Annals of Operations Research (2023) 325:521–538
https://doi.org/10.1007/s10479-021-04391-8

ORIGINAL RESEARCH

Filtering active moments in basketball games using data from players tracking systems

Tullio Facchinetti[1] · Rodolfo Metulini[2] · Paola Zuccolotto[3]

Accepted: 29 October 2021 / Published online: 16 November 2021

Abstract
In recent years, sport analytics evolved in the massive collection of data, especially from Global Positioning System (GPS) sensors installed in sport facilities or worn by the athletes. The largest amount of data are used to track locations and trajectories of players during their performance. Data analysis of positioning information during the actions of a game allows a deep characterization of the performance of single players and the whole team. Basketball is one of the team sports where analytics are becoming a fundamental asset. However, during a game, actions are interleaved with inactive periods (e.g., pauses or breaks). For a proper knowledge extraction on the game features, the analysis of players movements must be restricted to active periods only. This paper proposes an algorithm to automatically identify active periods by using players' tracking data in basketball. The algorithm is based on thresholds that apply to players kinematic parameters. The values of thresholds are identified by setting-up a "ground truth" extracted from the video analysis of the games and by developing a performance evaluation method derived from "Receiver Operating Characteristic" (ROC) curves. When tested on a number of real games, the method shows good performance. This algorithm, along with the identified parameters, could be adopted by practitioners to identify game active periods without the need for video analysis.

Keywords Sensor data · Sports analytics · Big data · Location tracking · Data mining

1 Introduction

Recent years registered the rise of new technologies for collecting, storing and processing an increasing volume of data. These technologies are beneficial in many fields of research, such as medicine, finance, climate change and urban planning. Chen et al. (2014) summarize

✉ Rodolfo Metulini
rmetulini@unisa.it

[1] Department of Industrial, Computer and Biomedical Engineering, University of Pavia., Via Ferrata, 1, 27100 Pavia, PV, Italy

[2] Department of Economics and Statistics (DISES), University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

[3] Department of Economics and Management, University of Brescia, Contrada Santa Chiara, 50, 25122 Brescia, BS, Italy

many big data opportunities. Similarly to the mentioned fields, increasing amount of data are being collected in every sport, which urgently calls for the development of new methods for the automatic processing and analysis of massive and complex datasets in sports as well. Recent years registered the growing interest in sports by researchers from many scientific domains, including machine learning, complex and network systems, statistics and computer vision. The interest is proved by the appearance of dedicated special issues (e.g., Brefeld and Zimmermann 2017; Berrar et al. 2019), workshops (Davis et al. 2018; Lucey et al. 2016), proceedings (Brefeld 2019) and spontaneous contributions in this field, covering a range of topics in several sports such as baseball (Bendtsen 2017), basketball (Yang et al. 2014; Nikolaidis 2015) cricket (Salman et al. 2017), ice hockey (Schulte et al. 2017) soccer (Cravo et al. 2014) and many others. Operations research is a fruitful scientific framework for the analysis of sports data, as it offers an approach to decision making based on the application of problem-solving techniques, which can be particularly useful in the sports context. As a matter of fact, in operations research literature we can find a number of important contributions to sport analytics (Swartz 2020), which can be considered an innovative application field that has considerably grown in the recent years. We find studies aimed at improving performance from several perspectives (just to cite a few recent examples, Goes et al. (2021) use player tracking data to model soccer team performance, Grassetti et al. (2020) propose a method for basketball lineup management), at analyzing sport rules (see Wright (2014) for a survey on a large number of studies covering 21 sports), procedures (e.g., Csató (2020) proposes an analysis of the UEFA Champions League seeding and Durán et al. (2021) focus on efficient referee assignment in Argentinean professional basketball leagues), ranking methods (e.g., Cea et al. 2020). The prediction of game outcomes is also considered (e.g, Song and Shi (2020) use a Bayesian dynamic forecasting procedure to predict NBA future games).

Special attention has been also devoted to the collection and analysis of location and trajectory data that are relevant for applications in fields that include, among others, human mobility (Giannotti and Pedreschi 2008), intelligent transportation (Pang et al. 2013) and animal ecology (Pappalardo and Simini 2018; Li et al. 2011). A spatial trajectory is a sequence of observations that capture the position of an object along a series of points; each observation is also associated with a time-stamp (Zheng and Zhou 2011). A range of widespread applications is based on position tracking by means of Geographical Positioning Systems (GPS). These technologies, despite having some limitations in the localization accuracy (treated, e.g, by Bermingham and Lee (2018)), are increasingly used in a variety of sport disciplines to track the movements of players as well as the ball inside the court. While the GPS technology is suitable for outdoor applications, many sports are played indoor – including basketball. Dedicated solutions exist that allow the tracking of indoor activities, with localization accuracy around 30 cm, as shown by Figueira et al. (2018). The errors do not propagate since every new measurement is always relative to the external localization infrastructure and not to the previous measured position. A comparison of the accuracy provided by different tracking technologies is carried out by Linke et al. (2018).

Tracking data, hereinafter also referred to as “sensor data”, and especially position data, are essential to coaches, experts and analysts, since they contain a full body of information on how the players (and, when available, the ball) move. In the academic literature, several aspects of sport science are based on the use of sensor data – see (Gudmundsson and Horton 2017; Horton 2018) for an overview of applications. For example Kostakis et al. (2017), by monitoring information regarding the position of the players in the field, passing the ball and coordinated moves, propose a method to segment activity streams into a sequence of recurrent modes that reflects different strategies adopted by the team.

Works on National Basketball Association (NBA) data are abundant. For example, Wu and Bornn (2017) provide a useful guide on how to manage with SportVU tracking data (https://www.statsperform.com/team-performance/football/optical-tracking/)—a technology developed by Stats Perform STATS (2018)—to produce visual analysis of offensive actions. Miller and Bornn (2017) use the same data to run a machine learning method that classifies NBA league strategies according to players' movements. Offensive movements and ball circulation are analyzed by D'Amour et al. (2015) to show that the more open shots opportunities can be generated with more frequent and faster movements of the ball. Tracking data has been also used by van Bommel and Bornn (2017) in order to adjust for scorekeeper bias (the inconsistency in the box score due to subjectivity of the scorekeeper) in recording NBA box scores.

Metulini (2017b, 2018) and Manisera et al. (2019) use tracking data from Italian professional basketball games, collected by an authorized private company (MYagonism), to split games into clusters of homogeneous spatial distances among players, looking for those with better team shooting performance. Metulini et al. (2018) apply a vector autoregressive model to show that larger surface area occupied by players is positively related to a large number of scored points by the team.

Localization systems such those of MYagonism track players' movements along the full game. However, during a basketball game, actions are interleaved with inactive periods (e.g., during pauses or breaks). Localization systems cannot distinguish between active and inactive periods. So, an issue that shall be addressed in order to enable the correct analysis of basketball games using tracking data is that of inactive moments. It is worth noting that this issue does not apply to NBA or other American leagues that are adopting the more advanced tracking system of SportVU. SportVU, using the above mentioned technology managed by Stats Perform, delivers to clients game events information attached to tracking data (Franks et al. 2015). So, hereinafter, we assume to deal with data not covered by SportVU and, for this reason, we develop the algorithm based on the International Basketball Federation (FIBA) rules. The data collected for a FIBA game often consists of a total of around 90–100 minutes, despite only 40 minutes are actually related to active play by regulation. Therefore, more than 50% of the real time may be associated to inactive periods. The analyzed data must be restricted to active periods only, in order to obtain meaningful indications (e.g., if one is interested on player's velocity, data is meaningless when he sits on the bench). In order to address this issue, the so-called *play-by-play data* (or *event-log*) are sometimes available, reporting the sequence of relevant games' events. However, play-by-play may not be available or some relevant information may not be recorded or not be usable. For example the three main providers of play-by-play in Europe, Realgm (https://basketball.realgm.com/), Fibalivestats (https://www.fibalivestats.com/u/FIPDP/1140342/pbp.html) and Fibaeurope (https://www.fiba.basketball/en/live-scores), report the game clock associated to information on the events that may be useful to our scope (such as fouls, start/end of time outs, start/end of quarter- and half-time intervals). However, game clock keeps track of the amount of "played" time elapsed (i.e., the 40 active minutes). Conversely, localization systems associate to the measurements their time-stamps. The time-stamp, differently from the game clock, is expressed in terms of "total" amount of time elapsed. So, information retrieved from play-by-play are generally not usable to detect inactive moments due to the inability to correctly match the two sources of data. A possible alternative solution to capture the information about active/inactive periods is to instruct a person to track this information in real-time during the game. Unfortunately, this often does not happen either due to organizational issues and cost impact. For these reasons, analysts require alternative, possibly automatic methods to identify active periods in a game.

This paper aims at providing a solution to the issue of retrieving active and inactive periods by developing a method for active game moments filtering using players' tracking data. The task of recognizing the activity of players by mean of their kinematics is attributable to the discipline of Human Activity Recognition (HAR). HAR aims at identifying the actions of an agent from a series of observations on the agents' actions. In this work, the agents are being represented by the players of the team moving on the court, while the action to recognize concerns whether the game is active or inactive, i.e., the players are actually playing or they are pausing. The literature on HAR is extensive, as shown in the surveys of Gavrila (1999) and Weinland et al. (2011). HAR was applied to the sport domain in several cases. Huang et al. (2012) apply data automation algorithms to sensor data for categorizing golf swing trajectories. A video segmentation algorithm is proposed in Jiang et al. (2004) to identify different elements of an image (playground, players, etc.) that suit to several sports. The model proposed in Jordan et al. (2009) is used to optimize the risk strategies of each decision maker to automate play calling strategies. Kautz et al. (2017) proposed a new Deep Convolutional Neural Network approach to prevent injuries in volleyball players. Learning trajectories in sports is another interesting aspect. Soekarjo et al. (2018) applied supervised classification algorithms on a trajectory dataset to classify the limb, and proposed a technique to identify for each strike in kickboxing. The analysis of Mehrasa et al. (2017) studies the dynamic interaction of multiple people using Deep Learning techniques. Recurrent Neural Network are used in Ramanathan et al. (2016) to detect the player who is responsible of the play-by-play event.

Metulini (2017a) proposed an algorithm to filter active periods in basketball games based on tracking data only. In that work, he tunes parameters in an heuristic way (i.e. he chooses the values which filters the game length to 40 minutes) that completely disregards the issue of selecting the ”correct” 40 minutes. In other words, with the chosen values for the parameters, one may end up (extreme case) with 40 minutes of inactive moments. So, the paper contains some shortcomings, the most important being the lack of verification against a “ground truth”. In this paper we aim to remedy to those shortcomings by introducing a video-based manually extracted ground truth, which allows to tune parameters by considering the accordance of the filtered moments with ”true” active moments, and by proposing a novel performance evaluation method for tuning the parameters that strongly reminds Receiver Operating Characteristic (ROC) curves. All in all, the algorithm consists on the use of kinematic parameters whose values undergoes a “tuning” strategy based on the use of a ground truth and on such a performance evaluation method. The algorithm is suitable in cases where

1. information on the movement of players on the court (with associated time-stamps) has been captured with the use of an appropriate localization technology; but
2. play-by-play is not available or some relevant information are not recorded or not usable; and
3. information regarding active/inactive periods are not explicitly tracked during the game (for example, nobody is in charge to track these events).

This paper is organized as follows. Section 2 describes the sensor dataset that was used and how it was pre-processed to prepare the data. Section 3 outlines the algorithm to filter out inactive moments, while Sect. 4 presents the tuning strategy to identify the values for the parameters of the algorithm. Section 5 shows the results and Sect. 6 concludes the paper.

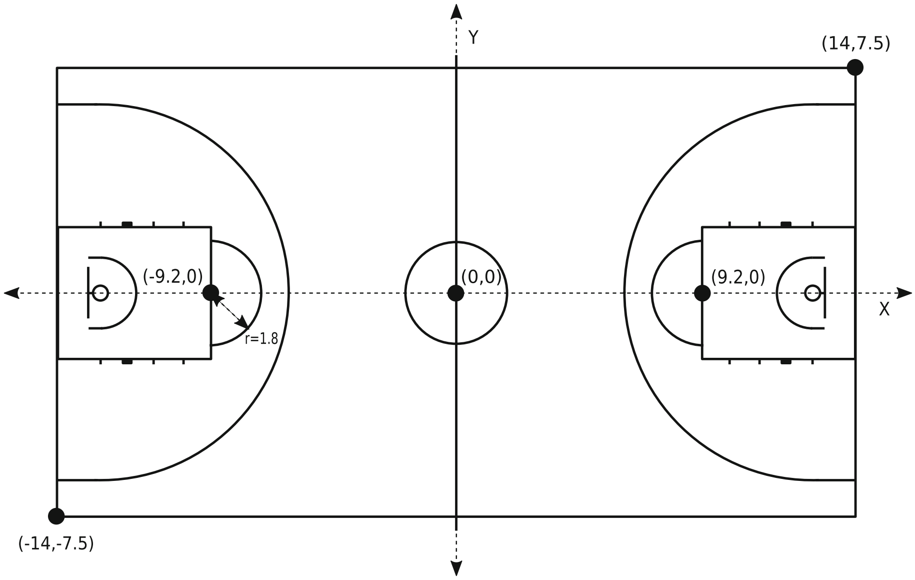


Fig. 1 International Basketball Federation (FIBA) court with relevant measures annotated

2 Context, dataset and feature extraction

The dataset used in this work relates to sensor tracking data of basketball players' movements during official games, provided by MYagonism.

Basketball is a sport played by two teams of five players each on a rectangular court. The objective is to shoot a ball through a hoop of 45 centimeters in diameter and mounted at a height of 3.05 meters to backboards at each end of the court. According to FIBA rules, court size is 28 × 15 meters (Fig. 1) and the game lasts 40 minutes, divided into four periods of 10 minutes each. There is a 2-minutes break after the first quarter and the third quarter of the match. After the first half, there is a 10 to 20 minutes half-time break. In addition to quarter-time and half-time breaks, there are other events that may generate inactive moments in basketball. Those are fouls, fouls with additional free-throws, free-throws, time-outs, players' injuries, technical problems (e.g. an electricity black out on the stadium).

The dataset refers to three games played by Italian professional basketball teams during the Italian Basketball Cup Final Eight 2017. In this paper, the three games will be indicated as case study 1 (CS1), case study 2 (CS2) and case study 3 (CS3). The GPS tracking system in use collects the position, the velocity and the acceleration of the player during the full game length, including those waiting on the bench, along the three axis: the x-axis is the court length, the y-axis is court width, and in the z-axis is the vertical one (e.g., it allows to study the height of players' jumps). On the whole, 10 (for CS1 and CS3) and 11 (for CS2) players of one team[1], rotating on the court, have been analysed. To the purpose of our analysis we do not consider accelerations and the z-axis. The measured positions are expressed in centimeters (cm). The estimated accuracy of each position measurements is around 30 cm.

[1] Aware of possible limitations, we decided to not consider data of both home and away team due to the presence of missing data (i.e. some players had not been tracked for the full game length).

Each measurement is marked by its time instant t. The tracking system is able to capture measurements at a sampling frequency of 50 Hz, corresponding to a measurement every 20 milliseconds (ms).

Considering all the analysed players, the system recorded measurements for a total of $525,901$, $545,143$ and $452,475$ different data samples for CS1, CS2 and CS3, respectively.

Measurements of different players are recorded at different time instants t. So, we let the set of measurements contain any time instant in which at least one player has been tracked, and we let each measurement to be composed of the information of all the players, attributing the last datum available to players not detected in t. As a result, the final set of measurements slightly decrease its dimension, since, in some cases, the original set contains players tracked at the same time instant.

The final set, that we call **X**(t), is made of $505,291$ time instants in CS1, $520,782$ in CS2 and $435,084$ in CS3. Time instants are not evenly spaced, so we denote with $\mathrm{T}(t)$ the actual time corresponding to instant t. The measurements made at instant t contain the following information:

- The vector of the position for the i-th player along the $x-$ and the $y-$ axis, denoted as $\mathbf{P}_i(t) = [p_i^x(t), p_i^y(t)]'$, measured in cm, where superscript x and y refer, respectively, to court length and court width;
- The vector of the velocity for the i-th player along the $x-$ and the $y-$ axis, denoted as $\mathbf{V}_i(t) = [v_i^x(t), v_i^y(t)]'$, measured in kilometres per hour (km/h).

On the basis of such information, additional features that are instrumental to the analysis have been obtained. In detail, we compute a measure of velocity for every single player i on the court at time t, comprehensive of both $x-$ and $y-$ axis velocities (that we will call S). We first assume that player i is on the court at time t if its coordinates, defined by the vector $\mathbf{P}_i(t)$, lie inside the rectangle delimited by the coordinates $(-14.0, -7.5)$ and $(+14.0, +7.5)$ (lower-left and upper-right corners, respectively, expressed in meters). We then denote with ζ_t the set of players on the court at time t, and we compute S as:

$$S_i(t) = \begin{cases} \sqrt{\mathbf{V}_i'(t)\mathbf{V}_i(t)}, & i \in \zeta_t \\ 0, & i \notin \zeta_t. \end{cases}$$

3 The filtering algorithm

This section describes the algorithm aimed at automatically identifying active and inactive periods during a basketball game using players' kinematic parameters. The algorithm aims at labelling as active or inactive the measurements from $\mathbf{X}(t)$ according to three different and sequential criteria (Fig. 2) based on players' position and velocity on the court.

The algorithm has been determined in cooperation with basketball experts, who suggested that, in order to properly take into account the inactive moments of the game, it is necessary to consider, primarily, the number of players on the court and their velocity.

Let x_t be the generic row of $\mathbf{X}(t)$ containing the measurements of all the players at time t:

1. According to the first criterion (A), the algorithm labels as inactive all the rows x_t when the number of players inside the court is different from 5.

$$\text{label } x_t = \begin{cases} \text{inactive}, & \text{if } |\zeta_t| \neq 5 \\ \text{active}, & \text{otherwise} \end{cases}$$

where $|\cdot|$ denotes cardinality.

This criterion allows to detect as inactive pre-match and post-match periods, half-time and quarter-time intervals, time-outs and so on.

2. The second criterion (B) aims to detect as inactive the rows x_t when a player is shooting a free-throw, by considering when his position on the court matches to the free-throw shooting area. Formally, we assume that a player i at time t is on the free-throw shooting area if the vector $\mathbf{P}_i(t)$ lies within a circle C_r with radius $r = 1.80m$ centred on the center of the free-throw area. Depending on the side of attack of the considered team, C_r is centred either on the coordinates [+9.2,0.0] or on the coordinates [-9.2,0.0], expressed in meters.

   The criterion considers how many consecutive seconds the player i lies within the free-throw shooting area. This latter variable will be referred to as $T_i^{ft}(t)$. If at least one player lies inside the free-throw area for a period $T_i^{ft}(t)$ of length equal or larger than parameter $\bar{T}^{ft}$, the algorithm detects as inactive the corresponding row x_t. Formally,

$$\text{label } x_t = \begin{cases} \text{inactive}, & \text{if } \exists i \in \zeta_t : \mathbf{P}_i(t) \in C_r \text{ and } T_i^{ft}(t) \geq \bar{T}^{ft} \\ \text{active}, & \text{otherwise} \end{cases}$$

   where

$$T_i^{ft}(t) = \begin{cases} \sum_{j=1}^{\infty} I_i(t-j) + I_i(t+j), & \text{if } \mathbf{P}_i(t) \in C_r \\ 0, & \text{otherwise} \end{cases}$$

$$I_i(t-j) = \begin{cases} T(t-j+1) - T(t-j), & \text{if } \mathbf{P}_i(t-j) \in C_r \\ & \text{and } \sum_{h=1}^{j} I_i(t-h) = T(t) - T(t-j) \\ 0, & \text{otherwise} \end{cases}$$

$$I_i(t+j) = \begin{cases} T(t+j) - T(t+j-1), & \text{if } \mathbf{P}_i(t+j) \in C_r \\ & \text{and } \sum_{h=1}^{j} I_i(t+h) = T(t+j) - T(t) \\ 0, & \text{otherwise.} \end{cases}$$

3. The third criterion (C) aims to detect as inactive the rows x_t when all the five players on the court are mostly "static", i.e., their velocity is below a given threshold $\bar{S}^{min}$, for a period $T_i^{spd}(t)$ of length equal or larger than $\bar{T}^{spd}$. Formally:

$$\text{label } x_t = \begin{cases} \text{inactive}, & \text{if } \forall i \in \zeta_t : S_i(t) \leq \bar{S}^{min} \text{ and } T_i^{spd}(t) \geq \bar{T}^{spd} \\ \text{active}, & \text{otherwise} \end{cases}$$

   where

$$T_i^{spd}(t) = \begin{cases} \sum_{j=1}^{\infty} I_i(t-j) + I_i(t+j), & \text{if } S_i(t) \leq \bar{S}^{min} \\ 0, & \text{otherwise} \end{cases}$$

$$I_i(t-j) = \begin{cases} T(t-j+1) - T(t-j), & \text{if } S_i(t-j) \leq \bar{S}^{min} \\ & \text{and } \sum_{h=1}^{j} I_i(t-h) = T(t) - T(t-j) \\ 0, & \text{otherwise} \end{cases}$$

$$I_i(t+j) = \begin{cases} T(t+j) - T(t+j-1), & \text{if } S_i(t+j) \leq \bar{S}^{min} \\ & \text{and } \sum_{h=1}^{j} I_i(t+h) = T(t+j) - T(t) \\ 0, & \text{otherwise.} \end{cases}$$

The outcome of the algorithm is the reduced set of measurements denoted as $\mathbf{X}_r(t)$, where rows labelled as inactive are removed.

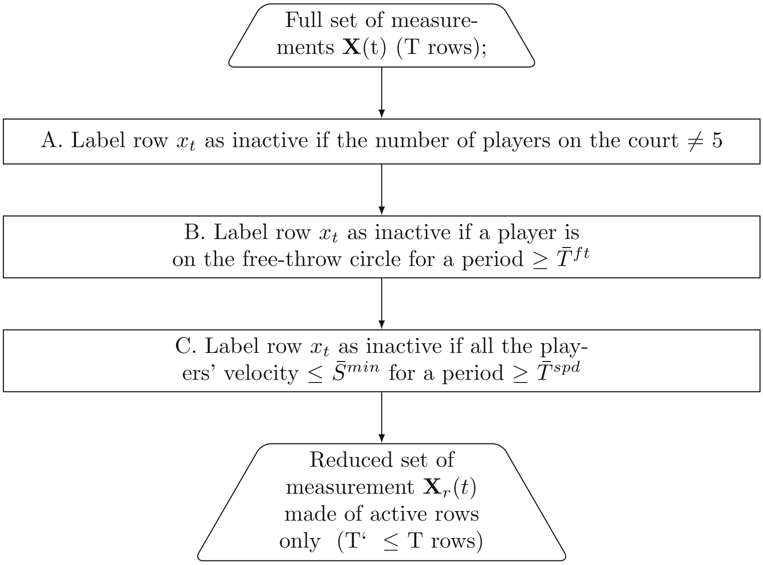

Full set of measurements $\mathbf{X}$(t) (T rows);
A. Label row x_t as inactive if the number of players on the court $\neq 5$
B. Label row x_t as inactive if a player is on the free-throw circle for a period $\geq \bar{T}^{ft}$
C. Label row x_t as inactive if all the players' velocity $\leq \bar{S}^{min}$ for a period $\geq \bar{T}^{spd}$
Reduced set of measurement $\mathbf{X}_r(t)$ made of active rows only (T' $\leq$ T rows)

Fig. 2 Flow chart representing the criteria used in the algorithm

4 Tuning parameters with Pseudo-ROC curves

4.1 Heuristic approach for tuning parameters

Steps B and C of the algorithm involve the determination of the parameters $\bar{T}^{ft}$, $\bar{S}^{min}$ and $\bar{T}^{spd}$. Different values for these parameters lead to a different reduced set $\mathbf{X_r}(t)$ and to a different filtered game length.

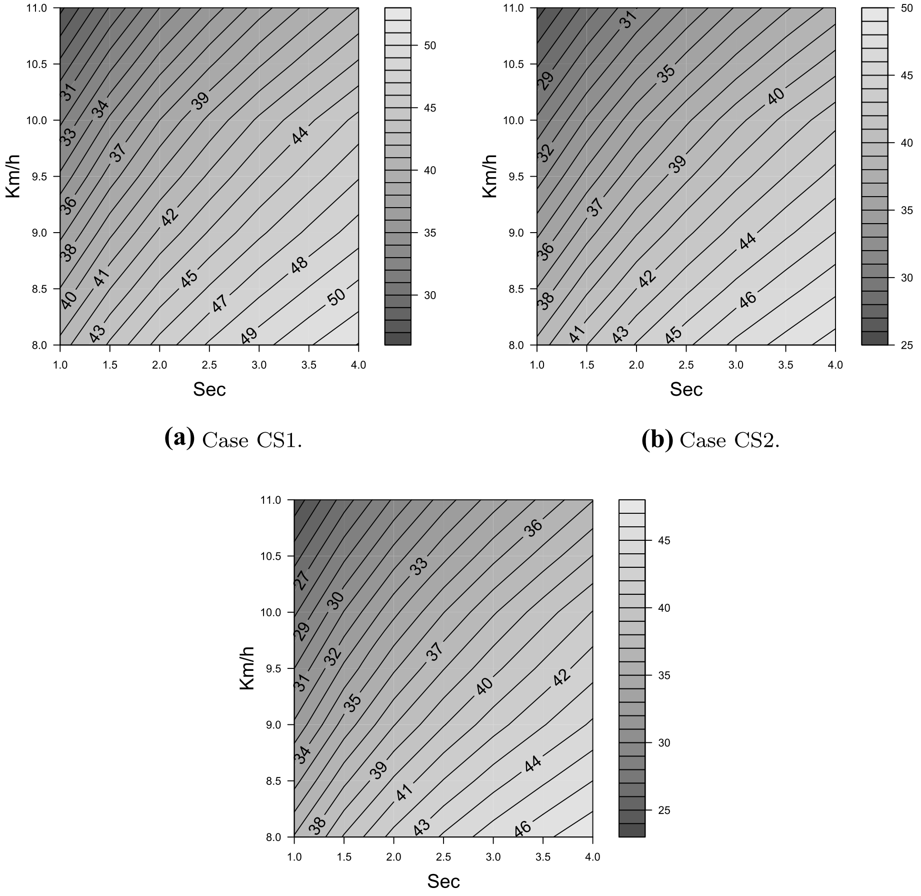
In this subsection we present the heuristic approach used by Metulini (2017a) for tuning the parameters. The determination of $\bar{S}^{min}$ and $\bar{T}^{spd}$ in criterion C is hard to obtain by just watching the video of the games, and it requires the adoption of a tuning strategy. However, to choose a value for $\bar{T}^{ft}$ is quite easy using a video-based analysis. So, we approximate the average duration of a free-throw interruption by watching the video of the games and by annotating basic descriptive statistics. In CS1, CS2 and CS3, the average duration (min/max) of a free-throw interruption is, respectively, 23.41, 21.88 and 26.93 (10/41, 3/54, 6/40) seconds[2]. According to these evidences we are confident that, when a player lies inside the free-throw line for at least 10 seconds, this is referring to an inactive period. Accordingly, $\bar{T}^{ft}$ is defined to be equal to 10 seconds.

About the parameters $\bar{S}^{min}$ and $\bar{T}^{spd}$ one could search for a value with the objective of obtaining a reduced set as close as possible to 40 minutes.

Different combinations of the two parameters are tested with respect to the filtered game length by using contour plots, by setting $\bar{T}^{ft}$=10.

Figure 3 displays contour plots for the three case studies. The charts report different levels of filtered game length (in minutes) as a function of different combinations of $\bar{S}^{min}$ and $\bar{T}^{spd}$. The contours are evaluated in the range [8, 11] km/h (`Km/h`) for $\bar{S}^{min}$ and in the range [1, 4]

[2] The high variability among single interruptions is increased by the fact that players can attempt either one or two free-throws, depending on the situation. The min values of 3 and 6 seconds are outliers, since, sometimes, it was not possible to correctly track the time due to a replay during the television broadcast.



(a) Case CS1.

(b) Case CS2.

(c) Case CS3.

Fig. 3 Contour plots showing the filtered game length (in minutes) obtained by the algorithm, subject to different parameters' $\bar{S}^{min}$ [km/h, y-axis] and $\bar{T}^{spd}$ [s, x-axis] combinations

seconds (`Sec`) for $\bar{T}^{spd}$. For all the three CS, the filtered game length increases as $\bar{T}^{spd}$ parameter increases, while it decreases as $\bar{S}^{min}$ parameter increases. Consistently over the three CS, the objective is satisfied when the parameters $\bar{S}^{min}$ and $\bar{T}^{spd}$ lie in the interval [9, 10] and [2, 3] seconds, respectively.

By setting the values of the parameters to the following valid combination: $\bar{T}^{ft}$=10, $\bar{S}^{min}$=10 and $\bar{T}^{spd}$=2.5, filtered game length of CS1 is 40 minutes and 15 seconds. Moreover, criterion A removes 126,265 rows (24.99% of the total rows of the full set), criterion B removes 1,832 rows (0.36%) and criterion C removes 170,779 rows (33.80%).

However, the tuning approach proposed here is heuristic, its main drawback being the lack of verification against a ground truth.

In order to address this shortcoming, we manually extract the ground truth by means of a video-based annotation of the games (see Sect. 4.2) and we propose a performance evaluation method – referred as "Pseudo-ROC" – to tune the parameters according to the ground truth (Sect. 4.3).

Table 1 Names of the event and description

Event	Description
Start free-throw	A player enters the circle to shoot a free-throw
Stop free-throw	A player exits the circle to shoot a free-throw
Start time-out	A time-out starts
Stop time-out	A time-out ends
Start half-time interval	An half-time interval starts
Stop half-time interval	An half-time interval ends
Start quarter-time interval	A quarter-time interval starts
Stop quarter-time interval	A quarter-time interval ends
Stop	The game stops for a generic reason
Start	The game starts/restarts after a generic stop

In other words, we seek the combination of $\bar{S}^{min}$ and $\bar{T}^{spd}$ allowing the best accordance with the ground truth.

4.2 The video-based annotation

The proposed performance evaluation strategy is based on the manual annotation of the available video footage of the considered games. This annotation will allow checking whether the active moments of games chosen by the algorithm actually correspond to the "real" active moments, and, viceversa, whether the inactive moments of games chosen by the algorithm actually correspond to the "real" inactive moments.

For CS1, CS2 and CS3, we detect the exact times in which the game is active or inactive. Although not necessary for our tuning strategy, we also detect the moments of free-throw shooting, time-outs, quarter-time and half-time intervals. The complete list of detected events with a description is reported in Table 1.

We produce a video-based annotation report for each case study. Each one looks as summarized in Table 2. The report displays when the action starts to be active (*action* = play) or starts to be inactive (*action* = stop) with reference to a given moment (*sec*). *active* is a dummy which assumes value 1 if the game starts to be active in that moment. *timeout*, *ft*, *quarter* and *half* are dummies assuming value 1 if the reason of the inactivity is, respectively, a time-out, the shooting of a free-throw, a quarter-time interval or an half-time interval. In the example in Table 2, the game starts at second 1 ($active$=1 and sec=1 in the first row). From second 1 to second 4 the game is active. At second 5 the game stops ($active$=0 at the second row of the table) due to a generic reason. At second 13 the game restarts (third row) and at second 47 the game stops because of a free-throw (ft = 1 in the fourth row).

4.3 The Pseudo-ROC method

To develop the Pseudo-ROC method we borrow the approach of the Receiver Operating Characteristic (ROC) curves, traditionally used to evaluate the goodness of a statistical prediction model for a dichotomous outcome (positive/negative or true/false). ROC is based on computing sensitivity and specificity, which quantify the performance of a set of binary classifications with respect to the ground truth. In particular, sensitivity measures the propor-

Table 2 A sample of the produced reports

Action	Sec	Active	Timeout	ft	Quarter	Half
Play	1	1	0	0	0	0
Stop	5	0	0	0	0	0
Play	13	1	0	0	0	0
Stop	47	0	0	1	0	0
–	–	–	–	–	–	–

tion of *true* positives, while specificity measures the proportion of *true* negatives. Generally, prediction methods return the probability of being positive for each observation. Then, the observation is predicted as positive if this probability exceeds a given threshold and negative otherwise. Accordingly, sensitivity and specificity are computed. ROC curve is then created by plotting the sensitivity against 1-specificity, at various threshold settings. The Area Under the Curve (AUC) is adopted to obtain an overall measure of goodness-of-fit based on the ROC curve (Zhou et al. 2009; Pepe 2003; Krzanowski and Hand 2009).

In our set-up, let $\tilde{\mathbf{X}}(\tilde{t})$ be the set of measurements obtained from $\mathbf{X}$(t) by aggregating index t at a frequency of 1 second $(\tilde{t})$[3].

We let $Y(\tilde{t})$ be the variable assuming value 1 if, according to the video report, the game is active at second $\tilde{t}$; 0 otherwise. Moreover, for a given combination of $\bar{S}^{min}$ and $\bar{T}^{spd}$, let $\hat{Y}_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})$ be the variable assuming value 1 in $\tilde{t}$ if the majority ($> 50\%$) of the observations x_t corresponding to that $\tilde{t}$ was labelled as active by the filtering algorithm; 0 otherwise.

Accordingly, we define the following variables:

- $TP_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})$ (true positives), which assumes value 1 if that $\tilde{t}$ is classified as active by both the video-based annotation and by the algorithm (i.e. $Y(\tilde{t}) = 1$ and $\hat{Y}_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t}) = 1$), 0 otherwise;
- $TN_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})$ (true negatives), which assumes value 1 if that $\tilde{t}$ is classified as inactive by both the video-based annotation and by the algorithm ($Y(\tilde{t}) = 0$ and $\hat{Y}_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t}) = 0$), 0 otherwise;
- $FP_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})$ (false positives), which assumes value 1 if that $\tilde{t}$ is classified as inactive by the video-based annotation and as active by the algorithm ($Y(\tilde{t}) = 0$ and $\hat{Y}_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t}) = 1$), 0 otherwise;
- $FN_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})$ (false negatives), which assumes value 1 if that $\tilde{t}$ is classified as active by the video-based annotation and as inactive by the algorithm ($Y(\tilde{t}) = 1$ and $\hat{Y}_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t}) = 0$), 0 otherwise.

For that given combination of $\bar{S}^{min}$ and $\bar{T}^{spd}$, sensitivity and specificity are computed using equations (1) and (2).

$$W(\bar{S}^{min}, \bar{T}^{spd}) = \frac{\sum_{\tilde{t}} TP_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})}{\sum_{\tilde{t}} TP_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t}) + \sum_{\tilde{t}} FN_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})} \tag{1}$$

$$Z(\bar{S}^{min}, \bar{T}^{spd}) = \frac{\sum_{\tilde{t}} TN_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})}{\sum_{\tilde{t}} TN_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t}) + \sum_{\tilde{t}} FP_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})} \tag{2}$$

In the illustrated situation, there are no threshold values to set for the underlying probabilities. In our case, the playing at every time instants is directly classified by our algorithm

[3] The aggregation of index t at a frequency of 1 second is necessary, since we match tracking data (expressed in ms) with video-based data (expressed in seconds).

as positive or negative. However, the binary classification of the variable $\hat{Y}_{\bar{S}^{min},\bar{T}^{spd}}(\tilde{t})$ (and consequently $W(\bar{S}^{min}, \bar{T}^{spd})$ and $Z(\bar{S}^{min}, \bar{T}^{spd})$) depends on parameters $\bar{S}^{min}$ and $\bar{T}^{spd}$. We thus propose to measure the performance of our algorithm by evaluating the AUC in terms of $W(\bar{S}^{min}, \bar{T}^{spd})$ and $Z(\bar{S}^{min}, \bar{T}^{spd})$ while changing the values of $\bar{S}^{min}$ and $\bar{T}^{spd}$ as thresholds. For this reason, we refer to the proposed method as "Pseudo-ROC" (PROC).

The method can be formally defined by the following 3 sequential steps:

1. For a given $\bar{S}^{min}$, let $PROC_{\bar{T}^{spd}}|\bar{S}^{min}$ be the PROC curve and $PAUC_{\bar{T}^{spd}}|\bar{S}^{min}$ the corresponding Area Under the Curve computed for different values of $\bar{T}^{spd}$ in the range [0, 20]; here $\bar{T}^{spd}$ is used as a threshold. $PAUC_{\bar{T}^{spd}}|\bar{S}^{min}$ is computed for the $\bar{S}^{min}$ in a sequence of values in [0,20].
2. We let ς be the value of $\bar{S}^{min}$ such that
$$\varsigma = \underset{\bar{S}^{min}}{\operatorname{argmax}}(PAUC_{\bar{T}^{spd}}|\bar{S}^{min})$$
3. Adopting the Youden index criteria (Youden 1950; Fluss et al. 2005; Liu 2012), for the chosen ς, we let τ be the value of $\bar{T}^{spd}$ such that
$$\tau = \underset{\bar{T}^{spd}}{\operatorname{argmax}}\ \Phi(\varsigma, \bar{T}^{spd})$$
$$\text{where } \Phi(\cdot, \cdot) = \ W(\cdot, \cdot) - 1 + Z(\cdot, \cdot).$$

The output of these steps are the values $[\varsigma, \tau]$ for the parameters $\bar{S}^{min}$ and $\bar{T}^{spd}$. In this 3-steps method, we first search for a value for the parameter $\bar{S}^{min}$ because we want to first detect a setting in which all the five players' velocity is less than a threshold. For a logical ordering, we then find the parameter $\bar{T}^{spd}$ by searching for how long the setting lasts.

5 Results and discussion

The PROC method has been tested over the three described case studies. Here we outline and we discuss the results.

Step 1

The pattern of $PAUC_{\bar{T}^{spd}}|\bar{S}^{min}$ as a function of $\bar{S}^{min}$ in the sequence of values in [0,20] at regular intervals of 0.25, for the three case studies is shown in Fig. 4.

Examples of PROC curves for some selected values of $\bar{S}^{min}$ are shown in Fig. 5a, b and c.

Step 2

The parameter ς results 9.25 km/h in CS1 and to 8.5 km/h in both CS2 and CS3, corresponding to $PAUC_{\bar{T}^{spd}}|\bar{S}^{min}$ equal to 0.8329, 0.7995 and 0.7671, respectively.

Step 3

In all the three case studies, the index $\Phi(\varsigma, \bar{T}^{spd})$ is larger for small values of $\bar{T}^{spd}$ (Fig. 6). The largest Youden index is found for τ = 2, 1 and 1.5 in CS1, CS2 and CS3 respectively.

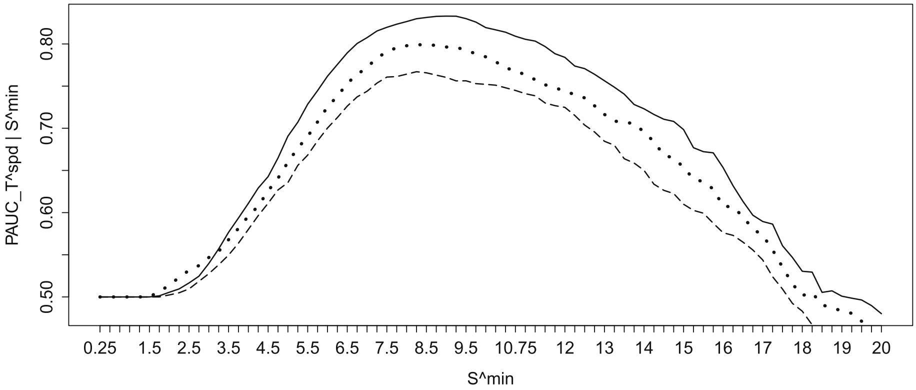
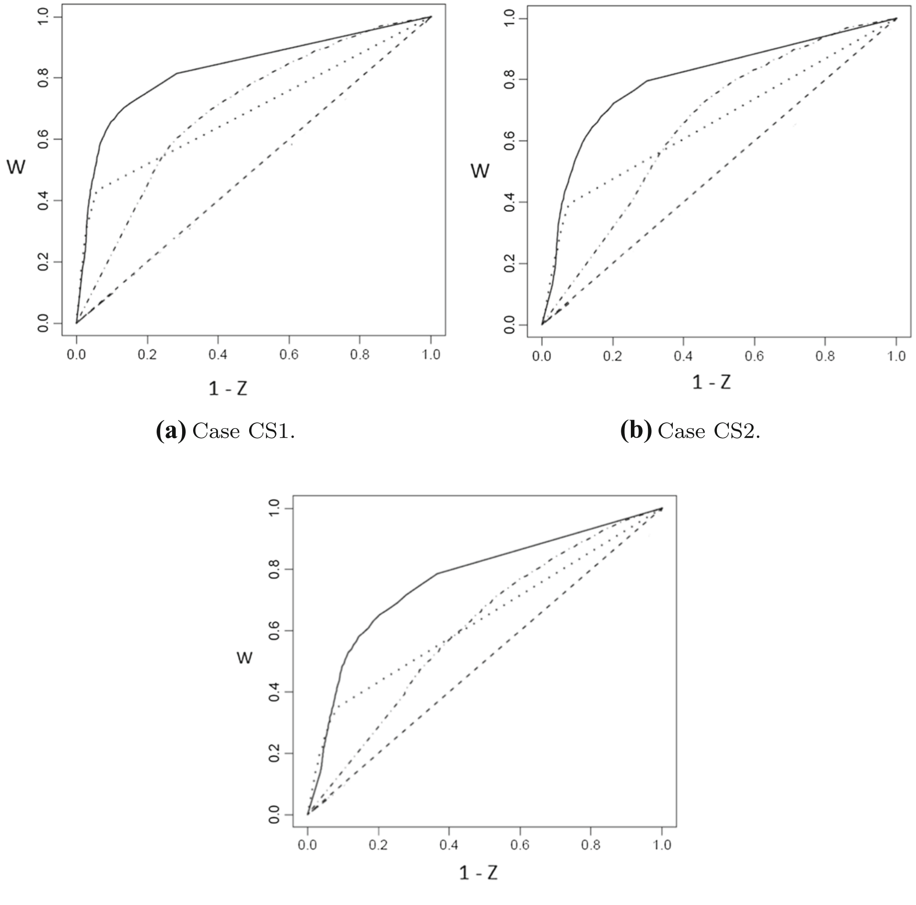


Fig. 4 Pattern of $PAUC_{\bar{T}spd}|\bar{S}^{min}$ as a function of $\bar{S}^{min}$. CS1 (solid line), CS2 (dotted line) and CS3 (longdash line)



(a) Case CS1.

(b) Case CS2.

(c) Case CS3.

Fig. 5 $PROC_{\bar{T}spd}|\bar{S}^{min}$ with $\bar{S}^{min} \in \{0.25, 5, 9.25, 15\}$, represeted by dashed line, dotted line, solid line and dotdash line, respectively

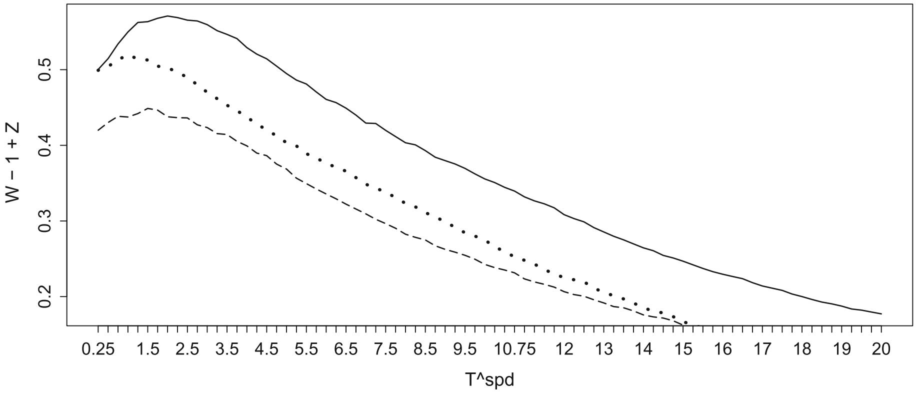


Fig. 6 Pattern of $\Phi(\varsigma, \bar{T}^{spd})$ as a function of $\bar{T}^{spd}$. CS1 (ς= 9.25, solid line), CS2 (ς= 8.5, dotted line) and CS3 (ς= 8.5, longdash line)

Consistently over the three case studies, our tuning strategy based on PROC curves leads to the choice of a set-up in which the rows classified as inactive are those in which all the five players' velocity is lower than 9 km/h for at least a few of seconds.

To prove to robustness of our results we also compute the $hit - rate$ for the three case studies. Hit-rate is a metric traditionally adopted for the classification of customers in credit scoring (e.g., Bensic et al. 2005). This measure ranges from 0 (worst classification) to 1 (best classification). We use hit-rate to measure the relative frequency of time instants correctly classified over the sample total, using the selected values for ς and τ:

$$hit\_rate_{\varsigma,\tau} = \frac{\sum_{\tilde{t}} TP_{\varsigma,\tau}(\tilde{t}) + \sum_{\tilde{t}} TN_{\varsigma,\tau}(\tilde{t})}{\sum_{\tilde{t}} TP_{\varsigma,\tau}(\tilde{t}) + \sum_{\tilde{t}} TN_{\varsigma,\tau}(\tilde{t}) + \sum_{\tilde{t}} FP_{\varsigma,\tau}(\tilde{t}) + \sum_{\tilde{t}} FN_{\varsigma,\tau}(\tilde{t})} \tag{3}$$

Hit-rate results, respectively for CS1, CS2 and CS3, 0.7865, 0.7568 and 0.7320.

Moreover, the filtered game length for CS1, CS2 and CS3 results, respectively, 41:06, 38:45 and 38:04 minutes.

Overall, the tuning strategy for the kinematic parameters of the algorithm proposed in this paper returns consistent results over the different games. The identified values make sense because when all the players on the court run slower than 9 km/h for two consecutive seconds it is reasonable to think that the game is paused.

Basketball analysts may consider to use this automatic algorithm along with the identified parameters when using tracking data of players in their analysis instead of themselves retrieving the play/pause information from the video. Since these analysts will not be able to validate the parameters, we suggest to start with the set of values [9,2] and then—whenever the filtered game length does not correspond to 40 minutes—to gradually move towards different values of the parameters with the aim of obtaining 40 minutes of game length.

6 Concluding remarks

The study of players' trajectories in basketball analytics is gaining relevance as in the era of big data it is possible to store and process huge amounts of data of any kind. Tracked data of basketball players feeds analysts with the availability of additional information to produce advanced statistics at support to game decisions. However, these data are relevant just during active moments of the game, but in some cases it is not possible to correctly analyse tracking data as they are, as tracking systems capture the full game length while the game is interleaved by pauses and breaks. We particularly refer to those cases when play-by-play (or event data) are not available, it is not possible to correctly match them with tracking data or they report some missing information, and nobody is in charge of manually annotating the moments of breaks and pauses. In these cases, analysts need for an automatic algorithm to filter out inactive moments of the game.

To do so, in this paper we proposed an algorithm which is based on players' kinematic parameters. The robustness of the strategy to tune the parameters benefits from the application of a "ground truth" coming from a video-based annotation of a sample of games and from the development of a new performance evaluation method which is similar to Receiving Operation Characteristic curves.

The identified values for the parameters has been found to be consistent along different case studies. Moreover, the robustness of the tuning strategy is confirmed by the analysis of the resulting filtered game length, which is always very close to 40 minutes. However, considering the limitation coming from the number of games available, in order to consider these values as "robust", it is desirable to validate the procedure using data coming from additional games, also checking whether they are sensitive to different game conditions (for example whether they change from regular season to play-offs, where the intensity of game play increases).

A criticism that could be raised to this work is the lack of a comparison to other existing methods. Our PROC approach is developed *ad-hoc* to filter inactive moments in basketball. As a matter of fact, to the best of our knowledge, "gold standard" methods for this aim do not exist at all. In fact, many works dealing with event detection in team sports rely on pre-processed data: SportVU technology, for example, supplies to clients the full event mappings associated to tracking data, so this kind of data does not need a filtering algorithm for inactive moments. While, potentially, this algorithm may be compared with the ones used by Stats Perform to pre-process SportVU data, actually, this turns out to be complicated, since complete details about how they pre-process data are not publicly available, for obvious reasons. In this respect, where tracking data with SportVU technology is available, one can face other research questions, such as detecting offensive and defensive assignments (Keshri et al. 2019) or characterize defensive skills (Franks et al. 2015). In other sports (e.g., in soccer, when the issue of inactive periods is less relevant compared to basketball) it is also possible to focus the attention on recognizing events without a filtering algorithm for inactive periods. For example, Khaustov and Mozgovoy (2020) propose a rule-based algorithm with an approach very similar to that of our paper, with the aim of identifying several basic types of events in soccer. Here the authors evaluate the quality of event detection by using recall, precision, and F1 score values, which can be considered an alternative to the PROC method presented in this paper. Again with reference to using algorithms to tag player tracking data with additional information in soccer, there are solutions that use feature reduction strategies to find the most likely formation patterns of a team associated with match events (Xinyu et al. 2013), machine learning approaches to determine how long the ball spends in the sphere of influence of a player (Link and Hoernig 2017), a temporal logic approach to detect "atomic" and "complex" events (Morra et al. 2020). However, all the proposals made in the field of soccer cannot be directly compared to ours, due to the obvious context differences.

The algorithm, along with the identified values for the players kinematic parameters may help basketball experts who want to analyse tracking data without watching to the video of the game. The generalization of the algorithm to other team sports is not straightforward, as it employs basketball-specific rules. It is actually possible to adopt the PROC approach to other sports, taking care of appropriate changes in the rules adopted by the algorithm. Concluding, it is worth noting that, despite the employment of our proposed approach requires high-frequency tracking data, nonentheless it works with just x- and y-axis position and velocity of the players of one team (so, it requires a minimum quantity of information).

Future works aims to develop a similar automatic algorithm to split the game into offensive and defensive actions by labelling measurements as "offensive" or "defensive" using players' tracking data.

Acknowledgements Research carried out in collaboration with the Big&Open Data Innovation Laboratory (BODaI-Lab), University of Brescia (project nr. 03-2016, title *Big Data Analytics in Sports*, http://bdsports.unibs.it), granted by Fondazione Cariplo and Regione Lombardia.

References

Bendtsen, M. (2017). Regimes in baseball players' career data. *Data Mining and Knowledge Discovery 31*(6), 1580–1621. https://doi.org/10.1007/s10618-017-0510-5

Bensic, M., Sarlija, N., & Zekic-Susac, M. (2005). Modelling smallbusiness credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management: International Journal, 13*(3), 133–150. https://doi.org/10.1002/isaf.261

Bermingham, L., & Lee, I. (2018). A probabilistic stop and move classifier for noisy gps trajectories. *Data Mining and Knowledge Discovery, 32*(6), 1634–1662. https://doi.org/10.1007/s10618-018-0568-8

Berrar, D., Lopes, P., Davis, J., & Dubitzky, W. (2019). Guest editorial: Special issue on machine learning for soccer. *Machine Learning, 108*(1), 1–7. https://doi.org/10.1007/s10994-018-5763-8

Brefeld, U., & Zimmermann, A. (2017). Guest editorial: Special issue on sports analytics. *Data Mining and Knowledge Discovery, 31*(6), 1577–1579. https://doi.org/10.1007/s10618-017-0530-1

Brefeld, U. (2019). Machine Learning and Data Mining for Sports Analytics. Springer. https://doi.org/10.1007/978-3-030-17274-9

Cea, S., Durán, G., Guajardo, M., Sauré, D., Siebert, J., & Zamorano, G. (2020). An analytics approach to the FIFA ranking procedure and the World Cup final draw. *Annals of Operations Research, 286*(1), 119–146.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Cravo, J., Almeida, F., Abreu, P. H., Reis, L. P., Lau, N., & Mota, L. (2014). Strategy planner: Graphical definition of soccer set-plays. *Data and Knowledge Engineering, 94*, 110–131. https://doi.org/10.1016/j.datak.2014.10.001

Csató, L. (2020). The UEFA Champions League seeding is not strategy-proof since the 2015/16 season. *Annals of Operations Research, 292*, 161–169.

D'Amour, A., Cervone, D., Bornn, L., & Goldsberry, K. (2015). Move or die: how ball movement creates open shots in the nba, MIT Sloan Sports Analytics Conference.

Davis, J., van Haaren, J., Kaytoue, M., & Zimmermann, A. (2018). Machine learning and data mining for sports analytics. https://dtai.cs.kuleuven.be/events/mlsa18/index.php.

Durán, G., Guajardo, M., & Gutiérrez, F. (2021). Efficient referee assignment in Argentinean professional basketball leagues using operations research methods. *Annals of Operations Research* 1–19.

Figueira, B., Gonçalves, B., Folgado, H., Masiulis, N., Calleja-González, J., & Sampaio, J. (2018). Accuracy of a basketball indoor tracking system based on standard bluetooth low energy channels (nbn23®). *Sensors, 18*(6), 1940. https://doi.org/10.3390/s18061940

Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the youden index and its associated cutoff point. *Biometrical Journal: Journal of Mathematical Methods in Biosciences, 47*(4), 458–472. https://doi.org/10.1002/bimj.200410135

Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *Annals of Applied Statistics, 9*(1), 94–121.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer vision and image understanding, 73*(1), 82–98. https://doi.org/10.1006/cviu.1998.0716

Giannotti, F., & Pedreschi, D. (2008). Mobility, data mining and privacy: Geographic knowledge discovery, Springer Science & Business Media. ISBN: 978-3-540-75176-2.

Goes, F. R., Kempe, M., van Norel, J., & Lemmink, K. A. P. M. (2021). Modelling team performance in soccer using tactical features derived from position tracking data. *IMA Journal of Management Mathematics*

Grassetti, L., Bellio, R., Di Gaspero, L., Fonseca, G., & Vidoni, P. (2020). An extended regularized adjusted plus-minus analysis for lineup management in basketball using play-by-play data, IMA Journal of Management Mathematics.

Gudmundsson, J., & Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys (CSUR), 50*(2), 22. https://doi.org/10.1145/3054132

Horton, M. (2018). Algorithms for the Analysis of Spatio-Temporal Data from Team Sports, PhD thesis, University of Sydney. URI: http://hdl.handle.net/2123/17755.

Huang, Y.-C., Chen, T.-L., Chiu, B.-C., Yi, C.-W., Lin, C.-W., Yeh, Y.-J., & Kuo, L.-C. (2012). Calculate golf swing trajectories from imu sensing data. In: Parallel Processing Workshops (ICPPW), 2012 41st International Conference on, IEEE (pp. 505-513). ISBN: 978-1-4673-2509-7.

Jiang, S., Ye, Q., Gao, W., & Huang, T. (2004). A new method to segment playfield and its applications in match analysis in sports video. In: Proceedings of the 12th annual ACM international conference on Multimedia, ACM (pp. 292-295). ISBN: 978-1-58113-893-1.

Jordan, J. D., Melouk, S. H., & Perry, M. B. (2009). Optimizing football game play calling, *Journal of Quantitative Analysis in Sports*, *5*(2). https://doi.org/10.2202/1559-0410.1176.

Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity recognition in beach volleyball using a deep convolutional neural network. *Data Mining and Knowledge Discovery, 31*(6), 1678–1705. https://doi.org/10.1007/s10618-017-0495-0

Keshri, S., Oh, M. H., Zhang, S., & Iyengar, G. (2019). Automatic event detection in basketball using HMM with energy based defensive assignment. *Journal of Quantitative Analysis in Sports, 15*(2), 141–153.

Khaustov, V., & Mozgovoy, M. (2020). Recognizing events in spatiotemporal soccer data. *Applied Sciences, 10*(22), 8046.

Kostakis, O., Tatti, N., & Gionis, A. (2017). Discovering recurring activity in temporal networks. *Data Mining and Knowledge Discovery, 31*(6), 1840–1871. https://doi.org/10.1007/s10618-017-0515-0

Krzanowski, W. J., & Hand, D. J. (2009). ROC curves for continuous data, Chapman and Hall/CRC. ISBN: 978-1-4398-0021-8.

Li, Z., Han, J., Ji, M., Tang, L.-A., Yu, Y., Ding, B., Lee, J.-G., & Kays, R. (2011). Movemine: Mining moving object data for discovery of animal movement patterns. *ACM Transactions on Intelligent Systems and Technology (TIST), 2*(4), 37. https://doi.org/10.1145/1989734.1989741

Link, D., & Hoernig, M. (2017). Individual ball possession in soccer. *PloS one, 12*(7), e0179953.

Linke, D., Link, D., Lames, M., & Ardigò, L. P. (2018). Validation of electronic performance and tracking systems epts under field conditions. *PLoS One 13*(7). https://doi.org/10.1371/journal.pone.0199519.

Liu, X. (2012). Classification accuracy and cut point selection. *Statistics in medicine, 31*(23), 2676–2686. https://doi.org/10.1002/sim.4509

Lucey, P., Morgan, S., Wiens, J., & Yue, Y. (2016). Kdd workshop on large-scale sports analytics. http://large-scale-sports-analytics.org/.

Manisera, M., Metulini, R., & Zuccolotto, P. (2019). Basketball analytics using spatial tracking data, *New Statistical Developments in Data Science* pp. 305-318. https://doi.org/10.1007/978-3-030-21158-5-23.

Mehrasa, N., Zhong, Y., Tung, F., Bornn, L., & Mori, G. (2017). Learning person trajectory representations for team activity analysis, arXiv preprint arXiv:1706.00893.

Metulini, R. (2017). Filtering procedures for sensor data in basketball. *Statistica and Applicazioni, 15*(2), 133–150. https://doi.org/10.26350/999999000007

Metulini, R. (2017). Spatio-temporal movements in team sports: A visualization approach using motion charts. *Electronic Journal of Applied Statistical Analysis, 10*(3), 809–831. https://doi.org/10.1285/i20705948v10n3p809

Metulini, R. (2018). Players movements and team shooting performance: a data mining approach for basketball. In: 49th Scientific meeting of the Italian Statistical Society, SIS2018 proceeding (pp. 681-688). ISBN-9788891910233.

Metulini, R., Manisera, M., & Zuccolotto, P. (2018). Modelling the dynamic pattern of surface area in basketball and its effects on team performance. *Journal of Quantitative Analysis in Sports, 14*(3), 117–130. https://doi.org/10.1515/jqas-2018-0041

Miller, A. C., & Bornn, L. (2017). Possession sketches: Mapping NBA strategies, MIT Sloan Sports Analytics Conference 2017.

Morra, L., Manigrasso, F., Canto, G., Gianfrate, C., Guarino, E., & Lamberti, F. (2020). Slicing and dicing soccer: Automatic detection of complex events from spatio-temporal data. In *International Conference on Image Analysis and Recognition* Springer.

Nikolaidis, Y. (2015). Building a basketball game strategy through statistical analysis of data. *Annals of Operations Research, 227*(1), 137–159. https://doi.org/10.1007/s10479-013-1309-4

Pang, L. X., Chawla, S., Liu, W., & Zheng, Y. (2013). On detection of emerging anomalous traffic patterns using gps data. *Data and Knowledge Engineering, 87*, 357–373. https://doi.org/10.1016/j.datak.2013.05.002

Pappalardo, L., & Simini, F. (2018). Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery, 32*(3), 787–829. https://doi.org/10.1007/s10618-017-0548-4

Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction, Medicine. ISBN: 978-0198565826.

Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2016). Detecting events and key actors in multi-person videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3043-3053). https://doi.org/10.1109/CVPR.2016.1.

Salman, M., Qaisar, S., & Qamar, A. M. (2017). Classification and legality analysis of bowling action in the game of cricket. *Data Mining and Knowledge Discovery, 31*(6), 1706–1734. https://doi.org/10.1007/s10618-017-0511-4

Schulte, O., Khademi, M., Gholami, S., Zhao, Z., Javan, M., & Desaulniers, P. (2017). A markov game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery, 31*(6), 1735–1757. https://doi.org/10.1007/s10618-017-0496-z

Soekarjo, K. M., Orth, D., Warmerdam, E., & Van Der Kamp, J. (2018). Automatic classification of strike techniques using limb trajectory data. In: International Workshop on Machine Learning and Data Mining for Sports Analytics, Springer (pp. 131-141). https://doi.org/10.1007/978-3-030-17274-911.

Song, K., & Shi, J. (2020). A gamma process based in-play prediction model for National Basketball Association games. *European Journal of Operational Research, 283*(2), 706–713.

STATS (2018). Sportvu system. Last visited: 2018-08-17.

Swartz, T. B. (2020). Where should i publish my sports paper? *The American Statistician, 74*(2), 103–108.

van Bommel, M., & Bornn, L. (2017). Adjusting for scorekeeper bias in nba box scores. *Data Mining and Knowledge Discovery, 31*(6), 1622–1642. https://doi.org/10.1007/s10618-017-0497-y

Weinland, D., Ronfard, R., & Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Computer vision and image understanding, 115*(2), 224–241. https://doi.org/10.1016/j.cviu.2010.10.002

Wright, M. (2014). OR analysis of sporting rules-A survey. *European Journal of Operational Research, 232*(1), 1–8.

Wu, S., & Bornn, L. (2017). Modeling offensive player movement in professional basketball. *The American Statistician, 72*(1), 72–79. https://doi.org/10.1080/00031305.2017.1395365

Xinyu W., Long S., Patrick L., Stuart M., & Sridha S. (2013). Large-scale analysis of formations in soccer. In *2013 international conference on digital image computing: Techniques and applications (DICTA)*, IEEE.

Yang, C. H., Lin, H. Y., & Chen, C. P. (2014). Measuring the efficiency of NBA teams: Additive efficiency decomposition in two-stage DEA. *Annals of Operations Research, 217*(1), 565–589. https://doi.org/10.1007/s10479-014-1536-3

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32–35. https://doi.org/10.1002/1097-0142

Zheng, Y., & Zhou, X. (2011). Computing with spatial trajectories. Springer. https://doi.org/10.1007/978-1-4614-1629-6

Zhou, X.-H., McClish, D. K., & Obuchowski, N. A. (2009). Statistical methods in diagnostic medicine, Vol. 569. Wiley. https://doi.org/10.1002/9780470906514.