L” group is characterized by a much lower probability of competing in all the races with respect to “1500: L-U” group. However, their histories are different since drop-in probability in “1500: L-U” group is much lower than the one in “1500: E-L” group. Interestingly, differences across drop-out probabilities of all groups are less evident, except for “5000: E-E” which is empty in mostly all iterations. This may indicate that drop-out, in comparison with the other factors, is a less discriminating one, but also that the model has difficulty in identifying groups with very short careers. This aspect, which is relevant from an inferential point of view, could be considered as a possible object of future investigation.

While studying the posterior distributions of parameters allows to characterize the groups found, understanding whether their characteristics are associated with better performances is a key aspect of the proposed approach. To do so, one strategy simply inspects posterior draws of the states. Figure 3.6, for example, shows performances on 800 meters race for groups “800: E-E”, “800: E-L”, and “800: L-U”. “800: E-E” and “800: E-L” are characterized by better performances, if compared with “800: L-U” group. This implies that, in athletes who compete more in the 800 meters race, an higher probability of drop-in appears to be associated with better performances. Differences between "800: E-E" and "800: E-L" are less visible. "800: E-E" group seems slightly better than "800: E-L", with improvements continuing until the age of 24. "800: E-L" is, on the contrary, characterized by performances which are nearly constant over time. On the basis of pure graphical comparison, it can be said that athletes who compete more in the 800 meters race improve over time and perform better if they start their career earlier in age and if they compete with an higher probability in the 1500 meters race (see Figure 3.5). Other comparisons are possible by considering other races and other groups. A complete view, however, would require comparing $27 = 9 \cdot 3$ different plots, which are reported in Appendix A.

Since, in the specified model, $\alpha_{p,t}^{(g)}$ describes the performance over time of group g in race p, the variable

$$\delta_{p,t}^{gg'} = \mathrm{I}(\alpha_{p,t}^{(g)} < \alpha_{p,t}^{(g')})$$

describes whether, during year t, group g is better than group g' in the considered race, for $g \neq g'$, since, in middle-distance races, a better performance is determined by a smaller time. An overall performance indicator can be obtained by considering

$$\Delta_p^{gg'} = \frac{1}{T} \sum_{t=1}^{T} \delta_{p,t}^{gg'} .$$

The performance indicator $\Delta_p^{gg'}$ indicates whether group g is better than g' on race p, as an average of $\delta_{p,t}^{gg'}$ over the entire period of observation. $\Delta_p^{gg'}$ is a discrete random variable with support $\{0, \frac{1}{T}, \dots, \frac{T-1}{T}, 1\}$. A value of $\Delta_p^{gg'}$ between 0.5 and 1 indicates that the performance of the g–th group is better than the performance of the g'–th one in the considered race for more than a half of the period of observation. The closer the value is to 1, the more were the times the performance was better. On the contrary, a value of $\Delta_p^{gg'}$ between 0 and 0.5 indicates an overall better performance of group g' with respect to the g–th one on the considered race, with a value near 0 that indicates stronger evidence on that. Thus, the variable $\Delta_p^{gg'}$ can be used for an overall comparison between groups, as an alternative to the graphical inspections previously explained. To do so, it is necessary to select a reference group to be used for comparison with the others. In this study, we use the groups "800: L-U", "1500: L-U", and "5000: L-U" for evaluating the 800, 1500, and 5000 meters races, respectively. Figure 3.7 shows the posterior distributions of $\Delta_p^{gg'}$ for the considered groups and races. Considering each graph marginally, the results suggest that groups "800: E-E" and "800: E-L" are

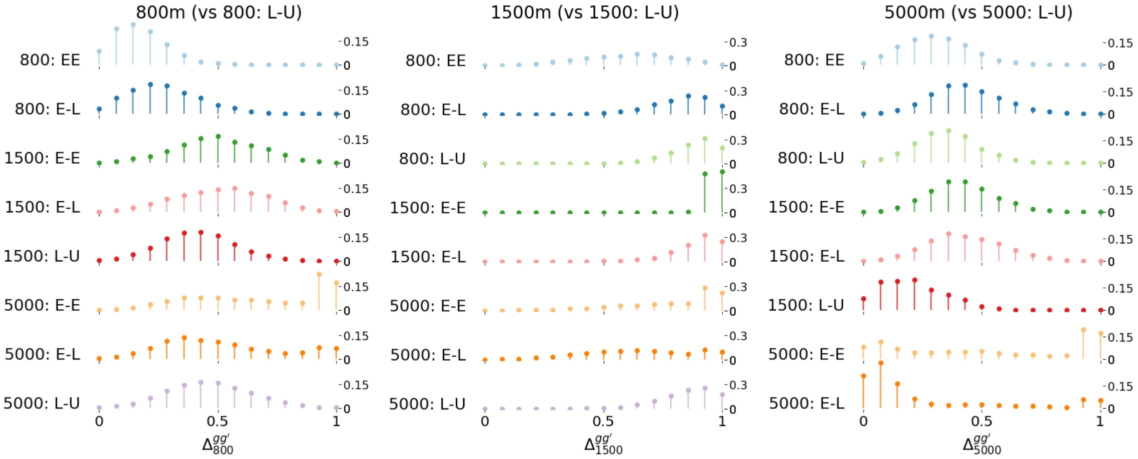

FIGURE 3.7: Posterior distributions of relative performance indicator $\Delta_p^{gg'}$ for 800, 1500, and 5000 meters races, where "800: L-U", "1500: L-U", and "5000: L-U" are used as reference groups, respectively.

overall better than the "800: L-U" one ($\Delta_{800}^{gg'} < 0.5$ with probability 0.98 and 0.88, respectively). On the contrary, there is no evidence for saying that group "800: L-U" is better than the others in 800 meters race. "1500: L-U" group seems to perform better than all other groups in 1500 meters race, if each distribution is considered marginally. Looking at the group characteristics in Figure 3.5, this result suggests that, in the 1500 meters race, better results are associated with an higher probability of competing not only in the reference race, but also in the others, despite the history of the group is characterized by a low probability of dropping-in. Finally, there is only slight evidence that the group "1500: E-L" is better than "1500: L-U". Note that multiple tests and comparisons were performed for evaluating hypothesis related to Figure 3.7. However, in order to validate these hypothesis, it is necessary to perform a joint test on them or possibly think about an appropriate correction. These further developments have to be considered for future research, together with algorithm improvements explained in Section 3.4.1.

3.6 Discussion

In this chapter, a model-based clustering method has been proposed for the analysis of multivariate time series. The clustering is achieved with a matrix state space model used to describe the athletes' performances over time. Different performances are linked to the observed values by means of a selection matrix involved in the measurement equation. The specific form given to the model permits the inclusion of various types of dependencies. Temporal dependence is considered through the state equation, which is general enough to include within it various models proposed in the literature of time series analysis (see, e.g., Durbin and Koopman, 2012). Cross-sectional dependence is included by means of the covariance matrices of errors and disturbances involved in both the measurement and the state equations. Since missing data patterns may be related to the observed performances of middle distance athletes, the presence or absence of data have been modeled using two other processes, describing athletes' attitudes and histories. Subjective priors were used to characterized 9 distinct profiles. A Gibbs sampling algorithm was derived. The real data application suggests that: (a) in the 800 meter race, late-entry into competition is associated with worse performances; (b) athletes who are more likely to participate in races other than their reference one have better overall performances. These results highlight the importance of starting careers at young ages and also that of versatility in competitions. A sample of athletes, regardless of their level, was used in the analyses. These results were derived from the analysis of the states, that have to be interpreted as an "average" behavior in the performances over time in the various groups. Therefore, these results are valid for common athletes, and not for specialized and high-level ones. However, in sports science interest is often directed toward this latter category. The analyses that are performed are typically the result of sample selection by researchers. An alternative approach in this context combines quantile regression methods (see, e.g., Koenker, 2005) with our proposal, that allows to consider different attitudes and histories of the athletes. Combining our approach with quantile regression allows to focus the attention on specific quantiles of the distributions, and therefore to identify the best performances (e.g., best 10%) based on all the observed data, accounting also the eventual presence of selection bias due to unobserved components (e.g., attitude and history). This should be considered as another possible future research goal, along with the other technical developments outlined throughout the chapter. U-shaped curves, which are typically adopted in studying performances of athletes (Haugen *et al.*, 2018), can be considered, in our model, by adopting the state space formulation of the regression model (Durbin and Koopman, 2012). This would simply require $\mathbf{Z} = \mathbf{Z}_t$ to be time dependent and $\mathbf{A}_t = \mathbf{A}$ to be constant over time. The approach adopted is more general and allows to capture various behaviors, as shown in Section 3.5. Finally, additional interesting aspects that were left out of this work, but that can be taken in consideration in future investigations, are: extensive simulation studies, analysis of female athletes, and analysis of other cohorts.

Chapter 4

Doubly-online changepoint detection for monitoring health status during sports activities

4.1 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, 2020b). 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 (2020c) 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, 2020a). Indeed, it is expected that in the near future, 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 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 *et al.* (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 4.1 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 the end of the activity. Figure 4.2

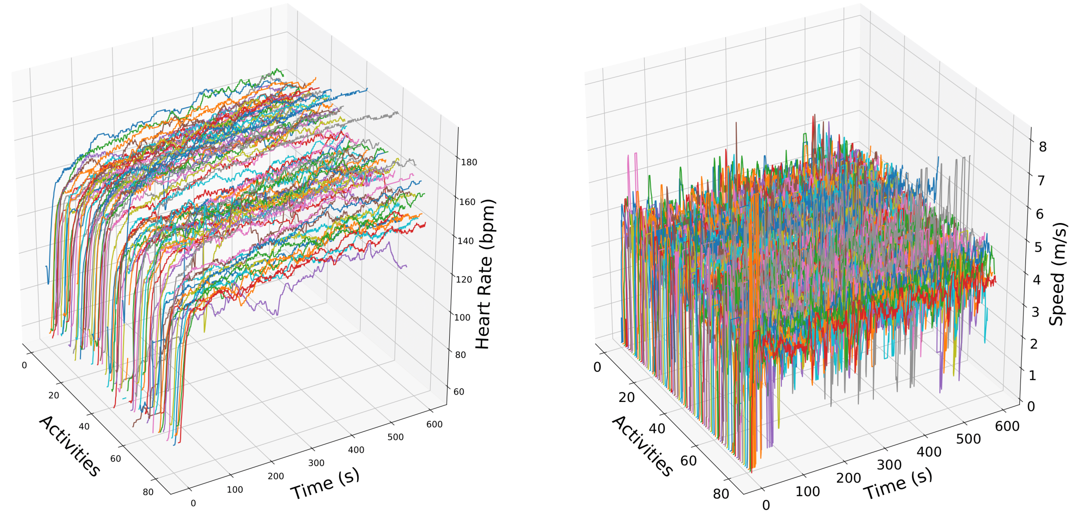

FIGURE 4.1: A sequence of activities performed by one athlete from our dataset.

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 4.2, the runner is interested in the behavior change at minute 5 of the current 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 *et al.* (2012); Yildirim *et al.* (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 *et al.*, 2020; Xie *et al.*, 2021; Haynes *et al.*, 2017).

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

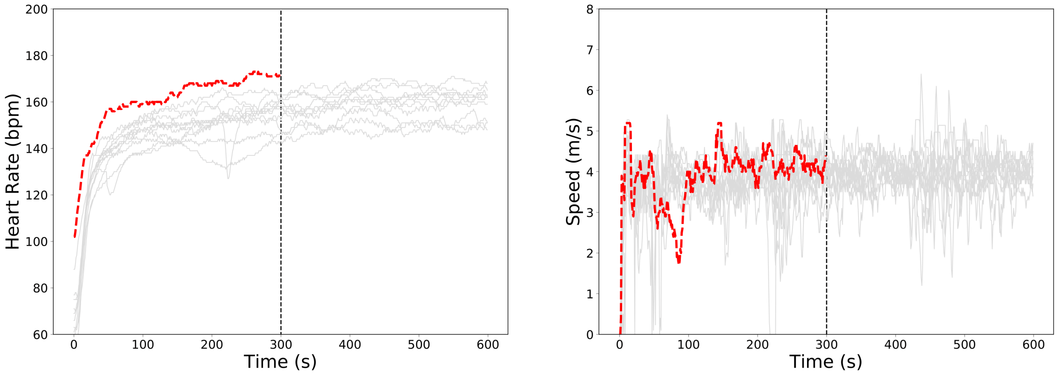

FIGURE 4.2: An example of the within-online setting. The red dashed lines indicate the current monitored activity, while gray lines denote previous activities. The vertical line marks the time at which our algorithm provides the posterior probability that the current activity is a changepoint.

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 *et al.*, 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. 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 *et al.* (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.

4.2 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

$$\mathbf{y}_{n,t} = \begin{bmatrix} \mathbf{Z}^{(S)}_{\boldsymbol{\theta}} & \mathbf{Z}^{(A)}_{\boldsymbol{\theta}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}^{(s)}_t \\ \boldsymbol{\alpha}_{n,t} \end{bmatrix} + \boldsymbol{\epsilon}_{n,t}, \tag{4.1}$$

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

$$\begin{bmatrix} \boldsymbol{\alpha}^{(s)}_{t+1} \\ \boldsymbol{\alpha}_{n,t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}^{(S)}_{\boldsymbol{\theta}} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{(A)}_{\boldsymbol{\theta}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}^{(s)}_t \\ \boldsymbol{\alpha}_{n,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}^{(s)}_t \\ \boldsymbol{\eta}_{n,t} \end{bmatrix}, \tag{4.2}$$

with $\boldsymbol{\eta}^{(s)}_t \overset{iid}{\sim} \mathrm{N}_M(\mathbf{0}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, $\boldsymbol{\eta}_{n,t} \overset{iid}{\sim} \mathrm{N}_K(\mathbf{0}, \boldsymbol{\Delta}_{\boldsymbol{\theta}})$, and $\boldsymbol{\alpha}^{(s)}_1 \overset{iid}{\sim} \mathrm{N}_M(\hat{\boldsymbol{\alpha}}^{(S)}_{1|0}, \mathbf{P}^{(S)}_{1|0})$ independent of $\boldsymbol{\alpha}_{n,1} \overset{iid}{\sim} \mathrm{N}_K(\hat{\boldsymbol{\alpha}}^{(A)}_{1|0}, \mathbf{P}^{(A)}_{1|0})$. 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. The elements $\mathbf{Z}^{(S)}_{\boldsymbol{\theta}}$, $\mathbf{Z}^{(A)}_{\boldsymbol{\theta}}$, $\mathbf{T}^{(S)}_{\boldsymbol{\theta}}$, and $\mathbf{T}^{(A)}_{\boldsymbol{\theta}}$ are non-stochastic design matrices with dimensions $P \times M$, $P \times K$, $M \times M$, and $K \times K$, 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 $\boldsymbol{\eta}^{(s)}_t$ 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 $\boldsymbol{\theta}$.

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 $n' \neq n$. With no information on the initial states, we adopt the diffuse state initialization technique, in which the means and variances are independent of $\boldsymbol{\theta}$, and the latter are supposed to be large (Durbin and Koopman, 2012).

Condition now on $S_{1:N}$ and assume further that the s-th segment ranges between the j_s-th and the k_s-th activity, so that its length is $m_s = k_s - j_s + 1$. We model this segment using the following equations:

$$\begin{bmatrix} \mathbf{y}_{j_s,t} \\ \mathbf{y}_{j_s+1,t} \\ \vdots \\ \mathbf{y}_{k_s,t} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{\boldsymbol{\theta}}^{(S)} & \mathbf{Z}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{Z}_{\boldsymbol{\theta}}^{(S)} & \mathbf{0} & \mathbf{Z}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \vdots \\ \vdots & \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{Z}_{\boldsymbol{\theta}}^{(S)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}_{\boldsymbol{\theta}}^{(A)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{j_s,t} \\ \boldsymbol{\alpha}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\alpha}_{k_s,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{j_s,t} \\ \boldsymbol{\epsilon}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\epsilon}_{k_s,t} \end{bmatrix}, \tag{4.3}$$

$$\begin{bmatrix} \boldsymbol{\alpha}_{t+1}^{(s)} \\ \boldsymbol{\alpha}_{j_s,t+1} \\ \boldsymbol{\alpha}_{j_s+1,t+1} \\ \vdots \\ \boldsymbol{\alpha}_{k_s,t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\boldsymbol{\theta}}^{(S)} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{\boldsymbol{\theta}}^{(A)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{T}_{\boldsymbol{\theta}}^{(A)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{j_s,t} \\ \boldsymbol{\alpha}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\alpha}_{k_s,t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_t^{(s)} \\ \boldsymbol{\eta}_{j_s,t} \\ \boldsymbol{\eta}_{j_s+1,t} \\ \vdots \\ \boldsymbol{\eta}_{k_s,t} \end{bmatrix}, \tag{4.4}$$

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

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 of the model, which has the conditional independence structure

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}, \boldsymbol{\alpha}_{1:T}^{1:N}, S_{1:N}) = p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}|\boldsymbol{\alpha}_{1:T}^{1:N}, S_{1:N})p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:T}^{1:N}|S_{1:N})p(S_{1:N}), \tag{4.5}$$

where $p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:T}^{1:N}|S_{1:N}) = p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:T}^{(1:S_N)}|S_{1:N})p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_{1:N,1:T})$. Conditional on segments $S_{1:N}$, Equations (4.3) and (4.4) 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 (4.5) we obtain the contribution of the s-th segment to the likelihood conditional on $S_{1:N}$ given by

$$\log p_{\boldsymbol{\theta}}(\mathbf{y}_{j_s:k_s,1:T}|S_{j_s:k_s}) = -\frac{1}{2}\sum_{t=1}^{T}\left(m_s P\log(2\pi) + \log|\mathbf{F}_{s,t}| + \boldsymbol{v}'_{j_s:k_s,t}(\mathbf{F}_{s,t})^{-1}\boldsymbol{v}_{j_s:k_s,t}\right), \tag{4.6}$$

where both the innovations vectors $\boldsymbol{v}_{j_s:k_s,t}$ and their respective covariance matrices $\mathbf{F}_{s,t}$ are outputs of the Kalman filter routine, reviewed in Chapter 2 and Appendix B. 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 $\boldsymbol{v}_{j_s:k_s,t}$ 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 *et al.* (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 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

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}|D_{1:N}) = \prod_{n=1}^{N} G^{D}_{\boldsymbol{\theta},n}(D_n), \tag{4.7}$$

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 (4.6), 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.

4.3 Estimation and changepoint detection

4.3.1 From batch to online EM algorithms

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

Let $\hat{\boldsymbol{\theta}}_{it}$ be 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

$$\begin{aligned} Q_{1:N}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{it}) &= \mathrm{E}_{\hat{\boldsymbol{\theta}}_{it}}\big[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}, \boldsymbol{\alpha}_{1:T}^{1:N}, D_{1:N})|\mathbf{y}_{1:N,1:T}\big] && (4.8)\\ &= \mathrm{E}_{\hat{\boldsymbol{\theta}}_{it}}\Bigg[\log p(D_{1:N}) + \mathrm{E}_{\hat{\boldsymbol{\theta}}_{it}}\big[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}, \boldsymbol{\alpha}_{1:T}^{1:N}|D_{1:N})|D_{1:N}, \mathbf{y}_{1:N,1:T}\big]|\mathbf{y}_{1:N,1:T}\Bigg] && (4.9)\end{aligned}$$

The expected value in Equation (4.8) is computed with respect to both $D_{1:N}$ and the latent features $\boldsymbol{\alpha}_{1:T}^{1:N}$, considered jointly, and involves the log-density augmented for both latent variables. Equation (4.9) involves an external and an internal expectation, which are computed with respect to the random variables $D_{1:N}$ and $\boldsymbol{\alpha}_{1:T}^{1:N}|D_{1:N}$, 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 $p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}, \boldsymbol{\alpha}_{1:T}^{1:N}, D_{1:N})$. 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

$$\hat{\boldsymbol{\theta}}_{it+1} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big[Q_{1:N}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{it})\big] = \boldsymbol{\Lambda}(\mathcal{Q}_{1:N}) \tag{4.10}$$

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 (4.8) and the M-step in Equation (4.10) 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

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

with $\iota_{\boldsymbol{\theta}'}(\mathbf{y}_{1,1:T}) = \mathrm{I}(D_1 = 1) + \mathrm{E}_{\boldsymbol{\theta}'}\big[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1,1:T}, \boldsymbol{\alpha}_{1:T}^{1}|D_1)|\mathbf{y}_{1,1:T}, D_1\big]$, 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 (4.9) computed using the observations up to activity n and the same argument calculated using the 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 (4.7) 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. $Q_{1:N}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathrm{E}_{\boldsymbol{\theta}'}\big[\sum_{n=1}^{N} \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})|\mathbf{y}_{1:N,1:T}\big]$, and therefore its sequential evaluation as new activities are observed.

We adopt the stochastic approximation proposed by Yildirim *et al.* (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

$$\begin{aligned}
\mathbf{S}_n(D_{1:n}, \boldsymbol{\theta}') :&= \sum_{j=1}^{n} \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{j,1:T}) \\
\mathbf{T}_n(D_{1:n}, \boldsymbol{\theta}') :&= \sum_{D_{1:(n-1)} \in \mathcal{D}_{1:(n-1)}} \mathbf{S}_n(D_{1:n}, \boldsymbol{\theta}') p_{\boldsymbol{\theta}}(D_{1:(n-1)}|\mathbf{y}_{1:(n-1),1:T}, D_n) \\
&= \sum_{D_{n-1} \in \mathcal{D}_{n-1}} \big[\mathbf{T}_{n-1}(D_{1:(n-1)}, \boldsymbol{\theta}') + \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})\big] p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}, D_n),
\end{aligned} \tag{4.11}$$

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

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

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 (4.11) proposed by Yildirim *et al.* (2013) becomes

$$\begin{aligned}
\mathbf{T}_{\gamma,n}(D_{1:n}; \hat{\boldsymbol{\theta}}_{n-1}) = \sum_{D_{n-1} \in \mathcal{D}_{n-1}} &\big[(1-\gamma_n)\mathbf{T}_{\gamma,n-1}(D_{1:(n-1)}; \hat{\boldsymbol{\theta}}_{n-2}) \\
&+ \gamma_n \iota_{\hat{\boldsymbol{\theta}}_{n-1}}(\mathbf{y}_{n,1:T})\big] p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}, D_n),
\end{aligned} \tag{4.12}$$

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(4.10). 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.

4.3.2 SMC approximation of the predicted probabilities

The between-online setting processes the data of the activities 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 *et al.* (2013) and derive the computations that lead to our algorithm for changepoint detection, details of which are given in Appendix B. Suppose that $p_{\boldsymbol{\theta}}(D_{n-1}, |\mathbf{y}_{1:(n-1),1:T})$ is known. The quantity

$$\begin{aligned} p_{\boldsymbol{\theta}}(D_n, |\mathbf{y}_{1:(n-1),1:T}) &= \sum_{D_{n-1}\in\mathcal{D}_{n-1}} p_{\boldsymbol{\theta}}(D_n, D_{n-1}|\mathbf{y}_{1:(n-1),1:T}) \\ &= \sum_{D_{n-1}\in\mathcal{D}_{n-1}} p(D_n|D_{n-1})p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}) \end{aligned} \tag{4.13}$$

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 (4.13) 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 $\mathcal{D}_{n-1}^{B} = \{d_{n-1}^{1}, \dots, d_{n-1}^{B}, d_{n-1}^{B}\}$ composed by the particles themselves. Consider then the *augmented support* $\mathcal{D}_{n}^{B\star}$ of dimension $2B$ defined as

$$\mathcal{D}_{n}^{B\star} = \{(1, d_{n-1}^{1}), (d_{n-1}^{1}+1, d_{n-1}^{1}), \dots, (1, d_{n-1}^{B}), (d_{n-1}^{B}+1, d_{n-1}^{B})\}.$$

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

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

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

$$p_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:n,1:T}) = \frac{\sum_{D_{n-1}\in\mathcal{D}_{n-1}^{B}} G_{\hat{\boldsymbol{\theta}}_{n-1},n}^{D}(D_n)p(D_n|D_{n-1})\eta_{n-1}^{B}(D_{n-1})}{\sum_{(D'_n,D'_{n-1})\in\mathcal{D}_{(n,n-1)}^{B}} G_{\hat{\boldsymbol{\theta}}_{n-1},n}^{D}(D'_n)p(D'_n|D'_{n-1})\eta_{n-1}^{B}(D'_{n-1})}, \tag{4.14}$$

over the supports $\mathcal{D}_{(n,n-1)}^{B}$ and $\mathcal{D}_n^B$, 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.

4.3.2.1 Maximization step and inner expectations

The maximization step in Equation (4.10) 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 $\mathbf{Z}_{\boldsymbol{\theta}}^{(S)}$, $\mathbf{Z}_{\boldsymbol{\theta}}^{(A)}$, $\mathbf{T}_{\boldsymbol{\theta}}^{(S)}$, $\mathbf{T}_{\boldsymbol{\theta}}^{(A)}$, $\boldsymbol{\Delta}_{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, 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 (4.9). 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 (4.3) and (4.4) 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 $L_{\boldsymbol{\theta}'}(\mathbf{y}_{j:n,1:T}) = \mathrm{E}_{\boldsymbol{\theta}'}\big[\log p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:T}, \boldsymbol{\alpha}^{j:n}_{1:T}|D_n)|\mathbf{y}_{j:n,1:T}, D_n\big]$, 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 $\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}^{j:n}_t|D_n, \mathbf{y}_{j:n,1:T}\big]$, $\mathrm{E}_{\boldsymbol{\theta}'}\big[(\boldsymbol{\alpha}^{j:n}_t)(\boldsymbol{\alpha}^{j:n}_t)'|D_n, \mathbf{y}_{j:n,1:T}\big]$, and $\mathrm{E}_{\boldsymbol{\theta}'}\big[(\boldsymbol{\alpha}^{j:n}_{t+1})(\boldsymbol{\alpha}^{j:n}_t)'|D_n, \mathbf{y}_{j:n,1:T}\big]$, which are computed using the standard Kalman filtering, smoothing, and lagged smoothing routines, reviewed in Appendix B (see, e.g., Shumway and Stoffer, 2017; Durbin and Koopman, 2012).

4.3.3 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

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

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 in the between online setting 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

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n, \mathbf{y}_{1:(n-1),1:T}) = \begin{cases} \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}, \mathbf{y}_{j:(n-1),1:T}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),1:T}|D_n)} & \text{if } D_n > 1 \\ p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n) & \text{if } D_n = 1 \end{cases} \tag{4.15}$$

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 indeces 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 (4.15) with

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

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 slowdowns 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 $\eta_n^B(D_n)$.

4.4 Simulation studies

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

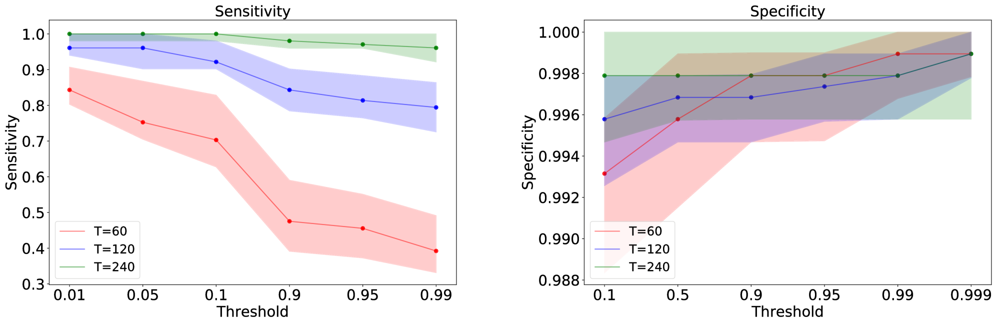

FIGURE 4.3: Medians and 90% confidence intervals of sensitivity and specificity with different thresholds obtained for 20 synthetic examples using our model for $T = 60, 120$ and 240.

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 $\boldsymbol{\alpha}_{t+1}^{(s)} = \mathbf{I}_P \otimes \begin{bmatrix} 0.95 & 1 \\ 0 & 0.90 \end{bmatrix} \boldsymbol{\alpha}_t^{(s)} + \boldsymbol{\xi}_t^{(s)}$, $\boldsymbol{\xi}_t^{(s)} \sim \mathrm{N}_{2P}(\mathbf{0}_{2P}, \sigma_\alpha^2(\mathbf{I}_P \otimes \boldsymbol{\Psi}_0))$ and the activity-specific states according to $\boldsymbol{\alpha}_{n,t+1} = \rho \cdot \boldsymbol{\alpha}_{n,t} + \boldsymbol{\xi}_{n,t}$, $\boldsymbol{\xi}_{n,t} \sim \mathrm{N}_P(\mathbf{0}_P, \sigma_d^2 \mathbf{I}_P)$. We then generated the observations $\mathbf{y}_{n,t} = \begin{bmatrix} \mathbf{I}_P \otimes \begin{bmatrix} 1 & 0 \end{bmatrix} & \mathbf{I}_P \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t^{(s)} \\ \boldsymbol{\alpha}_{n,t} \end{bmatrix} + \boldsymbol{\epsilon}_{n,t}$, $\boldsymbol{\epsilon}_{n,t} \sim \mathrm{N}_P(\mathbf{0}_P, \sigma_\epsilon^2 \mathbf{I}_P)$.

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 Appendix B together with alternative specifications of λ.

We estimate (i) the changepoints in the between-online setting by utilizing Equation (4.14) 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 4.3 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 4.3. It is reassuring 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 $T = 240$.

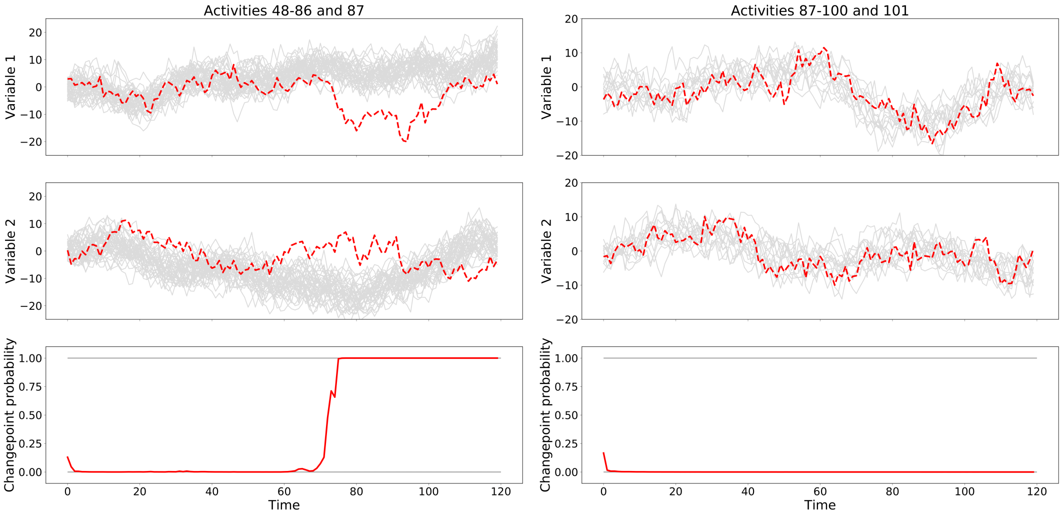

FIGURE 4.4: Two instantiations of our simulation with different activities and the respective filtered probabilities of changepoints. Red dashed line: activity that is being monitored; gray lines: previous activities since the last changepoint.

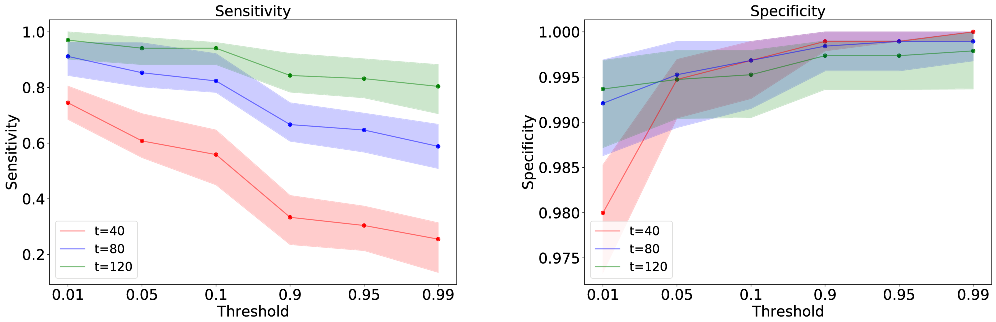

FIGURE 4.5: Medians and 90% confidence intervals of sensitivity and specificity evaluated for 20 synthetic examples using our model in the within-online setting.

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 4.4 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 4.5 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.

4.5 Case study

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 $y_{\mathrm{hr},n,t}$ be the heart rate in beats per minute and $y_{\mathrm{sp},n,t}$ 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

$$\begin{bmatrix} y_{\mathrm{hr},n,t} \\ y_{\mathrm{sp},n,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha^{(s)}_{\mathrm{hr},1,t} \\ \alpha^{(s)}_{\mathrm{hr},2,t} \\ \alpha^{(s)}_{\mathrm{sp},t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{\mathrm{hr},n,t} \\ \alpha_{\mathrm{sp},n,t} \end{bmatrix} + \begin{bmatrix} \upsilon_{\mathrm{hr},n,t} \\ \upsilon_{\mathrm{sp},n,t} \end{bmatrix}, \quad \begin{bmatrix} \upsilon_{\mathrm{hr},n,t} \\ \upsilon_{\mathrm{sp},n,t} \end{bmatrix} \sim \mathrm{N}_2(0, \boldsymbol{\Sigma}),$$

segment-specific state equations

$$\begin{bmatrix} \alpha^{(s)}_{\mathrm{hr},1,t+1} \\ \alpha^{(s)}_{\mathrm{hr},2,t+1} \\ \alpha^{(s)}_{\mathrm{sp},t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha^{(s)}_{\mathrm{hr},1,t} \\ \alpha^{(s)}_{\mathrm{hr},2,t} \\ \alpha^{(s)}_{\mathrm{sp},t} \end{bmatrix} + \begin{bmatrix} \xi^{(s)}_{\mathrm{hr},1,t} \\ \xi^{(s)}_{\mathrm{hr},2,t} \\ \xi^{(s)}_{\mathrm{sp},t} \end{bmatrix}, \quad \begin{bmatrix} \xi^{(s)}_{\mathrm{hr},1,t} \\ \xi^{(s)}_{\mathrm{hr},2,t} \\ \xi^{(s)}_{\mathrm{sp},t} \end{bmatrix} \sim \mathrm{N}_3(0, \boldsymbol{\Psi}),$$

and activity-specific state equations

$$\begin{bmatrix} \alpha_{\mathrm{hr},n,t+1} \\ \alpha_{\mathrm{sp},n,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \rho_{\mathrm{sp}} \end{bmatrix} \begin{bmatrix} \alpha_{\mathrm{hr},n,t} \\ \alpha_{\mathrm{sp},n,t} \end{bmatrix} + \begin{bmatrix} \xi_{\mathrm{hr},n,t} \\ \xi_{\mathrm{sp},n,t} \end{bmatrix}, \quad \begin{bmatrix} \xi_{\mathrm{hr},n,t} \\ \xi_{\mathrm{sp},n,t} \end{bmatrix} \sim \mathrm{N}_2(0, \boldsymbol{\Delta}),$$

with $\boldsymbol{\alpha}^{(s)}_1 = (\alpha^{(s)}_{\mathrm{hr},1,1}, \alpha^{(s)}_{\mathrm{hr},2,1}, \alpha^{(s)}_{\mathrm{sp},1})' \sim \mathrm{N}_3((80,0,0)', \mathrm{diag}(100,1,100))$, $\boldsymbol{\alpha}_{n,1} = (\alpha_{\mathrm{hr},n,1}, \alpha_{\mathrm{sp},n,t})' \sim \mathrm{N}_2(\mathbf{0}, 10 \cdot \mathbf{I}_2)$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Delta}$ are full covariance matrices, and ρ_{sp} is an autoregressive coefficient.

The segment-specific latent states that describe the physical condition and skills

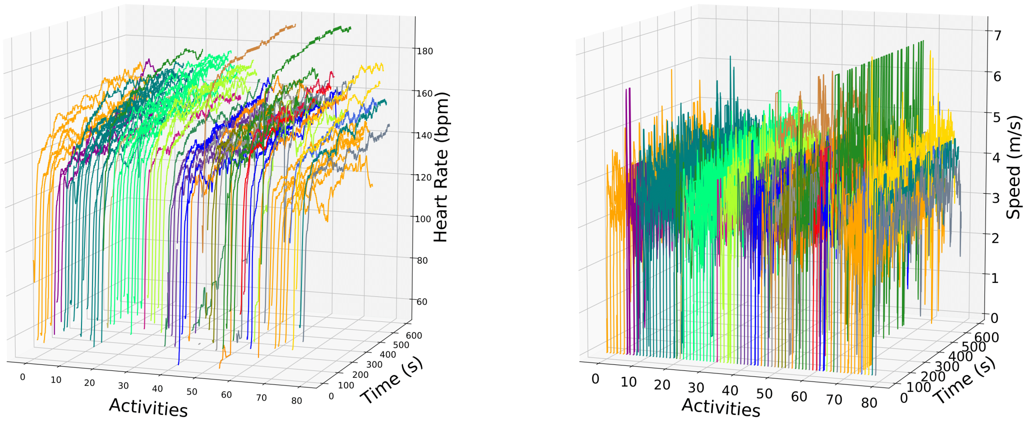

FIGURE 4.6: Segmentation of warm-up activities in the between-online setting for an athlete. The segmentation of the activities was obtained defining the changepoint as those activities for which the filtered distribution at the end of the activity is $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:n,1:T}) > 0.5$.

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 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. We set $\lambda = 0.5$ and, following the guidelines of Yildirim *et al.* (2013), we estimated the parameters of the model $\boldsymbol{\theta} = \{\boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{\Delta}, \rho_{\text{sp}}\}$ by repeating the EM algorithm 30 times in order to reach convergence. Figure 4.6 provides four instantiations of our results. We depict segments in the between-online setting, obtained according to the rule $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:n,1:T}) > 0.50$. The estimated number of changepoints is 34, of which 19 involve activities with a single activity segment. This interesting finding highlights the large variability between successive activities. Of these 19 changepoints, 15 are located in the last 43 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.

Figure 4.7 shows four instantiations of the within-online setting by presenting heart rate, speed, and changepoint probability $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:n,1:T})$ 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 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 changepoint probability is close to 1 after around 20 seconds of warm-up. Activity 32 is similar to activities 23–31

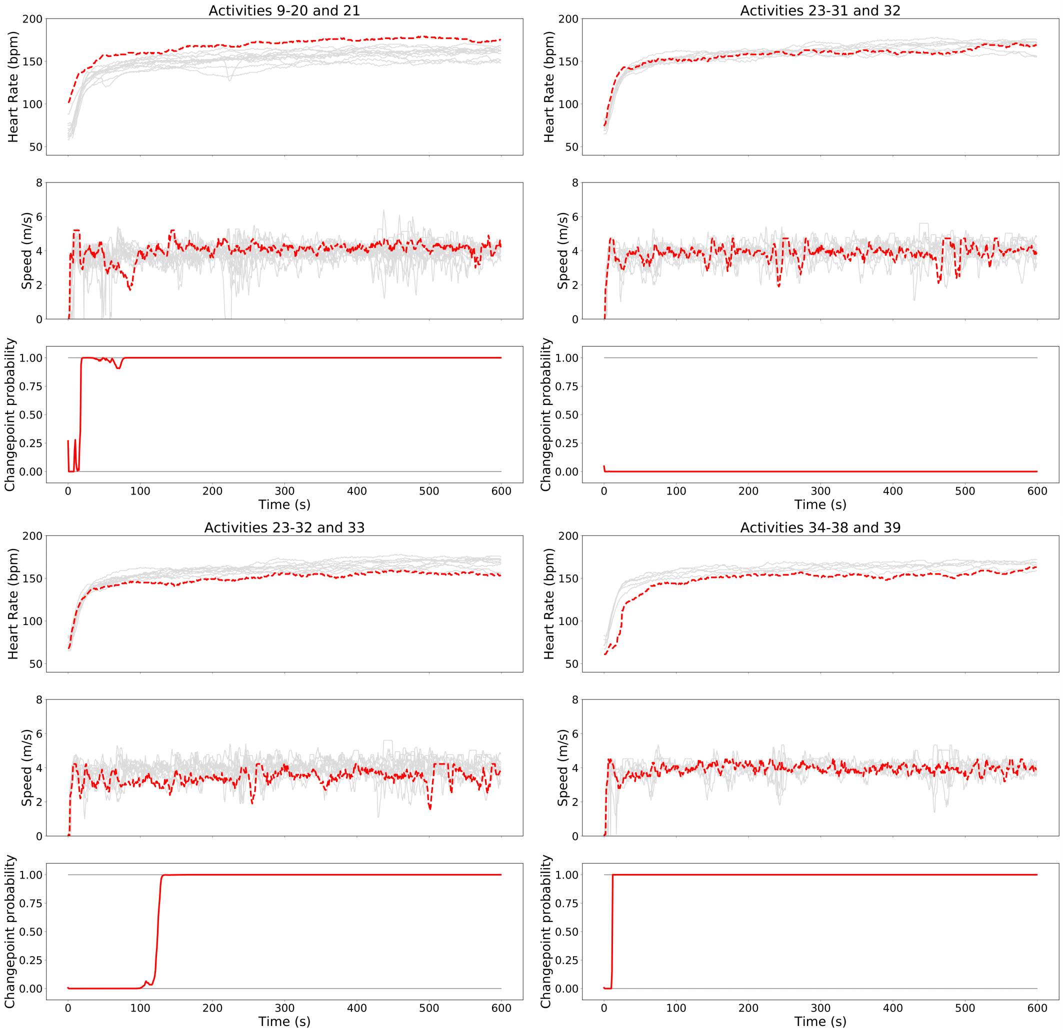

FIGURE 4.7: Selected activities in which the changepoint probability is being monitored. The gray lines in the background represent activities since the last changepoint, obtained according to the rule $\hat{p}_{\hat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:n,1:T}) > 0.5$.

and the changepoint probability is nearly 0 throughout the activity. The bottom left panel shows activity 33, for which the changepoint probability changes strongly after two minutes because both the heart rate and the speed tend to be lower than during other activities, corresponding to athlete putting in less effort. Finally, activity 39 is characterized by a lower heart rate, although the speed curve seems similar to those in previous activities; this indicates less effort and an improved state of well-being of the athlete.

4.6 Discussion

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 *et al.* (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 (4.4). One possible generalization could assume that the block transition matrix in Equation (4.4) 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.

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 *et al.* (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.

Conclusions

Discussion

The recent development of technology has brought researchers from different scientific domains into the world of sports performance analysis. Mathematicians, engineers, computer scientists, and statisticians are involved in different aspects of this field, both in developing technological tools useful to collect and use data, and to answer to research questions of different levels of complexity. The present work showed how statistics can be useful in various aspects related to the development of tools used in sports science. In fact, the chapters of this thesis reported some of the questions that emerge in sports performance analysis and showed how state space models can provide an answer in an exhaustive way. The proposed approaches are varied, and consist in graphical tools for the visualization of collected data, Bayesian clustering methods to analyze the evolution of athletes' careers, and the use of changepoint models to monitor athletes' health during activities.

More specifically, the first chapter introduces the main aspects discussed in sports science and sports performance analysis, reviewing also some recent statistical contributions on the topic, introducing some recent datasets, and detailing research opportunities in this field. In short, research opportunities range from the simple development of graphical tools to summarize data and perform exploratory data analysis, to the development of tools to describe phenomena involving various levels of complexity, related for example to the size of the problems or to the presence of temporal dependence, but also to the need of using collected data for monitoring athletes' performances in a real time environment.

The second chapter introduces the state space models and reviews some of the current proposals for dealing with the analysis of matrix-variate time series. The state space models are a wide class of models that allows for an unified treatment of several problems in the analysis of time series. In the chapter, both the vector and the matrix state space models are described, together with some other models present in the literature for dealing with matrix of observations, including the matrix-variate regression model, the matrix-variate autoregressive process, and the matrix-variate dynamic factor model. The main tools for inference have been introduced, including the Kalman filter and smoother, useful for both state estimation and likelihood evaluation. The usefulness and versatility of state space models have been shown in the next chapters, in which two different models were proposed for clustering athletes' careers and for early identifications of behavioural changes of the observed variables during sports activities.

More specifically, Chapter 3 describes a matrix state space model to cluster multivariate time series describing the performances over the years in 800, 1500, and 5000 meters races of a cohort of Italian middle-distance athletes. The states describing the performances of the athletes in different races are linked to the observed values by means of a selection matrix involved in the measurement equation. Since missing data patterns may be related to the observed performances of middle distance athletes, the presence or absence of data have been modeled using two other processes, describing athletes' attitudes and histories. Subjective priors were used to characterized 9 distinct profiles, corresponding to 9 distinct groups expected prior to the analysis. A Gibbs sampling algorithm was derived, and posterior analysis were carried out for evaluating the parameters involved in the model. Differences across groups were checked by means of a relative performance indicator describing whether the performances of a reference group were better or worse than the others in a given race during the years. Results suggest that: (a) in the 800 meter race, late-entry into competition is associated with worse performances; (b) athletes who are more likely to participate in races other than their reference one have better overall performances. The posterior analysis suggest the presence of groups different from those specified a priori, with one group left without athletes. Although relevant from an interpretive perspective, this result requires further analysis as future development, since it can be influenced both by the effective absence of athletes in that group, but also by the algorithm used for inference. Further developments include also the specification of alternative priors, coherent with the available prior knowledge of the phenomenon, along with an appropriate model's evaluation and diagnostic checking of the results.

Chapter 4 describes a new method for changepoint detection in a doubly-online framework, useful for early detection of distributional changes between a set of repeated running activities. The model combines and leverages tools from both the classical changepoint model and the linear state space models, the firsts allowing the use of an EM algorithm that involves a SMC approach with constant complexity for changepoint identification in a between-online setting, i.e. every time a new activity is fully observed, the seconds for monitoring the real time probability of the presence of a changepoint while a new activity is carried out. The results show the ability of the proposed approach in alarming the athletes on possible problems much before the end of the activities. Among the possible developments that can be considered, the inclusion time-dependent model matrices and of activity-specific covariates is possible without further methodological effort, provided that these elements are non-stochastic. A particularly appealing possible future development, that requires additional methodological effort, consists of considering 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, which is typical in sports performance analysis, as discussed in Chapter 1.

Future directions of research

Future developments that may follow the work presented in this thesis are manifold, in addition to some technical improvements of the tools proposed and refinement of some details on the models. The development of new applications using state space models is surely a starting point for new research activities. In addition to the simplest models, however, it is possible to combine state space models with other methodologies, such as the use of quantile regression tools or the generalization of the proposed models for data far away from Gaussian assumptions, such as counts or data with mixed nature. Sports scientists are often interested in aspects of distribution that are different from the central value, such as the top 5% performances of different athletes. Obviously, the use of statistical tools in sports performance analysis is not limited just to the use of state space. For example, the use of functional data analysis methods can be useful for statistical analysis in biomechanics, since, in this context, the observed data are often smooth curves that can be represented as functions. Injury prediction is a hot topic in this discipline, since, not only is an aspect related the health prevention of athletes, but also teams and people who invest in athletes may be interested in this. If, on one side, training and competition scheduling is something that has always been meticulously done by coaches and specialists, in recent years automated or data driven solutions for training are used for both professional and recreational sports. Studying and defining the relationships between training and competition, based on collected data, can further boost the development of these technique, especially in environments in which is possible the interaction of domain experts and technology. The exchange of ideas between statistics and sports science may open the door to new research directions in both disciplines.

Appendix A

Appendix for "Time series clustering of athletes' careers under informative missing data patterns"

Some details on the Gibbs Sampling algorithm

Cluster allocations and cluster probability

The full conditional of matrix $\mathbf{S}$ is

$$p(\mathbf{S}|\boldsymbol{\theta}, \mathcal{Y}, \mathcal{D}, \mathcal{D}^{\star}, \mathcal{A}) \propto p_{\boldsymbol{\theta}}(\mathcal{Y}|\mathcal{D}, \mathcal{A}, \mathbf{S}) p_{\boldsymbol{\theta}}(\mathcal{D}|\mathcal{D}^{\star}, \mathbf{S}) p_{\boldsymbol{\theta}}(\mathcal{D}^{\star}|\mathbf{S}) p_{\boldsymbol{\theta}}(\mathbf{S}).$$

Given the conditional independence assumptions expressed in Equations (3.7) and (3.10), together with the diagonal structure of $\boldsymbol{\Sigma}^C$, the q-th row of $\mathbf{S}$ can be updated independently of the others according to

$$\Pr(S_q = g|\boldsymbol{\theta}, \mathcal{A}, \mathcal{Y}, \mathcal{D}^{\star}, \mathcal{D}) = \frac{\exp\left[\sum_{t=1}^{T}(\mathcal{Q}^{y}_{qg,t} + \mathcal{Q}^{d}_{qg,t} + \mathcal{Q}^{d^{\star}}_{qg,t}) + \mathcal{Q}^{S}_{qg}\right]}{\sum_{j=1}^{G}\exp\left[\sum_{t=1}^{T}(\mathcal{Q}^{y}_{qj,t} + \mathcal{Q}^{d}_{qj,t} + \mathcal{Q}^{d^{\star}}_{qj,t}) + \mathcal{Q}^{S}_{qj}\right]},$$

where $\mathcal{Q}^{y}_{qg,t}$, $\mathcal{Q}^{d}_{qg,t}$, $\mathcal{Q}^{d^{\star}}_{qg,t}$, and $\mathcal{Q}^{S}_{qg}$ are log-densities obtained by isolating all components related to subject q and group g in the full conditional.Note that this step requires the completed data $\mathcal{Y}$ to be available, despite the missing values. Their obtaining is discussed in the next sections. Alternative samplers that marginalize out $\mathcal{A}$, or other elements involved in the model, such as $\widetilde{\mathcal{E}}$ and the probabilities associated to missing values and cluster allocations, are easily derivable. However, these would lead to intensive procedure, where each row of $\mathbf{S}$ is updated conditionally to the others, or where moves are performed only locally on the neighborhood of $\mathbf{S}$ (see Nobile and Fearnside, 2007; Titsias and Yau, 2017; Zanella, 2020). Once $\mathbf{S}$ is updated, π is updated according to

$$\pi|\mathbf{S} \sim \mathrm{Dir}_G(1/G + \sum_{q=1}^{Q} \mathrm{I}(S_q = 1), \ldots, 1/G + \sum_{q=1}^{Q} \mathrm{I}(S_q = G)).$$

Missing data probabilities

The full conditional of missing values probabilities are

$$\pi_{pg}|\mathcal{D}, \mathcal{D}^\star, \mathbf{S} \sim \mathrm{Be}(\alpha_{pg} + N'_{pg}, \beta_{pg} + N''_{pg}),$$
$$\pi^\star_{1g}|\mathcal{D}^\star, \mathbf{S} \sim \mathrm{Be}(\alpha^\star_{1g} + N^{\star\prime}_{1g}, \beta^\star_{1g} + N^{\star\prime\prime}_{1g}), \quad \pi^\star_{2g}|\mathcal{D}^\star, \mathbf{S} \sim \mathrm{Be}(\alpha^\star_{2g} + N^{\star\prime}_{2g}, \beta^\star_{2g} + N^{\star\prime\prime}_{2g})$$

where

$$\begin{aligned}
N'_{pg} &= \sum_{q=1}^{Q} \mathrm{I}(S_q = g) \sum_{t=1}^{T} \mathrm{I}(d_{pq,t} = 1)\mathrm{I}(d^\star_{q,t} = 1),\\
N''_{pg} &= \sum_{q=1}^{Q} \mathrm{I}(S_q = g) \sum_{t=1}^{T} \mathrm{I}(d_{pq,t} = 0)\mathrm{I}(d^\star_{q,t} = 1),\\
N^{\star\prime}_{1g} &= \sum_{q=1}^{Q} \mathrm{I}(S_q = g) \left[\mathrm{I}(d^\star_{q,1} = 1) + \sum_{t=1}^{T} \mathrm{I}(d^\star_{q,t-1} = 0)\mathrm{I}(d^\star_{q,t} = 1)\right],\\
N^{\star\prime\prime}_{1g} &= \sum_{q=1}^{Q} \mathrm{I}(S_q = g) \left[\mathrm{I}(d^\star_{q,1} = 0) + \sum_{t=2}^{T} \mathrm{I}(d^\star_{q,t-1} = 0)\mathrm{I}(d^\star_{q,t} = 0)\right],\\
N^{\star\prime}_{2g} &= \sum_{q=1}^{Q} \mathrm{I}(S_q = g) \left[\sum_{t=2}^{T} \mathrm{I}(d^\star_{q,t-1} = 1)\mathrm{I}(d^\star_{q,t} = 2)\right],\\
N^{\star\prime\prime}_{2g} &= \sum_{q=1}^{Q} \mathrm{I}(S_q = g) \left[\sum_{t=2}^{T} \mathrm{I}(d^\star_{q,t-1} = 1)\mathrm{I}(d^\star_{q,t} = 1)\right],
\end{aligned}$$

are counts that reflect the conditioning structure imposed in Equation (3.8), for $g = 1, \ldots, G$, and $p = 1, \ldots, P$. We note that hyper-parameter α_{pg}, β_{pg}, $\alpha^\star_{1g}$, $\beta^\star_{1g}$, $\alpha^\star_{2g}$, and $\beta^\star_{2g}$ are group-dependent, and reflect prior beliefs previously explained.

Covariance matrices

Updating steps for covariance matrices are standard updates for inverse Wishart and inverse gamma priors with Gaussian likelihood, and more specifically are

$$\mathbf{\Sigma}^R|\mathcal{Y}, \mathcal{A}, \mathbf{S} \sim \mathrm{IW}_P\big(\nu'_\sigma, \mathbf{\Sigma}'\big), \quad \mathbf{\Psi}_{800}|\mathcal{A} \sim \mathrm{IW}_F\big(\nu'_{800}, \mathbf{\Psi}'_{800}\big),$$

$$\mathbf{\Psi}_{1500}|\mathcal{A} \sim \mathrm{IW}_F\big(\nu'_{1500}, \mathbf{\Psi}'_{1500}\big), \quad \mathbf{\Psi}_{5000}|\mathcal{A} \sim \mathrm{IW}_F\big(\nu'_{5000}, \mathbf{\Psi}'_{5000}\big),$$

where

$$\nu_\sigma^{R\prime} = \nu_\sigma^R + QT, \quad \mathbf{\Sigma}^{R\prime} = \mathbf{\Sigma}_0^R + \sum_{t=1}^{T}(\mathbf{Y}_t - \mathbf{Z}\mathbf{A}_t\mathbf{S}^\top)(\mathbf{Y}_t - \mathbf{Z}\mathbf{A}_t\mathbf{S}^\top)^\top,$$

$$\nu'_{800} = \nu_{800} + 3(T-1), \quad \mathbf{\Psi}'_{800} = \mathbf{\Psi}^0_{800} + \sum_{g=1}^{3}\sum_{t=2}^{T}(\boldsymbol{\alpha}_t^{(g)} - \mathbf{T}\boldsymbol{\alpha}_{t-1}^{(g)})(\boldsymbol{\alpha}_t^{(g)} - \mathbf{T}\boldsymbol{\alpha}_{t-1}^{(g)})^\top,$$

$$\nu'_{1500} = \nu_{1500} + 3(T-1), \quad \mathbf{\Psi}'_{1500} = \mathbf{\Psi}^0_{1500} + \sum_{g=4}^{6}\sum_{t=2}^{T}(\boldsymbol{\alpha}_t^{(g)} - \mathbf{T}\boldsymbol{\alpha}_{t-1}^{(g)})(\boldsymbol{\alpha}_t^{(g)} - \mathbf{T}\boldsymbol{\alpha}_{t-1}^{(g)})^\top,$$

$$\nu'_{5000} = \nu_{5000} + 3(T-1), \quad \mathbf{\Psi}'_{5000} = \mathbf{\Psi}^0_{5000} + \sum_{g=7}^{9}\sum_{t=2}^{T}(\boldsymbol{\alpha}_t^{(g)} - \mathbf{T}\boldsymbol{\alpha}_{t-1}^{(g)})(\boldsymbol{\alpha}_t^{(g)} - \mathbf{T}\boldsymbol{\alpha}_{t-1}^{(g)})^\top,$$

State estimation and missing values To obtain a draw from the full conditional of the states $\mathcal{A}$, we adopt the simulation smoothing technique by Durbin and Koopman (2002), after applying the reduction by transformation technique presented in Section 2.4.3. In doing so, we first identify suitable transformations to reduce the vector of augmented observations $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$, and then introduce the steps required to get a draw from

$$p(\mathcal{A}|\boldsymbol{\theta}, \mathbf{S}, \mathcal{Y}, \mathcal{D}).$$

Suppose, for simplicity, that $\mathbf{Z}$ is full row-rank and that all groups have athletes, so that $\mathbf{S}$ is full column-rank. We can apply the reduction by transformation technique to vector form of the model by considering the following decomposition

$$(\mathbf{S} \otimes \mathbf{Z}) = (\mathbf{S}\mathbf{I}_G) \otimes (\mathbf{I}_P\mathbf{Z}) = (\mathbf{S} \otimes \mathbf{I}_P)(\mathbf{I}_G \otimes \mathbf{Z}).$$

If we consider $\mathbf{Z}_\dagger = (\mathbf{S} \otimes \mathbf{I}_P)$, and $\mathbf{A}' = \mathbf{Z}_\dagger^\top\mathbf{\Sigma}^{-1}$, state estimation can be applied to the reduced vector of observations $\mathbf{y}'_t = \mathbf{A}'\mathbf{y}_t$ which is of dimensions $PG \times 1$, where PG is typically such that $PG \ll PQ$, leading to larger speed-ups when $G \ll Q$.

States are then obtained by considering the model

$$\mathbf{y}'_t = \mathbf{A}'(\mathbf{S} \otimes \mathbf{Z})\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}'_t, \quad \boldsymbol{\varepsilon}'_t \sim \text{N}_{PG^+}(\mathbf{0}, \mathbf{\Sigma}'), \tag{A.1}$$

$$\boldsymbol{\alpha}_{t+1} = (\mathbf{U} \otimes \mathbf{T})\boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \text{N}_{FG}(\mathbf{0}, \mathbf{\Psi}^C \otimes \mathbf{\Psi}^R), \tag{A.2}$$

for $\boldsymbol{\alpha}_1 \sim \text{N}_{FG}(\boldsymbol{\alpha}_{1|0}, \mathbf{P}_{1|0})$ and $\mathbf{\Sigma}' = \mathbf{A}'\mathbf{\Sigma}\mathbf{A}'^\top$. In particular, we consider the following steps:

1. Obtain independent samples from $\boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_T$, $\boldsymbol{\xi}'_1, \dots, \boldsymbol{\xi}'_{T-1}$, and $\boldsymbol{\alpha}_1$ from the respective distributions in Equations (A.1) and (A.2);

2. Use Equations (A.1) and (A.2) with the samples obtained in Step 1 for obtaining fictional observations $\bar{\mathbf{y}}'_1, \dots, \bar{\mathbf{y}}'_1$ and fictional states $\bar{\boldsymbol{\alpha}}_1, \dots, \bar{\boldsymbol{\alpha}}_T$;

3. Transform the vectors of observations as $\bar{\bar{\mathbf{y}}}'_t = \mathbf{y}'_t - \bar{\mathbf{y}}'_t$, for $t = 1, \dots, T$;

4. Use $\bar{\bar{\mathbf{y}}}'_1, \dots, \bar{\bar{\mathbf{y}}}'_T$ to obtain $\bar{\bar{\boldsymbol{\alpha}}}_{1|T}, \dots, \bar{\bar{\boldsymbol{\alpha}}}_{T|T}$ by means of a Kalman smoothing recursion under the model in Equations (A.1) and (A.2);

5. Use $\tilde{\boldsymbol{\alpha}}_t = \bar{\bar{\boldsymbol{\alpha}}}_{t|T} + \bar{\boldsymbol{\alpha}}_t$, for $t = 1, \dots, T$, as a sample from the full conditional of the states.

These steps are nothing else than Algorithm 2 in Durbin and Koopman (2002), suitably modified to obtain single draws from the full conditional $p(\mathcal{A}|\boldsymbol{\theta}, \mathrm{S}, \mathcal{Y}, \mathcal{D})$. In the case $\mathbf{Z}$ or $\mathbf{S}$ are not full rank matrices, further reduction can be obtained. $\hat{\mathbf{A}}_{1|0}$ is obtained by simple update of Gaussian prior with Gaussian likelihood and known covariance matrix.

Once $\mathcal{A}$ is known, draws from

$$p(\widetilde{\mathcal{E}}|\mathcal{A}, \mathrm{S}, \boldsymbol{\theta}, \mathcal{Y}^\star, \mathcal{D}, \mathcal{D}^\star),$$

are obtained using simple rules of multivariate Gaussian. In particular, consider $\mathbf{y}_{\cdot q,t}$, and suppose that athlete q belongs to group g, for which the states $\boldsymbol{\alpha}_t^{(g)}$ are known. If all observations were available, then the vector of errors would be $\boldsymbol{\varepsilon}_{\cdot q,t} = \mathbf{y}_{\cdot q,t} - \mathbf{Z}\boldsymbol{\alpha}_t^{(g)}$ deterministically. However, missing values are present, so we can consider

$$\boldsymbol{\varepsilon}_{\cdot q,t} = \begin{bmatrix} \boldsymbol{\varepsilon}^\star_{\cdot q,t} \\ \tilde{\boldsymbol{\varepsilon}}_{\cdot q,t} \end{bmatrix} \sim \mathrm{N}_P\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_\star & \widetilde{\boldsymbol{\Sigma}}_\star^\top \\ \widetilde{\boldsymbol{\Sigma}}_\star & \widetilde{\boldsymbol{\Sigma}} \end{bmatrix} \right), \tag{A.3}$$

and obtain the errors $\tilde{\boldsymbol{\varepsilon}}_{\cdot q,t}$ associated with missing values by conditioning, according to

$$\tilde{\boldsymbol{\varepsilon}}_{\cdot q,t}|\boldsymbol{\theta}, \mathrm{s}_{q\cdot}, \boldsymbol{\varepsilon}^\star_{\cdot q,t} \sim \mathrm{N}_{\tilde{P}}\big(- \widetilde{\boldsymbol{\Sigma}}_\star \boldsymbol{\Sigma}_\star^{-1} \boldsymbol{\varepsilon}^\star_{\cdot q,t}, \widetilde{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{\Sigma}}_\star \boldsymbol{\Sigma}_\star^{-1} \widetilde{\boldsymbol{\Sigma}}_\star^\top \big).$$

Some other details on the results

In Figure A.1 shows the posterior probabilities describing athletes' attitudes and histories derived from the sample of $\mathbb{Q}_1(\mathbf{S})$. The plot can be used for comparison with Figure 3.5. Figures A.2, A.3, and A.4 show performances for the groups on the observed races. Quantiles are based on the obtained sample and consider the presence of uncertainty present in $\mathbf{S}$. On the contrary, data are shown based on a posterior summary $\mathbf{S}$, that does not account for all the uncertainty present. This explains why median performances are not always centered with the data, and also why some groups have exploding quantiles. In fact, the two groups with anomalous behavior are characterized by iterations in

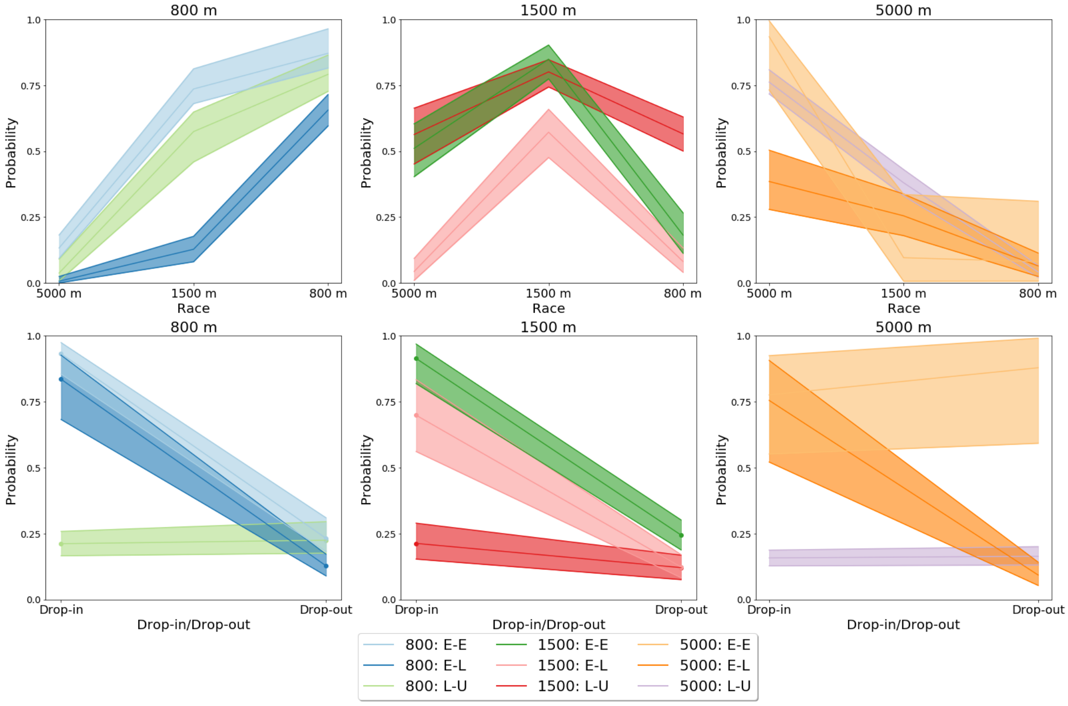

FIGURE A.1: Posterior probabilities describing athletes' attitudes and histories derived from the sample of $\mathbb{Q}_1(\mathbf{S})$.

which the group is lacking or has few athletes. This also motivates the method's ability to construct trajectories despite appearing to be data-less.

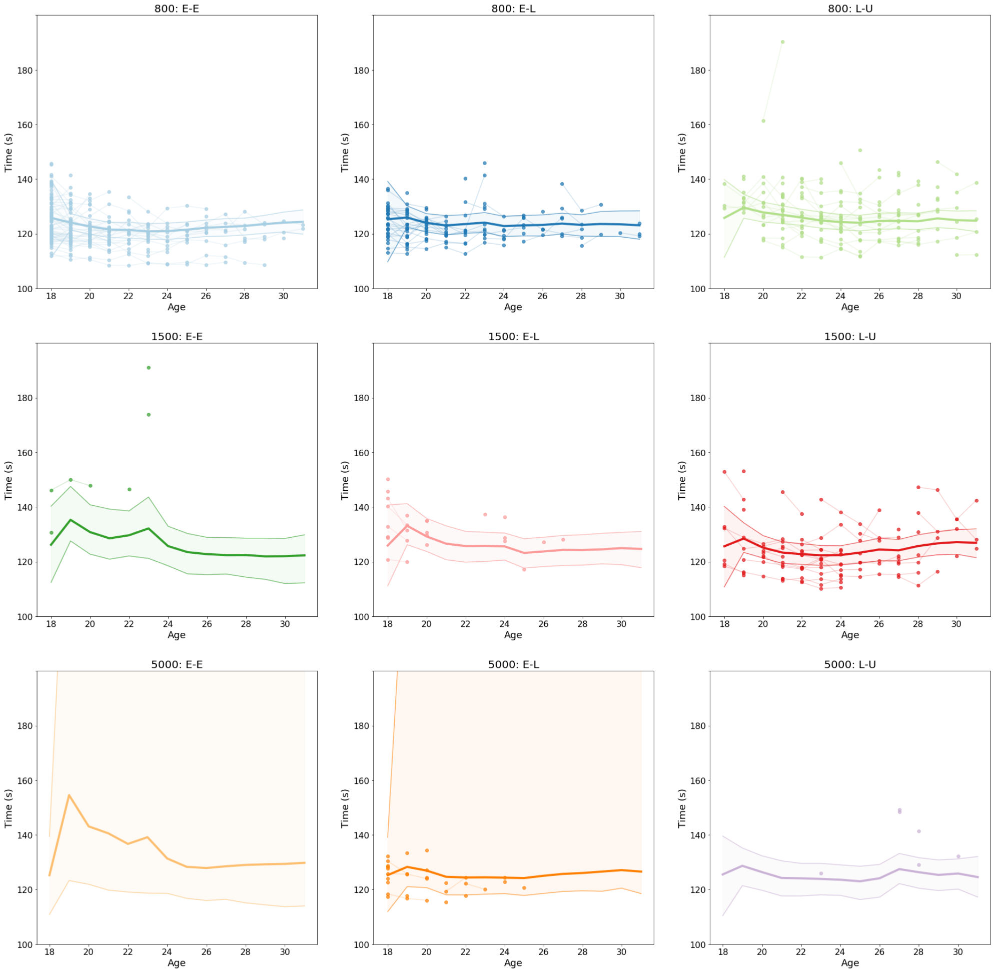

FIGURE A.2: Performances on 800 meters race for the groups. Thicker lines denote posterior medians of the states. Colored bands denote the respective 90% pointwise posterior credible intervals. Observed data are represented in the background, according to athletes' MAP cluster allocations.

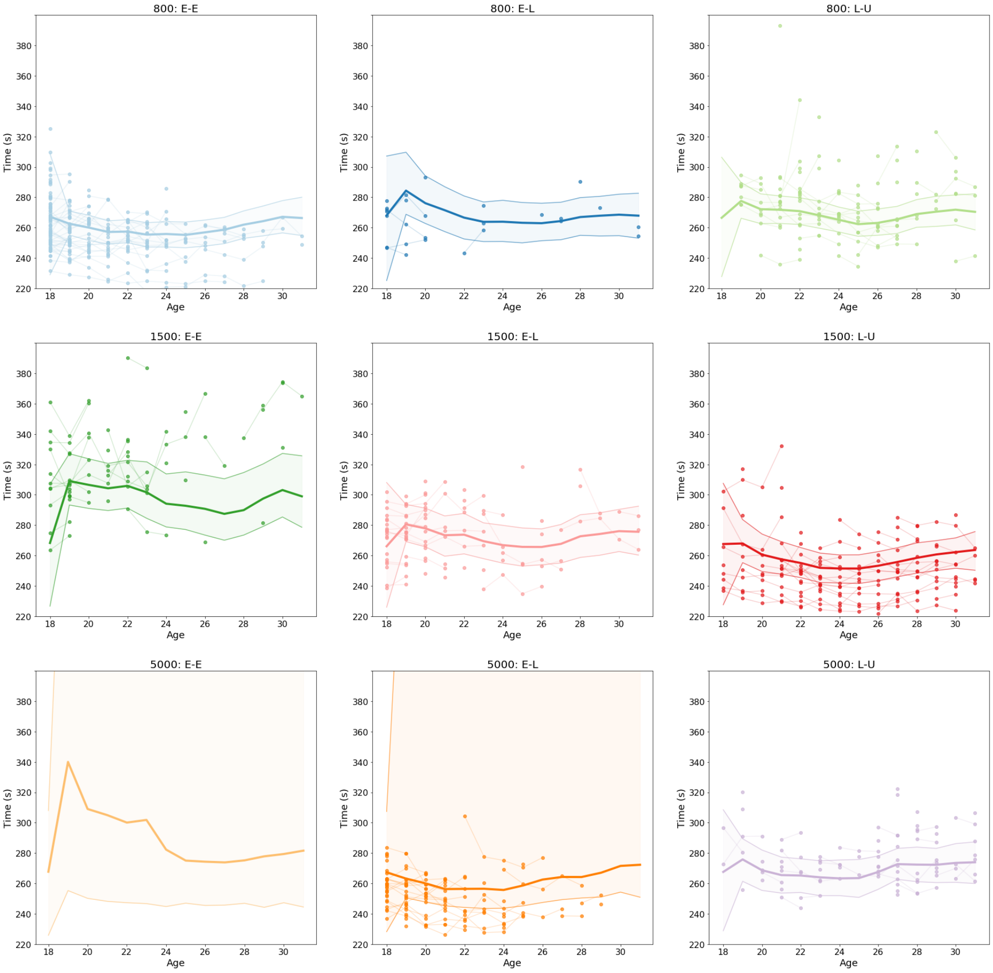

Figure A.3: Performances on 1500 meters race for the groups. Thicker lines denote posterior medians of the states. Colored bands denote the respective 90% pointwise posterior credible intervals. Observed data are represented in the background, according to athletes' MAP cluster allocations.

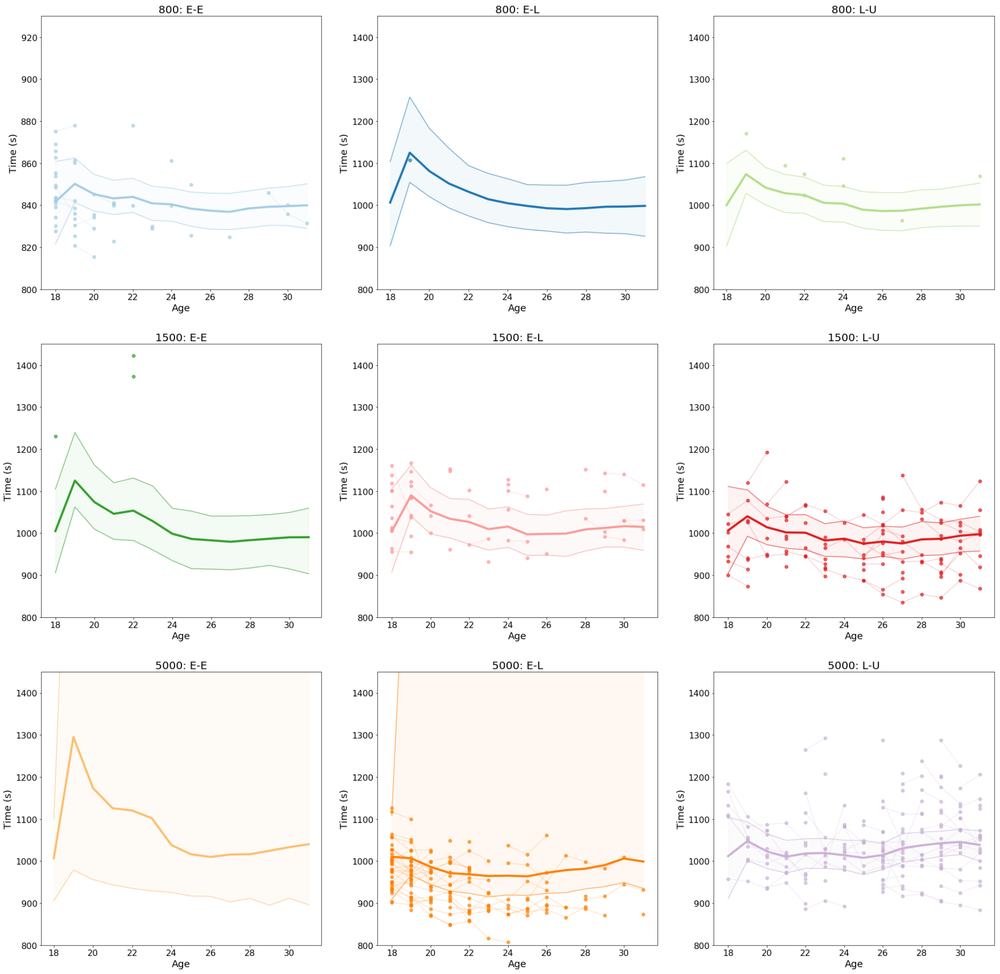

Figure A.4: Performances on 5000 meters race for the groups. Thicker lines denote posterior medians of the states. Colored bands denote the respective 90% pointwise posterior credible intervals. Observed data are represented in the background, according to athletes' MAP cluster allocations.

Appendix B

Appendix for “Doubly-online changepoint detection for monitoring health status during sports activities”

Proofs and derivations

Writing conditional likelihood in term of potentials. Set $j_1 = 0$ and $j_{s+1} = N$ and consider then the chain

$$S_{1:N} = (S_1 = 1, \dots, S_{j_2} = 1, S_{j_2+1} = 2, S_{j_2+1} = 2, \dots, S_{j_3} = 2, \dots,\\ S_{j_s+1} = s, \dots, S_{j_{s+1}} = s).$$

Note that the knowledge of $S_{1:N}$ is equivalent to know the whole sequence

$$D_{1:N} = (D_1 = 1, \dots, D_{j_2} = j_2,\\ D_{j_2+1} = 1, \dots, D_{j_3} = j_3 - j_2, \dots, D_{j_s+1} = 1, \dots, N - j_s),$$

and, therefore, $p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}|S_{1:N}) = p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}|D_{1:N})$, which reduces to

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N,1:T}|S_{1:N}) = \prod_{m=1}^{s} p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):j_{m+1},1:T}|S_{(j_m+1):j_{m+1}}),$$

by leveraging the independence assumption between activities of different segments. Now, consider the m-th segment, starting at index $j_m + 1$ and ending at index j_{m+1} having length $D_{j_{m+1}} = j_{m+1} - j_m$. Then the joint distribution of the m-th segment conditional on $S_{1:N}$ becomes

$$p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):j_{m+1},1:T}|S_{(j_m+1):j_{m+1}}) = p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):j_{m+1},1:T}|D_{j_{m+1}} = j_{m+1} - j_m). \tag{A.1}$$

We then note that

$$\begin{aligned}
&p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):j_{m+1},1:T}|D_{j_{m+1}} = j_{m+1} - j_m) \\
&\quad = p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1),1:T}|D_{j_m+1} = 1) \times \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):(j_m+2),1:T}|D_{j_m+2} = 2)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1),1:T}|D_{j_m+1} = 1)} \\
&\qquad \times \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):(j_m+3),1:T}|D_{j_m+3} = 3)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):(j_m+2),1:T}|D_{j_m+2} = 2)} \\
&\qquad \times \ldots \times \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):j_{m+1},1:T}|D_{j_{m+1}} = j_{m+1} - j_m)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{(j_m+1):(j_{m+1}-1),1:T}|D_{j_{m+1}-1} = j_{m+1} - j_m - 1)},
\end{aligned}$$

where the multiplicands in the previous expression are the potentials, defined as follows:

$$G_{\boldsymbol{\theta},n}(D_n) = p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:T}|D_{1:n}, \mathbf{y}_{1:(n-1),1:T}) = \begin{cases} \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:n,1:T}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),1:T}|D_{n-1})}, & \text{if } D_n > 1 \\ p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:T}|D_n), & \text{if } D_n = 1. \end{cases} \tag{A.2}$$

Plugging the multiplicands in terms of the potentials in equation (A.1), we obtain the conditional likelihood as a product of N potentials. Remember that if $D_n = d$, with $d > 1$, then D_{n-1} with probability 1.

Forward smoothing technique

Here we provide the step by step derivation of the forward smoothing technique (see the manuscript for definitions and details)

Recursive formula. Recall that

$$\begin{aligned}
p_{\boldsymbol{\theta}}(D_{1:(n-1)}|\mathbf{y}_{1:(n-1),1:T}, D_n) &= p_{\boldsymbol{\theta}}(D_{1:(n-2)}, D_{n-1}|\mathbf{y}_{1:(n-1),1:T}, D_n) \\
&= p_{\boldsymbol{\theta}}(D_{1:(n-2)}|D_{n-1}, \mathbf{y}_{1:(n-2),1:T}, \mathbf{y}_{n-1,1:T}, D_n) \\
&\qquad \times p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}, D_n) \\
&= p_{\boldsymbol{\theta}}(D_{1:(n-2)}|D_{n-1}, \mathbf{y}_{1:(n-2),1:T}) p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}, D_n),
\end{aligned}$$

and it follows that

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

$$= \sum_{D_{n-1} \in \mathcal{D}_{n-1}} \Big[\sum_{D_{1:(n-2)} \in \mathcal{D}_{1:(n-2)}} (\mathbf{S}_{n-1}(D_{1:(n-1)}, \boldsymbol{\theta}') + \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T}))\Big]$$

$$\times\, p_{\boldsymbol{\theta}}(D_{1:(n-2)} | D_{n-1}, \mathbf{y}_{1:(n-2),1:T}) p_{\boldsymbol{\theta}}(D_{n-1} | \mathbf{y}_{1:(n-1),1:T}, D_n)$$

$$= \sum_{D_{n-1} \in \mathcal{D}_{n-1}} \Big[\sum_{D_{1:(n-2)} \in \mathcal{D}_{1:(n-2)}} (\mathbf{S}_{n-1}(D_{1:(n-1)}, \boldsymbol{\theta}') p_{\boldsymbol{\theta}}(D_{1:(n-2)} | D_{n-1}, \mathbf{y}_{1:(n-2),1:T})$$

$$+ \sum_{D_{1:(n-2)} \in \mathcal{D}_{1:(n-2)}} p_{\boldsymbol{\theta}}(D_{1:(n-2)} | D_{n-1}, \mathbf{y}_{1:(n-2),1:T}) \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})\Big] p_{\boldsymbol{\theta}}(D_{n-1} | \mathbf{y}_{1:(n-1),1:T}, D_n)$$

$$= \sum_{D_{n-1} \in \mathcal{D}_{n-1}} \big[\mathbf{T}_{n-1}(D_{1:(n-1)}, \boldsymbol{\theta}') + \iota_{\boldsymbol{\theta}'}(\mathbf{y}_{n,1:T})\big] p_{\boldsymbol{\theta}}(D_{n-1} | \mathbf{y}_{1:(n-1),1:T}, D_n),$$

since by definition

$$\mathbf{T}_{n-1}(D_{1:(n-1)}, \boldsymbol{\theta}') = \sum_{D_{1:(n-2)} \in \mathcal{D}_{1:(n-2)}} (\mathbf{S}_{n-1}(D_{1:(n-1)}, \boldsymbol{\theta}') p_{\boldsymbol{\theta}}(D_{1:(n-2)} | D_{n-1}, \mathbf{y}_{1:(n-2),1:T})$$

and $\sum_{D_{1:(n-2)} \in \mathcal{D}_{1:(n-2)}} p_{\boldsymbol{\theta}}(D_{1:(n-2)} | D_{n-1}, \mathbf{y}_{1:(n-2),1:T}) = 1$.

Expected value through recursion. The expected value over $D_{1:n}$

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

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

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

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

where the last expression is obtained by plugging

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

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

Sequential Monte Carlo - Filtering recursion

Here we provide the expressions for the filtering recursion, that provides the exact *predicted, filtered* and *smoothed* probabilities that represent the core algorithms of the SMC technique adopted in the paper approximate the predicted probability. Once the predicted probabilities are obtained by the SMC approximations, they are simply plugged into the expressions here derived.

Predicted probability.

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

where, by the assumptions on the model $p_{\boldsymbol{\theta}}(D_n|D_{n-1}, \mathbf{y}_{1:(n-1),1:T}) = p_{\boldsymbol{\theta}}(D_n|D_{n-1})$.

Filtered probability. Given the predicted probability $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})$, the direct application of the Bayes formula provides a closed form expression for the *filtered probability.* Indeed

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

where $G_{\boldsymbol{\theta},n}(D_n) = p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:T}|D_n, \mathbf{y}_{1:(n-1),1:T})$, by definitions of potential.

Smoothed probability. The knowledge of the predicted probability $p_{\boldsymbol{\theta}}(D_{n-1}|\mathbf{y}_{1:(n-2),1:T})$ allows to compute the smoothed probability

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

Within online setting

Real time probabilities. We should monitor the quantity $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{n,1:t}, \mathbf{y}_{1:(n-1),1:T})$, for any $t < T$. This quantity is exactly

$$\begin{aligned}
p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{n,1:t}, \mathbf{y}_{1:(n-1),1:T}) &= \frac{p_{\boldsymbol{\theta}}(D_n, \mathbf{y}_{n,1:t}|\mathbf{y}_{1:(n-1),1:T})}{\sum_{D'_n\in\mathcal{D}^n} p_{\boldsymbol{\theta}}(D'_n, \mathbf{y}_{n,1:t}|\mathbf{y}_{1:(n-1),1:T})} && \text{(A.3)} \\
&= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n, \mathbf{y}_{1:(n-1),1:T})p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})}{\sum_{D'_n\in\mathcal{D}^n} p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D'_n, \mathbf{y}_{1:(n-1),1:T})p_{\boldsymbol{\theta}}(D'_n|\mathbf{y}_{1:(n-1),1:T})}. && \text{(A.4)}
\end{aligned}$$

The distribution $p_{\boldsymbol{\theta}}(D_n|\mathbf{y}_{1:(n-1),1:T})$ is obtained by SMC approximations as described in Algorithm 2. We now focus on the quantity

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

which becomes

$$\begin{aligned}
p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n, \mathbf{y}_{1:(n-1),1:T}) &= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}, \mathbf{y}_{1:(n-1),1:T}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:T}|D_n)} && \text{(A.5)} \\
&= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:n,1:t}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:t}|D_n)} \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):T}|\mathbf{y}_{1:n,1:t}, D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):T}|\mathbf{y}_{1:(n-1),1:t}, D_n)}, && \text{(A.6)}
\end{aligned}$$

for $D_n > 1$. Let us now consider equation (A.5). For each D_n, we need to run twice the Kalman filter recursion over the entire time domain $1 : T$, where, in the numerator of equation (A.5), the observations $\mathbf{y}_{n,(t+1):T}$ are treated as missing values (as they are not yet observed). Note also that, by the independence assumption between different segments, just 2 runs of the Kalman filter involving the last D_n activities are used for deriving the ratio in equation (A.5) (and not all activities). Although not directly expressed in (A.5), if $D_n > 1$, this fact can be understood by noting that equation (A.5) is

$$\begin{aligned}\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t},\mathbf{y}_{1:(n-1),1:T}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:T}|D_n)} &= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|\mathbf{y}_{1:(n-1),1:T},D_n)\prod_{k=1}^{n-1}G_{\boldsymbol{\theta},k}(D_k)}{\prod_{k=1}^{n-1}G_{\boldsymbol{\theta},k}(D_k)}\\ &= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|\mathbf{y}_{1:(n-1),1:T},D_n)\prod_{k=1}^{j-1}G_{\boldsymbol{\theta},k}(D_k)\prod_{k=j}^{n-1}G_{\boldsymbol{\theta},k}(D_k)}{\prod_{k=1}^{j-1}G_{\boldsymbol{\theta},k}(D_k)\prod_{k=j}^{n-1}G_{\boldsymbol{\theta},k}(D_k)}\\ &= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|\mathbf{y}_{1:(n-1),1:T},D_n)\prod_{k=j}^{n-1}G_{\boldsymbol{\theta},k}(D_k)}{\prod_{k=j}^{n-1}G_{\boldsymbol{\theta},k}(D_k)}\\ &= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t},\mathbf{y}_{j:(n-1),1:T}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{j:(n-1),1:T}|D_{n-1})},\end{aligned}$$

with $j = \max(1, n - D_n + 1)$, and $D_{n-1} = D_n - 1$ with probability 1. Nevertheless, for large T and small t the computation might be an unfeasible step, as it requires to go ahaed with Kalman filtering to the end of previous activities. Therefore, we fix $k \geq 0$ and equation (A.6) can factorizes as follows:

$$\begin{aligned}&\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:n,1:t}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:t}|D_n)}\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):T}|\mathbf{y}_{1:n,1:t},D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):T}|\mathbf{y}_{1:(n-1),1:t},D_n)}\\ &= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:n,1:t}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:t}|D_n)}\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:n,1:t},D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:(n-1),1:t},D_n)}\\ &\quad\times\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+k+1):T}|\mathbf{y}_{1:n,1:t},\mathbf{y}_{1:(n-1),(t+1):(t+k)},D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+k+1):T}|\mathbf{y}_{1:(n-1),1:(t+k)},D_n)}\\ &\propto \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:n,1:t}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:t}|D_n)}\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:n,1:t},D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:(n-1),1:t},D_n)},\end{aligned} \tag{A.7}$$

where we have assumed 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 $D_n > 1$. It turns out that equation (A.7) becomes

$$\begin{aligned}&\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:n,1:t}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:t}|D_n)}\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:n,1:t},D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:(n-1),1:t},D_n)}\\ &= \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:n,1:t}|D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),1:t}|D_n)}\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:n,1:t},D_n)}{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:(n-1),(t+1):(t+k)}|\mathbf{y}_{1:(n-1),1:t},D_n)}.\end{aligned}$$

For $D_n = 1$, $p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n,\mathbf{y}_{1:(n-1),1:T}) = p_{\boldsymbol{\theta}}(\mathbf{y}_{n,1:t}|D_n)$ can be obtained by running the Kalman filter recursion up time t.

Sufficient statistics for segment-specific state space model

Consider the model for the s-th segment and consider the s-th segment with $m = k_s - j_s + 1$ activities that range between j_s and k_s. Denote with

$$\mathbf{Y}_t = \Big[\mathbf{y}_{j_s,t} \quad \mathbf{y}_{j_s+1,t} \quad \cdots \quad \mathbf{y}_{k_s,t}\Big], \qquad \mathbf{A}_t^{(s)} = \boldsymbol{\alpha}_t^{(s)}, \qquad \mathbf{A}_t^{(A)} = \Big[\boldsymbol{\alpha}_{j_s,t} \quad \cdots \quad \boldsymbol{\alpha}_{k_s,t}\Big].$$

The model for the s-th segment has the following completed likelihood

$$\begin{aligned}
&\prod_{t=1}^{T}\Bigg[\det(\boldsymbol{\Sigma})^{-\frac{m}{2}}\det(\mathbf{I}_m)^{-\frac{P}{2}} \\
&\times \exp\Big\{-\frac{1}{2}\mathrm{tr}\big[\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_t - \mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top - \mathbf{Z}^{(A)}\mathbf{A}_t^{(A)})(\mathbf{Y}_t - \mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top - \mathbf{Z}^{(A)}\mathbf{A}_t^{(A)})^\top\big]\Big\}\Bigg] \\
&\prod_{t=1}^{T-1}\det(\boldsymbol{\Psi})^{-\frac{1}{2}}\det(1)^{-\frac{m}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big[\boldsymbol{\Psi}^{-1}(\mathbf{A}_{t+1}^{(s)} - \mathbf{T}^{(S)}\mathbf{A}_t^{(s)})(\mathbf{A}_{t+1}^{(s)} - \mathbf{T}^{(S)}\mathbf{A}_t^{(s)})^\top\big]\Big\} \\
&\det(\mathbf{P}_{1|0}^{(S)})^{-\frac{1}{2}}\det(1)^{-\frac{m}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big[\mathbf{P}_{1|0}^{(S)^{-1}}(\mathbf{A}_1^{(s)} - \widehat{\mathbf{A}}_{1|0}^{(s)})(\mathbf{A}_1^{(s)} - \widehat{\mathbf{A}}_{1|0}^{(s)})^\top\big]\Big\} \\
&\prod_{t=1}^{T-1}\det(\boldsymbol{\Delta})^{-\frac{1}{2}}\det(1)^{-\frac{m}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big[\boldsymbol{\Delta}^{-1}(\mathbf{A}_{t+1}^{(A)} - \mathbf{T}^{(A)}\mathbf{A}_t^{(A)})(\mathbf{A}_{t+1}^{(A)} - \mathbf{T}^{(A)}\mathbf{A}_t^{(A)})^\top\big]\Big\} \\
&\det(\mathbf{P}_{1|0}^{(A)})^{-\frac{m}{2}}\det(\mathbf{I}_m)^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big[\mathbf{P}_{1|0}^{(A)^{-1}}(\mathbf{A}_1^{(A)} - \widehat{\mathbf{A}}_{1|0}^{(A)})(\mathbf{A}_1^{(A)} - \widehat{\mathbf{A}}_{1|0}^{(A)})^\top\big]\Big\}.
\end{aligned}$$

We now associate and compute the expectations for obtaining the set of sufficient statistics for $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Delta}$, $\mathbf{Z}^{(S)}$, $\mathbf{Z}^{(A)}$, $\mathbf{T}^{(S)}$, $\mathbf{T}^{(A)}$, where, for easy of notation, we will drop the conditioning on $S_{j_s:k_s}$ in the expectations.

SS for $\boldsymbol{\Sigma}$ (`SSE`). Let us define the following quantities:

$$\mathtt{SSE} = \sum_{t=1}^{T}\mathrm{E}_{\boldsymbol{\theta}'}\big[(\mathbf{Y}_t - \mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top - \mathbf{Z}^{(A)}\mathbf{A}_t^{(A)})(\mathbf{Y}_t - \mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top - \mathbf{Z}^{(A)}\mathbf{A}_t^{(A)})^\top|\mathbf{Y}_{1:T}\big].$$

then, for each $t = 1, \dots, T$, we need to compute the expectations:

1. $\mathbf{Y}_t\mathbf{Y}_t^\top$;
2. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{Y}_t(\mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top)^\top|\mathbf{Y}_{1:T}\big] = -\mathbf{Y}_t\mathbf{1}_m\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]^\top{\mathbf{Z}^{(S)}}^\top$;
3. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{Y}_t{\mathbf{A}_t^{(A)}}^\top{\mathbf{Z}^{(A)}}^\top|\mathbf{Y}_{1:T}\big] = -\mathbf{Y}_t\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}|\mathbf{Y}_{1:T}\big]^\top{\mathbf{Z}^{(A)}}^\top$;
4. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top\mathbf{Y}_t^\top|\mathbf{Y}_{1:T}\big] = -\mathbf{Z}^{(S)}\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathbf{1}_m^\top\mathbf{Y}_t^\top$;
5. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top\mathbf{1}_m{\mathbf{A}_t^{(s)}}^\top{\mathbf{Z}^{(S)}}^\top|\mathbf{Y}_{1:T}\big]$
$= m\cdot\mathbf{Z}^{(S)}(\mathbf{V}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big] + \mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]^\top){\mathbf{Z}^{(S)}}^\top$;
6. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top(\mathbf{A}_t^{(A)})^\top{\mathbf{Z}^{(A)}}^\top|\mathbf{Y}_{1:T}\big] = \mathbf{Z}^{(S)}\Big[\sum_{j=j_s}^{k_s}\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}\boldsymbol{\alpha}_{j,t}^\top|\mathbf{Y}_{1:T}\big]\Big]{\mathbf{Z}^{(A)}}^\top$
$= \mathbf{Z}^{(S)}\Big[\sum_{j=j_s}^{k_s}\big[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]^\top + \mathrm{Cov}_{\boldsymbol{\theta}'}(\boldsymbol{\alpha}_t^{(s)}, \boldsymbol{\alpha}_{j,t})\big]\Big]{\mathbf{Z}^{(A)}}^\top$;
7. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{Z}^{(A)}\mathbf{A}_t^{(A)}\mathbf{Y}_t^\top|\mathbf{Y}_{1:T}\big] = -\mathbf{Z}^{(A)}\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}|\mathbf{Y}_{1:T}\big]\mathbf{Y}_t^\top$;
8. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{Z}^{(A)}\mathbf{A}_t^{(A)}(\mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}_m^\top)^\top|\mathbf{Y}_{1:T}\big] = \mathbf{Z}^{(A)}\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}\mathbf{1}_m{\mathbf{A}_t^{(s)}}^\top|\mathbf{Y}_{1:T}\big]{\mathbf{Z}^{(S)}}^\top$
$= \mathbf{Z}^{(A)}\mathrm{E}_{\boldsymbol{\theta}'}\big[\sum_{j=j_s}^{k_s}\boldsymbol{\alpha}_{j,t}{\boldsymbol{\alpha}_t^{(s)}}^\top|\mathbf{Y}_{1:T}\big]{\mathbf{Z}^{(S)}}^\top$
$= \mathbf{Z}^{(A)}\Big[\sum_{j=j_s}^{k_s}\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]^\top + \mathrm{Cov}_{\boldsymbol{\theta}'}(\boldsymbol{\alpha}_{j,t}, \boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T})\Big]{\mathbf{Z}^{(S)}}^\top$;
9. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{Z}^{(A)}\mathbf{A}_t^{(A)}{\mathbf{A}_t^{(A)}}^\top{\mathbf{Z}^{(A)}}^\top|\mathbf{Y}_{1:T}\big] = \mathbf{Z}^{(A)}\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}{\mathbf{A}_t^{(A)}}^\top|\mathbf{Y}_{1:T}\big]{\mathbf{Z}^{(A)}}^\top$
$= \mathbf{Z}^{(A)}\Big[\sum_{j=j_s}^{k_s}\mathrm{E}_{\boldsymbol{\theta}'}[\boldsymbol{\alpha}_{j,t}\boldsymbol{\alpha}_{j,t}^\top|\mathbf{Y}_{1:T}]\Big]{\mathbf{Z}^{(A)}}^\top$
$= \mathbf{Z}^{(A)}\Big[\sum_{j=j_s}^{k_s}\mathrm{E}_{\boldsymbol{\theta}'}[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}]\mathrm{E}_{\boldsymbol{\theta}'}[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}]^\top + \mathbf{V}_{\boldsymbol{\theta}'}[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}]\Big]{\mathbf{Z}^{(A)}}^\top$.

SS for $\boldsymbol{\Psi}$ (`SSA`). Let us define

$$\mathtt{SSA} = \sum_{t=1}^{T-1}\mathrm{E}_{\boldsymbol{\theta}'}\big[(\mathbf{A}_{t+1}^{(s)} - \mathbf{T}^{(S)}\mathbf{A}_t^{(s)})(\mathbf{A}_{t+1}^{(s)} - \mathbf{T}^{(S)}\mathbf{A}_t^{(s)})^\top|\mathbf{Y}_{1:T}\big],$$

then, for each $t = 1, \dots, T-1$, we need to compute the following expectations:

1. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(s)}\mathbf{A}_{t+1}^{(s)}{}^{\top}|\mathbf{Y}_{1:T}\big] = \mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]^{\top}$
$+ \mathrm{Var}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]$;
2. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_{t+1}^{(s)}\mathbf{A}_{t}^{(s)}{}^{\top}\mathbf{T}^{(S)}{}^{\top}|\mathbf{Y}_{1:T}\big] = -\Bigg[\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t}^{(s)}|\mathbf{Y}_{1:T}\big]^{\top}$
$+ \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(s)}, \mathbf{A}_{t}^{(s)}|\mathbf{Y}_{1:T}\big]\Bigg]\mathbf{T}^{(S)}{}^{\top}$;
3. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{T}^{(S)}\mathbf{A}_{t}^{(s)}\mathbf{A}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big] = -\mathbf{T}^{(S)}\Bigg[\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t}^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]^{\top}$
$+ \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t}^{(s)}, \mathbf{A}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]\Bigg]$;
4. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{T}^{(S)}\mathbf{A}_{t}^{(s)}\mathbf{A}_{t}^{(s)}{}^{\top}\mathbf{T}^{(S)}{}^{\top}|\mathbf{Y}_{1:T}\big] = \mathbf{T}^{(S)}\Bigg[\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t}^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t}^{(s)}|\mathbf{Y}_{1:T}\big]^{\top}$
$+ \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t}^{(s)}, \mathbf{A}_{t}^{(s)}|\mathbf{Y}_{1:T}\big]\Bigg]\mathbf{T}^{(S)}{}^{\top}$.

SS for $\boldsymbol{\Delta}$ (`SSI`). Define

$$\texttt{SSI} = \sum_{t=1}^{T-1}\mathrm{E}_{\boldsymbol{\theta}'}\big[(\mathbf{A}_{t+1}^{(A)} - \mathbf{T}^{(A)}\mathbf{A}_{t}^{(A)})(\mathbf{A}_{t+1}^{(A)} - \mathbf{T}^{(A)}\mathbf{A}_{t}^{(A)})^{\top}|\mathbf{Y}_{1:T}\big],$$

then, for each $t = 1, \dots, T-1$, we need to compute the following expectations:

1. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_{t+1}^{(A)}\mathbf{A}_{t+1}^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big]$
$= \sum_{j=j_s}^{k_s}\Bigg[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t+1}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t+1}|\mathbf{Y}_{1:T}\big]^{\top} + \mathbf{V}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t+1}|\mathbf{Y}_{1:T}\big]\Bigg]$
2. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_{t+1}^{(A)}\mathbf{A}_{t}^{(A)}{}^{\top}\mathbf{T}^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big]$
$= -\sum_{j=j_s}^{k_s}\Bigg[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t+1}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]^{\top} + \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t+1}, \boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\Bigg]\mathbf{T}^{(A)}{}^{\top}$
3. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{T}^{(A)}\mathbf{A}_{t}^{(A)}\mathbf{A}_{t+1}^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big]$
$= -\mathbf{T}^{(A)}\sum_{j=j_s}^{k_s}\Bigg[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t+1}|\mathbf{Y}_{1:T}\big]^{\top} + \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}, \boldsymbol{\alpha}_{j,t+1}|\mathbf{Y}_{1:T}\big]\Bigg]$
4. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{T}^{(A)}\mathbf{A}_{t}^{(A)}\mathbf{A}_{t}^{(A)}{}^{\top}\mathbf{T}^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big]$
$= \mathbf{T}^{(A)}\sum_{j=j_s}^{k_s}\Bigg[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]^{\top} + \mathbf{V}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\Bigg]\mathbf{T}^{(A)}{}^{\top}$.

SS for $\mathbf{Z}^{(S)}$ (`SSZs_num` and `SSZs_den`). Let us define

$$\mathtt{SSZs\_num} = \sum_{t=1}^{T} \mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(s)}\mathbf{1}_m^\top(\mathbf{Y}_t - \mathbf{Z}^{(A)}\mathbf{A}_t^{(A)})^\top|\mathbf{Y}_{1:T}\big]$$
$$\mathtt{SSZs\_den} = \sum_{t=1}^{T} \mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}\mathbf{1}_m^\top\mathbf{1}_m{\mathbf{A}_t^{(s)}}^\top|\mathbf{Y}_{1:T}\big],$$

then, we need to compute the following expectations:

1. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(s)}\mathbf{1}_m^\top\mathbf{Y}_t^\top|\mathbf{Y}_{1:T}\big] = -\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathbf{1}_m^\top\mathbf{Y}_t^\top$
2. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}\mathbf{1}_m^\top{\mathbf{A}_t^{(A)}}^\top{\mathbf{Z}^{(A)}}^\top|\mathbf{Y}_{1:T}\big] = \sum_{j=j_s}^{k_s}\big[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]^\top$

   $+ \quad \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}, \boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\big]{\mathbf{Z}^{(A)}}^\top$
3. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}\mathbf{1}_m^\top\mathbf{1}_m\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big] = m\cdot\Big[\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]^\top + \mathbf{V}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(s)}|\mathbf{Y}_{1:T}\big]\Big].$

SS for $\mathbf{Z}^{(A)}$ (`SSZi_num` and `SSZi_den`). Let us define the following quantities

$$\mathtt{SSZi\_num} = \sum_{t=1}^{T} \mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(A)}(\mathbf{Y}_t - \mathbf{Z}^{(S)}\mathbf{A}_t^{(s)}\mathbf{1}^\top)^\top|\mathbf{Y}_{1:T}\big]$$
$$\mathtt{SSZi\_den} = \sum_{t=1}^{T} \mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}{\mathbf{A}_t^{(A)}}^\top|\mathbf{Y}_{1:T}\big],$$

then, we need to compute the following expectations:

1. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(A)}\mathbf{Y}_t^\top|\mathbf{Y}_{1:T}\big] = -\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}|\mathbf{Y}_{1:T}\big]\mathbf{Y}_t^\top$
2. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}\mathbf{1}_m{\mathbf{A}_t^{(s)}}^\top{\mathbf{Z}^{(S)}}^\top|\mathbf{Y}_{1:T}\big]$

   $= \Bigg[\sum_{j=j_s}^{k_s}\big[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]^\top + \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}, \boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]\big]\Bigg]{\mathbf{Z}^{(S)}}^\top$
3. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}{\mathbf{A}_t^{(A)}}^\top|\mathbf{Y}_{1:T}\big] = \sum_{j=j_s}^{k_s}\Big[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]^\top + \mathbf{V}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\Big].$

SS for $\mathbf{T}^{(S)}$ (`SSTs_num` and `SSTs_den`). Let us define

$$\mathtt{SSTs\_num} = \sum_{t=1}^{T-1} \mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(S)}{\mathbf{A}_{t+1}^{(S)}}^\top|\mathbf{Y}_{1:T}\big]$$
$$\mathtt{SSTs\_den} = \sum_{t=1}^{T-1} \mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(S)}{\mathbf{A}_t^{(S)}}^\top|\mathbf{Y}_{1:T}\big].$$

then we need to compute the following expectations:

1. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(S)}\mathbf{A}_{t+1}^{(S)}{}^{\top}|\mathbf{Y}_{1:T}\big] = -\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]^{\top} - \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}, \boldsymbol{\alpha}_{t+1}^{(s)}|\mathbf{Y}_{1:T}\big]$
2. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}\boldsymbol{\alpha}_t^{(s)}{}^{\top}|\mathbf{Y}_{1:T}\big] = \mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big]^{\top} + \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_t^{(s)}, \boldsymbol{\alpha}_t^{(s)}|\mathbf{Y}_{1:T}\big],$

for each $t = 1, \ldots, T-1$.

SS for $\mathbf{T}^{(A)}$ (`SSTi_num` and `SSTi_den`). Let us define

$$\begin{aligned}\texttt{SSTi\_num} &= \sum_{t=1}^{T-1} \mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(A)}\mathbf{A}_{t+1}^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big]\\ \texttt{SSTs\_den} &= \sum_{t=1}^{T-1} \mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}\mathbf{A}_t^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big],\end{aligned}$$

then we need to compute the following expectations

1. $\mathrm{E}_{\boldsymbol{\theta}'}\big[-\mathbf{A}_t^{(A)}\mathbf{A}_{t+1}^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big] = -\sum_{j=j_s}^{k_s}\Big[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t+1}^{\top}|\mathbf{Y}_{1:T}\big]$
   $+ \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}, \boldsymbol{\alpha}_{j,t+1}|\mathbf{Y}_{1:T}\big]\Big]$
2. $\mathrm{E}_{\boldsymbol{\theta}'}\big[\mathbf{A}_t^{(A)}\mathbf{A}_t^{(A)}{}^{\top}|\mathbf{Y}_{1:T}\big] = \sum_{j=j_s}^{k_s}\Big[\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\mathrm{E}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}^{\top}|\mathbf{Y}_{1:T}\big] + \mathrm{Cov}_{\boldsymbol{\theta}'}\big[\boldsymbol{\alpha}_{j,t}, \boldsymbol{\alpha}_{j,t}|\mathbf{Y}_{1:T}\big]\Big],$

for each $t = 1, \ldots, T-1$. In all previous equations, the expectations involving the latent states are provided by the Kalman filter, smoother and lagged smoother recursions provided by algorithm 1.

Maximization step

Maximization step for simulation studies

In the simulation study it is required to estimate the parameters $\boldsymbol{\theta} = (\sigma_\epsilon^2, \sigma_\alpha^2, \sigma_d^2, \rho)$. The maximization step, with m time series of length T and of P measurement equations,

$$\begin{gathered}\widehat{\sigma}_\epsilon^2 = \mathrm{tr}(\mathrm{SSE})/(mTP), \quad \widehat{\sigma}_\alpha^2 = \mathrm{tr}(\mathrm{SSA}(\mathbf{I}_P \otimes \boldsymbol{\Psi}_0^{-1}))/(2P(T-1))\\ \widehat{\sigma}_d^2 = \mathrm{tr}(\mathrm{SSI})/(mP(T-1)), \quad \widehat{\rho} = -\mathrm{tr}(\mathrm{SSI\_num})/\mathrm{tr}(\mathrm{SSI\_den}),\end{gathered}$$

where SSE, SSA, SSI, SSI_num and SSI_den have been calculated in the previous section.

Maximization step for real data application

In the application it is required to estimate the parameters $\boldsymbol{\theta} = (\boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{\Delta}, \rho_{\text{sp}})$. The maximization step, with m time series of length T and of P measurements, computes

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}} &= \text{SSE}/(mT), \quad \widehat{\boldsymbol{\Psi}} = \text{SSA}/(T-1), \\ \widehat{\boldsymbol{\Delta}} &= \text{SSI}/(m(T-1)), \quad \rho_{\text{sp}} = -\text{SSI\_num}[1,1]/\text{SSI\_den}[1,1], \end{aligned}$$

where SSE, SSA, SSI, SSI_num and SSI_den have been calculated in the previous section.

Selection of the tuning parameters

This Section presents additional results concerning the choice of the tuning parameters (k^{*}, λ). Specifically, for selected running activities, we provide the changepoint probabilities for different choices of $k^{\star} = (0, 5, 10, 15, 30, 600)$ and $\lambda = (0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99)$. Results are illustrated in Figure B.1-B.3. The first two figures have been obtained by running the within online EM algorithm several times, one for each combination of $(k^{\star}, \lambda)$. The main findings outline that, in our setting, the values of $(k^{\star}, \lambda)$ do not highly affect the results in terms of real time detection of a changepoint. The third figure is obtained by running the between online EM algorithm, by varying the value of λ. This figure highlights the robustness of our result with respect to the choice of λ. Indeed, only extreme values of λ change slightly the segmentation of the activities.

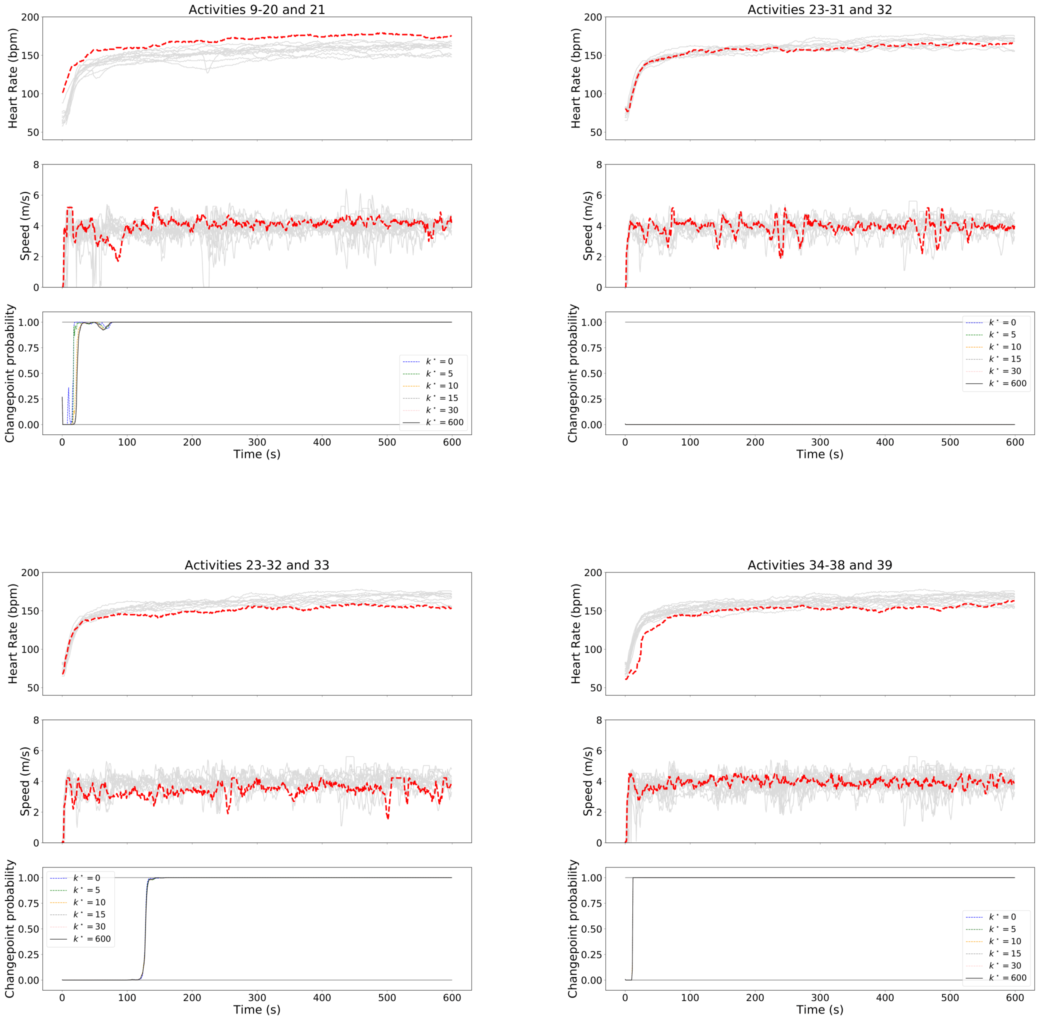

Figure B.1: Changepoint probabilities for different choices of $k^\star$.

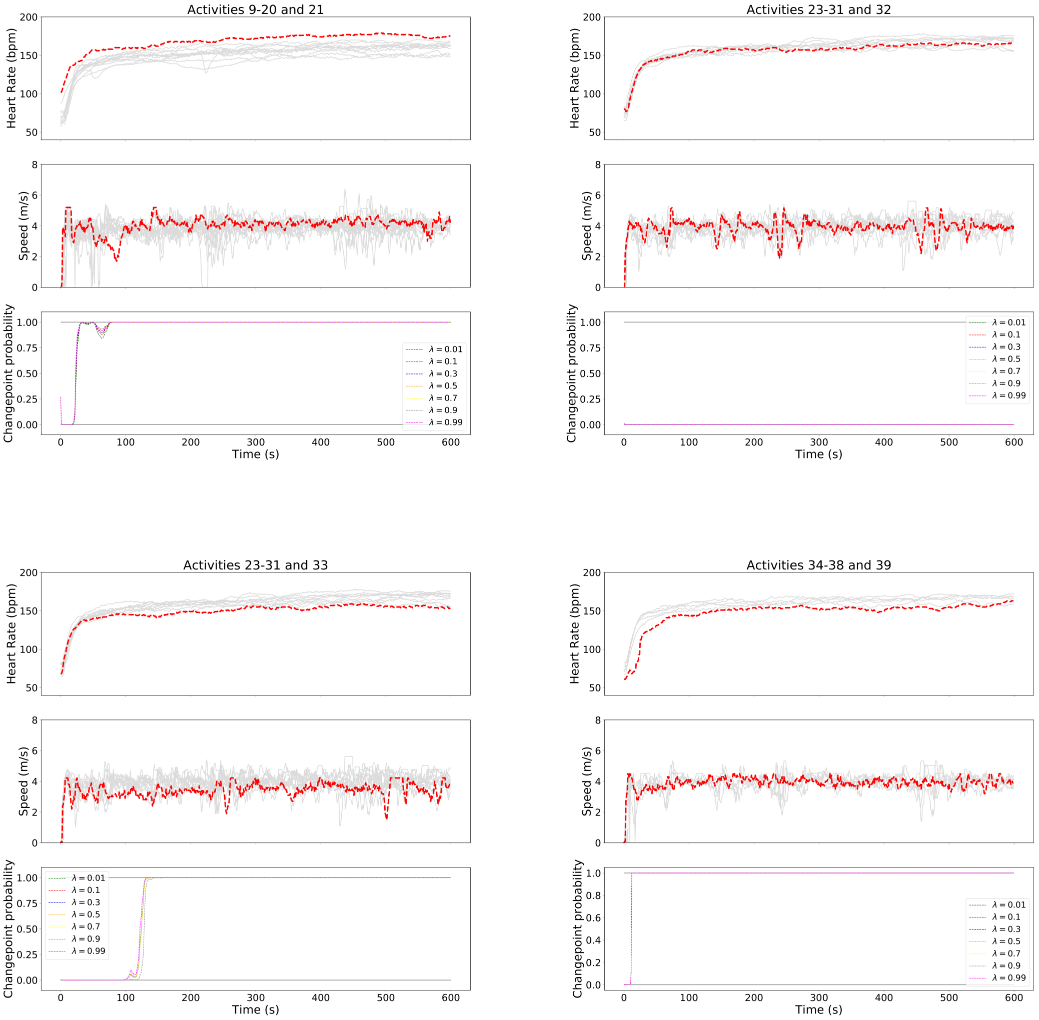

FIGURE B.2: Changepoint probabilities for different choice of λ.

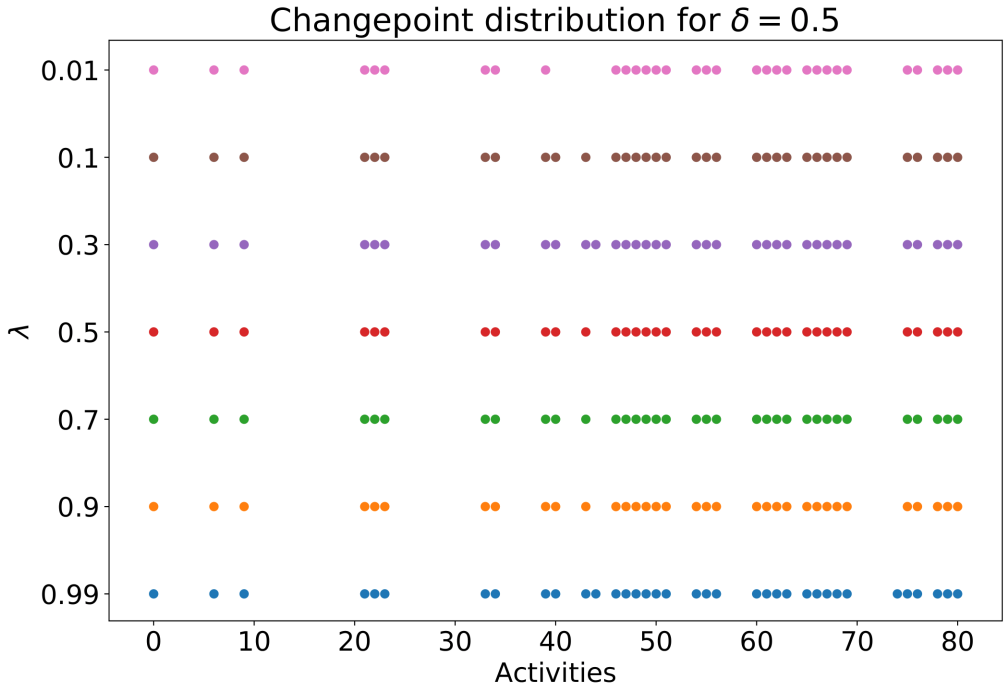

Figure B.3: Changepoint distribution for different choices of λ.

Algorithms

```
Result: Smoothed States $\widehat{\boldsymbol{\alpha}}_{t|T}$, Smoothed Variances $\mathbf{P}_{t|T}$, Smoothed Lagged Covariances
        $\mathbf{P}_{t,t-1|T}$, for $t = 1, \dots, T$ stored in the respective array State, Cov, Cov_lag, and
        the conditional log-likelihood llik
Input:
data: $\mathbf{Y}_{1:T} = \mathbf{y}_{j_s:k_s,1:T}$;
StrMatr: $\mathbf{Z}^{(S)}$, $\mathbf{T}^{(S)}$, $\mathbf{Z}^{(A)}$, $\mathbf{T}^{(A)}$;
param: $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Delta}$;
InitVal: $\widehat{\boldsymbol{\alpha}}^{(S)}_{1|0}$, $\mathbf{P}^{(S)}_{1|0}$, $\widehat{\boldsymbol{\alpha}}^{(A)}_{1|0}$, $\mathbf{P}^{(A)}_{1|0}$;
Initialization:
$m = k_s - j_s + 1$; $\texttt{llik} = 0$; $\mathbf{Z} = \begin{bmatrix}\mathbf{1}_m \otimes \mathbf{Z}^{(S)} & \mathbf{I}_m \otimes \mathbf{Z}^{(A)}\end{bmatrix}$; $\mathbf{T} = \text{blkdiag}(\mathbf{T}^{(S)}, \mathbf{I}_m \otimes \mathbf{T}^{(A)})$;
$\mathbf{H} = \mathbf{I}_m \otimes \boldsymbol{\Sigma}$; $\mathbf{G} = \text{blkdiag}(\boldsymbol{\Psi}, \mathbf{I}_m \otimes \boldsymbol{\Delta})$; $\widehat{\boldsymbol{\alpha}}_{1|0} = (\widehat{\boldsymbol{\alpha}}^{(S)}_{1|0}{}^\top, \mathbf{1}_m^\top \otimes \widehat{\boldsymbol{\alpha}}^{(A)\top}_{1|0})^\top$;
 $\mathbf{P}_{1|0} = \text{blkdiag}(\mathbf{P}^{(S)}_{1|0}, \mathbf{I}_m \otimes \mathbf{P}^{(A)}_{1|0})$;
Kalman Filter:
for $t$ = 1,2, ..., $T$ do
    $\boldsymbol{v}_t = \mathbf{Y}_t - \mathbf{Z}\widehat{\boldsymbol{\alpha}}_{t|t-1}$
    $\mathbf{F}_t = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}^\top + \mathbf{H}$
    $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{Z}^\top\mathbf{F}_t^{-1}$
    $\widehat{\boldsymbol{\alpha}}_{t|t} = \widehat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{K}_t\boldsymbol{v}_t$
    $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}^\top\mathbf{F}_t^{-1}\mathbf{Z}\mathbf{P}_{t|t-1}$
    $\widehat{\boldsymbol{\alpha}}_{t+1|t} = \mathbf{T}\widehat{\boldsymbol{\alpha}}_{t|t-1}$
    $\mathbf{P}_{t+1|t} = \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{T}^\top + \mathbf{G}$,
    $\texttt{llik} = \texttt{llik} - \frac{mP}{2}\log(2\pi) - \frac{1}{2}\big(\text{logdet}(\mathbf{F}_t) - \boldsymbol{v}_t^\top\mathbf{F}_t^{-1}\boldsymbol{v}_t\big)$
end
Kalman Filter:
for $t$ = $T$,..., 1 do
    $\mathbf{J}_{t-1} = \mathbf{P}_{t-1|t-1}\mathbf{T}^\top\mathbf{P}_{t|t-1}^{-1}$
    $\widehat{\boldsymbol{\alpha}}_{t-1|T} = \widehat{\boldsymbol{\alpha}}_{t-1|t-1} + \mathbf{J}_{t-1}(\widehat{\boldsymbol{\alpha}}_{t|T} - \widehat{\boldsymbol{\alpha}}_{t|t-1})$
    $\mathbf{P}_{t-1|T} = \mathbf{P}_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{P}_{t|T} - \mathbf{P}_{t|t-1})\mathbf{J}_{t-1}^\top$
end
Lag-one Covariance Smoother:
$\mathbf{P}_{T,T-1|T} = \mathbf{T}\mathbf{P}_{T-1|T-1} - \mathbf{K}_T\mathbf{Z}\mathbf{T}\mathbf{P}_{T-1|T-1}$
for $t$ = $T$,..., 2 do
    $\mathbf{P}_{t-1,t-2|T} = \mathbf{P}_{t-1|t-1}\mathbf{J}_{t-2}^\top + \mathbf{J}_{t-1}(\mathbf{P}_{t,t-1|T} - \mathbf{T}\mathbf{P}_{t-1|t-1})\mathbf{J}_{t-2}^\top$
end
```

Algorithm 1: Computing Kalman quantities `State, Cov, Cov_lag, llik = KalRec(data, StrMatr, param, InitVal)`

The within online EM algorithm is a modification of the between online version, where the computation of the filtered probabilities in equation (A.8) are modified according to equation (A.3).

```
Result: Filtered probabilities FiltProb and parameters trace tr_params
Input: data: $\mathbf{Y}_{1:T} = \mathbf{y}_{1:N,1:T}$; StrMatr: $\mathbf{Z}^{(S)}$, $\mathbf{T}^{(S)}$, $\mathbf{Z}^{(A)}$, $\mathbf{T}^{(A)}$;
param0: $\boldsymbol{\theta}_0$, and the function $g(\cdot)$ that links $\boldsymbol{\theta}$ with the elements of the of the model ($\lambda_{\boldsymbol{\theta}}$,
  $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Delta}$, $\mathbf{Z}^{(S)}$, $\mathbf{T}^{(S)}$, $\mathbf{Z}^{(A)}$, $\mathbf{T}^{(A)}$); InitVal: $\widehat{\boldsymbol{\alpha}}^{(S)}_{1|0}$, $\mathbf{P}^{(S)}_{1|0}$, $\widehat{\boldsymbol{\alpha}}^{(A)}_{1|0}$, $\mathbf{P}^{(A)}_{1|0}$; maximizer: the
  function $\boldsymbol{\Lambda}$; stepwise: the function $\gamma_n$, $n \geq 1$; B: the value $B$;
EM Algorithm
for $n$ = 1,..., $N$ do
    E-step:
    if $n = 1$ then
        - Initialize $\widehat{\boldsymbol{\theta}}_1 = \boldsymbol{\theta}_0$; set $\mathbf{T}_1^{\widehat{\boldsymbol{\theta}}_1}(D_1) = \iota_{\widehat{\boldsymbol{\theta}}_1}(\mathbf{y}_{1,1:T})$; set $\mathcal{D}_1^B = \{1, \dots, 1\}$, and
          $\eta_1^B(D_1) = 1$, for $D_1 = 1$;
    if $n > 1$ then
        - Create $\mathcal{D}_n^{B\star} = \{(1, d_{n-1}^1), (d_{n-1}^1 + 1, d_{n-1}^1), \dots, (1, d_{n-1}^B), (d_{n-1}^B + 1, d_{n-1}^B)\}$;
        - For all $(d_n^b, d_{n-1}^b) \in \mathcal{D}_n^{B\star}$ compute the weights

            $W(d_n^b, d_{n-1}^b) = p_{\widehat{\boldsymbol{\theta}}_{n-1}}(d_n^b|d_{n-1}^b) G^D_{\widehat{\boldsymbol{\theta}}_{n-1},n-1}(d_{n-1}^b)\eta_{n-1}^B(d_{n-1}^b)$

          to sample $B$ independent particles and store them in $\mathcal{D}^B_{(n,n-1)}$;
        - Obtain from $\mathcal{D}^B_{(n,n-1)}$ the marginal $\eta_n^B(D_n) = \sum_{b=1}^B \delta_{D_n}(d_n, d_{n-1})$ with support
          $\mathcal{D}_n^B = \{d_n^1, \dots, d_n^B\}$ and $\delta_{D_n}(d_n, d_{n-1}) = 1$ if $D_n = d_n$, 0 otherwise;
        - Compute

            $p_{\widehat{\boldsymbol{\theta}}_{1:(n-1)}}(D_n|\mathbf{y}_{1:n,1:T}) = \frac{\sum_{D_{n-1} \in \mathcal{D}^B_{n-1}} G^D_{\widehat{\boldsymbol{\theta}}_{n-1},n}(D_n) p_{\widehat{\boldsymbol{\theta}}_{n-1}}(D_n|D_{n-1})\eta^B_{n-1}(D_{n-1})}{\sum_{(D'_n, D'_{n-1}) \in \mathcal{D}^B_{(n,n-1)}} G^D_{\widehat{\boldsymbol{\theta}}_{n-1},n}(D'_n) p_{\widehat{\boldsymbol{\theta}}_{n-1}}(D'_n|D'_{n-1})\eta^B_{n-1}(D'_{n-1})}$    (A.8)

          and

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

          with supports $\mathcal{D}_n^B$ and $\mathcal{D}^B_{n-1|n}$ respectively ;
        - Compute

            $\mathbf{T}^{\widehat{\boldsymbol{\theta}}_{n-1}}_{\gamma,n}(D_{1:n}) = \sum_{D_{n-1} \in \mathcal{D}^B_{n-1|n}} \left[(1-\gamma_n)\mathbf{T}^{\widehat{\boldsymbol{\theta}}_{n-2}}_{\gamma,n-1}(D_{1:(n-1)}) + \gamma_n \iota_{\widehat{\boldsymbol{\theta}}_{n-1}}(\mathbf{y}_{n,1:T})\right]$
                $\times p_{\widehat{\boldsymbol{\theta}}_{1:(n-1)}}(D_{n-1}|\mathbf{y}_{1:(n-1),1:T}, D_n)$

          and

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

    M-step
    Compute $\widehat{\boldsymbol{\theta}}_n = \boldsymbol{\Lambda}(\mathcal{Q}_n)$ and set $(\lambda_{\widehat{\boldsymbol{\theta}}_n}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{\Delta}, \mathbf{Z}^{(S)}, \mathbf{T}^{(S)}, \mathbf{Z}^{(A)}, \mathbf{T}^{(A)}) = g(\widehat{\boldsymbol{\theta}}_n)$
end
```

Algorithm 2: Between-Online changepoints detection `FiltProb, tr_params = ONChpntDet(data, StrMatr, param0, InitVal, maximizer, B)`

Bibliography

Abbiss, C. R. and Laursen, P. B. (2008) Describing and Understanding Pacing Strategies during Athletic Competition. *Sports Medicine* **38**(3), 239–252.

Allen, T. and Goff, J. E. (2018) Resources for sports engineering education. *Sports Engineering* **21**(4), 245–253.

Alvero-Cruz, J. R., Carnero, E. A., García, M. A. G., Alacid, F., Correas-Gómez, L., Rosemann, T., Nikolaidis, P. T. and Knechtle, B. (2020) Predictive performance models in long-distance runners: A narrative review. *International Journal of Environmental Research and Public Health* **17**(21), 8289.

Aminikhanghahi, S. and Cook, D. J. (2017) A survey of methods for time series change point detection. *Knowledge and Information Systems* **51**(2), 339–367.

Baca, A. (2014) *Computer science in sport: research and practice.* Routledge.

Bartolucci, F. and Murphy, T. B. (2015) A finite mixture latent trajectory model for modeling ultrarunners' behavior in a 24-hour race. *Journal of Quantitative Analysis in Sports* **11**(4), 193–203.

Boccia, G., Moisè, P., Franceschi, A., Trova, F., Panero, D., La Torre, A., Rainoldi, A., Schena, F. and Cardinale, M. (2017) Career performance trajectories in track and field jumping events from youth to senior success: The importance of learning and development. *PLOS ONE* **12**(1), 1–15. Publisher: Public Library of Science.

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. *International journal of sports physiology and performance* **12**(S2), 161–170.

Buchheit, M. (2014) Monitoring training status with hr measures: do all roads lead to Rome? *Frontiers in physiology* **5**, 73.

Bussmann, G. (1999) How to prevent "dropout" in competitive sport. *IAAF new studies in athletics* **14**(1), S. 23–29.

Camomilla, V., Bergamini, E., Fantozzi, S. and Vannozzi, G. (2018) Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review. *Sensors* **18**(3).

Cardinale, M. and Varley, M. C. (2017) Wearable training-monitoring technology: applications, challenges, and opportunities. *International journal of sports physiology and performance* **12**(s2), S2–55.

Caron, F., Doucet, A. and Gottardo, R. (2012) On-line changepoint detection and parameter estimation with application to genomic data. *Statistics and Computing* **22**(2), 579–595.

Carvalho, C. M. and West, M. (2007) Dynamic matrix-variate graphical models. *Bayesian Analysis* **2**(1), 69 – 97.

Cece, V., Guillet-Descas, E., Nicaise, V., Lienhart, N. and Martinent, G. (2019) Longitudinal trajectories of emotions among young athletes involving in intense training centres: Do emotional intelligence and emotional regulation matter? *Psychology of Sport and Exercise* **43**, 128–136.

de Chaumaray, M. D. R., Marbac, M. and Navarro, F. (2020) Mixture of hidden Markov models for accelerometer data. *The Annals of Applied Statistics* **14**(4), 1834–1855.

Chen, E. Y., Tsay, R. S. and Chen, R. (2020) Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association* **115**(530), 775–793.

Chen, R., Xiao, H. and Yang, D. (2021a) Autoregressive models for matrix-valued time series. *Journal of Econometrics* **222**(1, part B), 539–560.

Chen, R., Yang, D. and Zhang, C.-H. (2021b) Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association* **0**(0), 1–23.

Chib, S. (1998) Estimation and comparison of multiple change-point models. *Journal of Econometrics* **86**(2), 221–241.

Choukroun, D., Weiss, H., Bar-Itzhack, I. Y. and Oshman, Y. (2006) Kalman filtering for matrix estimation. *IEEE Transactions on Aerospace and Electronic Systems* **42**(1), 147–159.

Coh, M., Zvan, M., Boncina, N. and Stuhec, S. (2019) Biomechanical model of hurdle clearance in 100m hurdle races: a case study. *Journal of Anthropology of Sport and Physical Education* **3**(4), 3–6.

Cook, R. D., Li, B., Chiaromonte, F. and Su, Z. (2010) Envelope models for parsimonious and efficient multivariate linear. *Statistica Sinica* **20**(3), 999–1010.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological* **39**(1), 1–38. With discussion.

Ding, S. and Cook, D. R. (2018) Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(2), 387–408.

Dolmeta, P., Argiento, R. and Montagna, S. (2021) Bayesian garch modeling of functional sports data. *arXiv preprint arXiv:2101.08175* .

Dong, J. G. (2016) The role of heart rate variability in sports physiology. *Experimental and therapeutic medicine* **11**(5), 1531–1536.

Durbin, J. and Koopman, S. J. (2002) A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89**(3), 603–615.

Durbin, J. and Koopman, S. J. (2012) *Time series analysis by state space methods*. Second edition, volume 38 of *Oxford Statistical Science Series*. Oxford University Press, Oxford. ISBN 978-0-19-964117-8.

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**(6), 435–453.

Emig, T. and Peltonen, J. (2020) Human running performance from real-world big data. *Nature Communications* **11**(1), 4936.

Enoksen, E. (2011) Drop-out rate and drop-out reasons among promising norwegian track and field athletes: A 25 year study. In *Scandinavian sport studies forum*. Malmö University.

Faiss, R., Saugy, J., Zollinger, A., Robinson, N., Schuetz, F., Saugy, M. and Garnier, P.-Y. (2020) Prevalence estimate of blood doping in elite track and field athletes during two major international events. *Frontiers in Physiology* **11**, 160.

Fearnhead, P. and Liu, Z. (2007) On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(4), 589–605.

Fister, I. J., Rauter, S., Fister, D. and Fister, I. (2017) A collection of sport activity datasets for data analysis and data mining 2017a. *Technical report 2017a* .

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 Medicine* **10**(1), 1–45.

Frick, H. and Kosmidis, I. (2017) tracker: Infrastructure for running and cycling data from gps-enabled tracking devices in R. *Journal of Statistical Software* **82**(7), 1–29.

Frühwirth-Schnatter, S., Malsiner-Walli, G. and Grün, B. (2020) Generalized mixtures of finite mixtures and telescoping sampling. *arXiv preprint arXiv:2005.09918* .

Frühwirth-Schnatter, S. (2011) Panel data analysis: a survey on model-based clustering of time series. *Advances in Data Analysis and Classification* **5**(4), 251–280.

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. *TrAC Trends in Analytical Chemistry* **135**, 116148.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014) *Bayesian data analysis.* Third edition. Texts in Statistical Science Series. CRC Press, Boca Raton, FL. ISBN 978-1-4398-4095-5.

van Gent, R. N., Siem, D., van Middelkoop, M., van Os, A. G., Bierma-Zeinstra, S. M. A. and Koes, B. W. (2007) Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review. *British journal of sports medicine* **41**(8), 469–480. Edition: 2007/05/01 Publisher: BMJ Group.

George, E. I. (2006) *Bayesian Model Selection.* John Wiley Sons, Ltd. ISBN 9780471667193.

Grathwohl, W., Swersky, K., Hashemi, M., Duvenaud, D. and Maddison, C. J. (2021) Oops i took a gradient: Scalable sampling for discrete distributions. *arXiv preprint arXiv:2102.04509* .

Grundy, T., Killick, R. and Mihaylov, G. (2020) High-dimensional changepoint detection via a geometrically inspired mapping. *Statistics and Computing* **30**(4), 1155–1166.

Gupta, A. K. and Nagar, D. K. (2000) *Matrix variate distributions.* Volume 104 of *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics.* Chapman & Hall/CRC, Boca Raton, FL. ISBN 1-58488-046-5.

Halson, S. L. (2014) Monitoring training load to understand fatigue in athletes. *Sports medicine (Auckland, N.Z.)* **44 Suppl 2**(Suppl 2), S139–S147. Publisher: Springer International Publishing.

Halson, S. L., Peake, J. M. and Sullivan, J. P. (2016) Wearable technology for athletes: Information overload and pseudoscience? *International Journal of Sports Physiology and Performance* **11**(6), 705 – 706.

Haugen, T. A., Solberg, P. A., Foster, C., Morán-Navarro, R., Breitschädel, F. and Hopkins, W. G. (2018) Peak age and performance progression in world-class track-and-field athletes. *International Journal of Sports Physiology and Performance* **13**(9), 1122 – 1129.

Haynes, K., Fearnhead, P. and Eckley, I. A. (2017) A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing* **27**(5), 1293–1305.

Hernandez, A. E., Mattarella-Micke, A., Redding, R. W., Woods, E. O. and Beilock, S. (2011) Age of acquisition in sport: Starting early matters. *The American journal of psychology* **124**(3), 253.

Hill, D. W. (1993) The critical power concept. *Sports medicine* **16**(4), 237–254.

Hsu, N.-J., Huang, H.-C. and Tsay, R. S. (2021) Matrix autoregressive spatio-temporal models. *Journal of Computational and Graphical Statistics* **0**(0), 1–13.

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. *Journal of the American Statistical Association* **114**(526), 553–564.

Jacques, J. and Samardžić, S. (2021) Analyzing cycling sensors data through ordinal logistic regression with functional covariates. Working paper or preprint.

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. *Statistical Science. A Review Journal of the Institute of Mathematical Statistics* **30**(3), 328–351.

Karlis, D., Ntzoufras, I. and Repoussis, P. (2021) Mathematics meet sports. *IMA Journal of Management Mathematics* **32**(4), 381–383.

Kasper, K. (2019) Sports Training Principles. *Current Sports Medicine Reports* **18**(4).

Kenneally, M., Casado, A. and Santos-Concejero, J. (2018) The effect of periodization and training intensity distribution on middle- and long-distance running performance:

A systematic review. *International Journal of Sports Physiology and Performance* **13**(9), 1114 – 1121.

Koenker, R. (2005) *Quantile regression.* Volume 38 of *Econometric Society Monographs.* Cambridge University Press, Cambridge. ISBN 978-0-521-60827-5; 0-521-60827-9.

Kosmidis, I. and Passfield, L. (2015) Linking the performance of endurance runners to training and physiological effects via multi-resolution elastic net. *arXiv preprint arXiv:1506.01388* .

Kunkel, D. and Peruggia, M. (2020) Anchored Bayesian Gaussian mixture models. *Electronic Journal of Statistics* **14**(2), 3869–3913.

Lam, C., Yao, Q. and Bathia, N. (2011) Estimation of latent factors for high-dimensional time series. *Biometrika* **98**(4), 901–918.

Larsen, C. and Alfermann, D. (2017) *Understanding dropout in the athlete development process*, pp. 325–335. Routledge International Handbooks. United Kingdom: Routledge. ISBN 9781138951778. Forventes at udkomme 23 March 2017.

Leroy, A., Marc, A., Dupas, O., Rey, J. L. and Gey, S. (2018) Functional data analysis in sport science: Example of swimmers' progression curves clustering. *Applied Sciences* **8**(10), 1766.

Liu, J. S. (1994) The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* **89**(427), 958–966.

Mackie, J. (2016) *cycleRtools: Tools for Cycling Data Analysis.* R package version 1.1.1.

Magnus, J. R. and Neudecker, H. (2019) *Matrix differential calculus with applications in statistics and econometrics.* John Wiley & Sons.

Maharaj, E. A., D'Urso, P. and Caiado, J. (2019) *Time series clustering and classification.* Chapman and Hall/CRC.

Malsiner-Walli, G., Frühwirth-Schnatter, S. and Grün, B. (2017) Identifying mixtures of mixtures using bayesian estimation. *Journal of Computational and Graphical Statistics* **26**(2), 285–295. PMID: 28626349.

Mooses, M., Jürimäe, J., Mäestu, J., Purge, P., Mooses, K. and Jürimäe, T. (2013) Anthropometric and physiological determinants of running performance in middle- and long-distance runners. *Kinesiology* **45**(2), 154–162.

Müller, H. and Glad, B. (2014) Technology in athletics. *New studies in Athletics* **29**(3), 7.

Nobile, A. and Fearnside, A. T. (2007) Bayesian finite mixtures with an unknown number of components: the allocation sampler. *Statistics and Computing* **17**(2), 147–162.

Paulich, M., Schepers, M., Rudigkeit, N. and Bellusci, G. (2018) Xsens mtw awinda: Miniature wireless inertial-magnetic motion tracker for highly accurate 3d kinematic applications. *Xsens: Enschede, The Netherlands* pp. 1–9.

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). *European heart journal* **42**(1), 17–96.

Pinot, J. and Grappe, F. (2015) A six-year monitoring case study of a top-10 cycling grand tour finisher. *Journal of sports sciences* **33**(9), 907–914.

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

Pol, R., Balagué, N., Ric, A., Torrents, C., Kiely, J. and Hristovski, R. (2020) Training or Synergizing? Complex Systems Principles Change the Understanding of Sport Processes. *Sports Medicine - Open* **6**(1), 28.

Pradier, F. M., Ruiz, J. R., Francisco and Perez-Cruz, F. (2016) Prior Design for Dependent Dirichlet Processes: An Application to Marathon Modeling. *PLOS ONE* **11**(1), 1–28. Publisher: Public Library of Science.

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**(8), e13494.

Puchowicz, M. J., Mizelman, E., Yogev, A., Koehle, M. S., Townsend, N. E. and Clarke, D. C. (2018) The critical power model as a potential tool for anti-doping. *Frontiers in Physiology* **9**, 643.

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**(3), 507–527.

Quintana, J. M. (1987) *Multivariate Bayesian Forecasting Models.* Ph.D. thesis, University of Warwick.

R Core Team (2020) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rana, M. and Mittal, V. (2021) Wearable sensors for real-time kinematics analysis in sports: A review. *IEEE Sensors Journal* **21**(2), 1187–1207.

Rauch, H. E., Tung, F. and Striebel, C. T. (1965) Maximum likelihood estimates of linear dynamic systems. *AIAA journal* **3**(8), 1445–1450.

Rauter, S. and Fister, I. (2015) *A collection of sport activity files for data analysis and data mining.* Ph.D. thesis, Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko.

Richter, C., O'Reilly, M. and Delahunt, E. (2021) Machine learning in sports science: challenges and opportunities. *Sports Biomechanics* **0**(0), 1–7. PMID: 33874846.

Robert, C. P. and Casella, G. (2004) *Monte Carlo statistical methods.* Second edition. Springer Texts in Statistics. Springer-Verlag, New York. ISBN 0-387-21239-6.

Rogers, M., Li, L. and Russell, S. J. (2013) Multilinear dynamical systems for tensor time series. In *Advances in Neural Information Processing Systems*, eds C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, volume 26. Curran Associates, Inc.

Samé, A. and Govaert, G. (2017) Segmental dynamic factor analysis for time series of curves. *Statistics and Computing* **27**(6), 1617–1637.

Sandford, G. N. and Stellingwerff, T. (2019) "Question your categories": the misunderstood complexity of middle-distance running profiles with implications for research methods and application. *Frontiers in Sports and Active Living* **1**, 28.

Santos-Fernandez, E., Wu, P. and Mengersen, K. L. (2019) Bayesian statistics meets sports: a comprehensive review. *Journal of Quantitative Analysis in Sports* **15**(4), 289–312.

Sargent, C., Lastella, M., Halson, S. L. and Roach, G. D. (2014) The impact of training schedules on the sleep and fatigue of elite athletes. *Chronobiology International* **31**(10), 1160–1168.

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. *Frontiers in physiology* **9**, 639.

Schütz, F. and Zollinger, A. (2018) Abps: an r package for calculating the abnormal blood profile score. *Frontiers in physiology* **9**, 1638.

Severini, T. A. (2020) *Analytic methods in sports: Using mathematics and statistics to understand data from baseball, football, basketball, and other sports.* Crc Press.

Shumway, R. H. and Stoffer, D. S. (2017) *Time series analysis and its applications.* Fourth edition. Springer Texts in Statistics. Springer, Cham. ISBN 978-3-319-52451-1; 978-3-319-52452-8. With R examples.

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. *Arrhythmia & electrophysiology review* **7**(4), 247.

Skiba, P. F., Fulford, J., Clarke, D. C., Vanhatalo, A. and Jones, A. M. (2015) Intramuscular determinants of the ability to recover work capacity above critical power. *European journal of applied physiology* **115**(4), 703–713.

Sottas, P.-E., Robinson, N. and Saugy, M. (2010) The athlete's biological passport and indirect markers of blood doping. *Doping in sports: Biochemical principles, effects and analysis* pp. 305–326.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2014) The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(3), 485–493.

Statista (2020a) *eServices Report 2020 - Fitness.* `https://www.statista.com/study/36674/fitness-report/`.

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

Statista (2020c) *Wearables dossier.* `https://www.statista.com/study/15607/wearables-statista-dossier/`.

Su, Z., Zhu, G., Chen, X. and Yang, Y. (2016) Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika* **103**(3), 579–593.

Titsias, M. K., Sygnowski, J. and Chen, Y. (2020) Sequential changepoint detection in neural networks with checkpoints. *arXiv preprint arXiv:2010.03053* .

Titsias, M. K. and Yau, C. (2017) The Hamming ball sampler. *Journal of the American Statistical Association* **112**(520), 1598–1611. PMID: 29456460.

Tsay, R. S. and Chen, R. (2018) *Nonlinear time series analysis.* Volume 891. John Wiley & Sons.

Tuyls, K., Omidshafiei, S., Muller, P., Wang, Z., Connor, J., Hennes, D., Graham, I., Spearman, W., Waskett, T., Steel, D. *et al.* (2021) Game plan: What ai can do for football, and what football can do for ai. *Journal of Artificial Intelligence Research* **71**, 41–88.

Vermeulen, E. and Venkata, S. (2018) Big data in sport analytics: applications and risks. In *Industrial Engineering and Operations Management (Presidencia). Proceedings of the International Conference on Industrial Engineering and Operations Management. Conferencia llevada a cabo en IEOM Society International. Pretoria/Johannesburg, South Africa. Recuperado de https://bit. ly/3ojotk9.*

Villena Gonzales, W., Mobashsher, A. T. and Abbosh, A. (2019) The progress of glucose monitoring—a review of invasive to minimally and non-invasive techniques, devices and sensors. *Sensors* **19**(4).

Viroli, C. (2012) On matrix-variate regression analysis. *Journal of Multivariate Analysis* **111**, 296–309.

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*, pp. 186–220. Cham: Springer International Publishing. ISBN 978-3-030-16272-6.

Wang, D., Liu, X. and Chen, R. (2019) Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics* **208**(1), 231–248.

Wang, H. and West, M. (2009) Bayesian analysis of matrix normal graphical models. *Biometrika* **96**(4), 821–834.

Weippert, M., Petelczyc, M., Thürkow, C., Behrens, M. and Bruhn, S. (2021) Individual performance progression of german elite female and male middle-distance runners. *European Journal of Sport Science* **21**(3), 293–299. PMID: 32107979.

West, M. and Harrison, J. (1997) *Bayesian forecasting and dynamic models.* Second edition. Springer Series in Statistics. Springer-Verlag, New York. ISBN 0-387-94725-6.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. ISBN 978-3-319-24277-4.

Xie, L., Zou, S., Xie, Y. and Veeravalli, V. V. (2021) Sequential (quickest) change detection: Classical results and new directions. *IEEE Journal on Selected Areas in Information Theory* **2**(2), 494–514.

Yildirim, S., Singh, S. S. and Doucet, A. (2013) An online expectation-maximization algorithm for changepoint models. *Journal of Computational and Graphical Statistics* **22**(4), 906–926.

Zanella, G. (2020) Informed proposals for local mcmc in discrete spaces. *Journal of the American Statistical Association* **115**(530), 852–865.