Assessing the Amount of Data per Second to Measure Tactical Variables in Team Sports

Markel Rico-González*  
University of the Basque, SPAIN

Asier Los Arcos  
University of the Basque, SPAIN

Carlos D. Gómez-Carmona  
University of Extremadura, SPAIN

Daniel Rojas-Valverde  
University National of Costa Rica, COSTA RICA

José Pino-Ortega  
University of Murcia, SPAIN

Abstract

Sampling frequency of microsensors that measure the position of the players in team sports is a variable that could affect the accuracy of the measurement. The aim of the present study was to assess the impact of the sampling frequency on the measurement of a collective tactical behaviour variable: the total area (TA). Sixteen young U16 male soccer players participated in the study. They carried out three controlled tasks. Tactical variable was measured by a radio ultra-wide band technology (IMU; WIMU PRO\textsuperscript{Tm}, RealTrack Systems, Almeria, Spain). For TA different sampling frequencies were applied (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz). Trivial differences ($p>0.05$) were found between the TA values across the different amounts of inserted data per second across Task 1 (ES=0.04-0.08), Task 2 (ES=0.01-0.09) and Task 3 (ES=-0.03-0.04). Also High to perfect ICCs (0.91-1) and linear correlations ($r=0.961-1; p<0.01$) were found among the TA values obtained through all sampling frequencies. The sampling frequency (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz) does not affect the measurement of total area during tactical behaviour analysis but does significantly affect the change in centroid position measurement. Thus, it is recommended the use of 1 Hz to measure TA, but further studies should analyse the impact of lower than 1 Hz and greater than 10 Hz sampling frequencies to measure this collective tactical behavior.

To cite this article: González et al. (2023). Assessing the amount of data per second to measure tactical variables in team sports. Journal of Coaching and Sports Science, 2(2), 68-77

INTRODUCTION

Positioning tracking systems technologies were originally intended for military and scientific use, but in recent years they have been used for a wide variety of applications (Malone et al., 2017). For example, the sports area has been a new niche of development for outdoor and indoor tracking systems (Frencken, W & Lemmink, K, 2009; Malone et al., 2017; Passos et al., 2008). These new applications have motivated improvements for the positional, computational and image analyses of these types of instruments. Application of position tracking systems in sport were driven by studies led by Schmidt, O’ Brien and Sysko (1999), who opened up new lines of research regarding intra- and inter-person coordination, making it possible to assess tactical behaviour in sport (Schmidt, O’ Brien, & Sysko, 1999). Although the authors proposed these methods as tactical measures, they were later called micro-level measures because they only quantified two players. The dyad analysis was proposed for basketball (Schmidt, O’Brien, & Sysko, 1999) but were also developed for individual racket sports (Palut & Zanone, 2005). Years later they were used to measure team sports as well (Passos et al., 2008; Yue et al., 2008). While some authors continued with dyad analysis, Schöllhorn (2003) proposed a triad analysis consisting of a) covered area by several or all players, b) common centre of gravity of several or all team members, and c) geometric shape which is formed by several or all team members. Therefore, this analysis method revealed that all changes of spatial parameters over time provided fruitful information about the behaviour of a team as a whole (Schöllhorn, 2003). However, this line of research was not further developed until 2008 (Yue et al., 2008). From this year on, supported by technological development, several published studies...
have appeared on this topic and many variables of spatial positioning tracking in team sports have been analysed in order to measure tactical behaviour (Low et al., 2020; Rico-González, Los Arcos, et al., 2020; Rico-González, Pino-Ortega, et al., 2020).

Several collective tactical variables have been classified into three geometric primitives (i.e. point, distance and polygon) (Rico-González, Los Arcos, et al., 2020; Rico-González, Pino-Ortega, et al., 2020). Among other tactical variables, the centroid (i.e. represented as a point), also named the geometrical centre (Yue et al., 2008) or centre of gravity (Schöllhorn, 2003; Travassos et al., 2012) of the team, and TA (total area) surface area i.e. represented as a polygon have been commonly measured in order to assess tactical behaviour in team sports (Low et al., 2020). The centroid represents, in a single variable, the relative positioning of each team in forward-backward and side-to-side movements (Araújo & Davids, 2016). The change in centroid position (CCP) is the distance in metres between two consecutive measured points of the centroid as the mid-point of the polygon. The TA represents the total field coverage of each team (Frencken & Lemmink, 2009) and is habitually used along with the centroid to assess team behaviour in team sports (Barnabé et al., 2016; Frencken et al., 2011; Frencken & Lemmink, 2009; Palucci Vieira et al., 2018). TA is defined as total square metres of a polygon described by players as its vertex point. It is also used to assess inter-team coordination through measurement of coupling stretch, relative phase (Lames, Ertmer, & Walter, 2010; Silva et al., 2014) and pressure index if compared to the team’s centroids distance (Frencken & Lemmink, 2009). This variable expresses the relationship between the tactical shapes adopted and spaces exploited by both teams to support analysis of how they vary over time (Barnabé et al., 2016). In addition, it has been used as a pressure index (Frencken et al., 2011; Frencken & Lemmink, 2009). Furthermore, the importance of increasing TA for the attacking sub-groups has been suggested to destabilise the opposing team and create shooting opportunities (Duarte et al., 2012).

These variables are based on positional data (Rico-González, Arcos, et al., 2020). Currently, FIFA collectively labels an assortment of competing tracking technologies that that differ in their methods or protocols as Electronic Performance and Tracking Systems (EPTS). Using EPTS, one of the parameters that researchers can modify according to their needs is the sampling rate of data collected per second, called “raw data” and expressed in hertz (Hz) (Winter, 2009). Among other factors, the sample rate capacity, which varies between EPTS technologies (Malone et al., 2017; Rico-González, Los Arcos, et al., 2020), influences the accuracy of the reported position of individual players on the pitch (Duarte et al., 2010; Frencken et al., 2010; Leser et al., 2011) and, consequently, the accuracy of team behaviour variables (Rico-González, Los Arcos, et al., 2020). Deciding before the investigation which sampling frequency is to be used when recording the data (i.e. raw data) is fundamental to avoid violating the Nyquist sampling theorem. The theorem shows that the sampling frequency must be at least twice as high as the highest frequency given by the signal itself (i.e. if the amount of Hz is too low, errors or bias will occur in the recording) (Winter, 2009). However, when the sampling frequency is too high, the signal may be distorted by noise, which increases linearly with frequency (Winter, 2009). In this sense, lowpass digital filtering of noisy signals has been an important procedure because the objective of any filtering technique is to attenuate noise and leave the true signal unaffected and stable (Winter, 2009). Once recorded, the data can go through a process of data reduction using algorithms, producing software-derived data (Malone et al., 2017).

To date, it is not clear what sampling frequency is suitable to measure collective tactical behaviour in team sports (Rico-González, Los Arcos, et al., 2020). The most commonly used sampling frequencies ranged from 0.4 to 100 Hz, from 0.4 to 50 Hz and from 1 to 30 Hz to measure the GC, the distance between two points and the area, respectively (Rico-González, Los Arcos, et al., 2020). Since the magnitude of the tactical behaviour variables differ according to their characteristics (i.e. a single point or occupied space), we hypothesised that different sampling frequencies are needed according to the magnitude of the variable to be measured. In fact, Rico-González et al., (2020) proposed a set of standard items to assess the quality of the methodology, and one of these criteria suggested the use of different sampling rates for each variable. However, to our knowledge, no study has assessed the impact of the sampling frequency on the outcomes of tactical behaviour variables during controlled tasks. Therefore, the aim of the study was to assess the impact of the sampling frequency on the measurement of collective tactical behaviour (i.e. CCP and the TA).
METHOD

Participants
Data was collected from sixteen young male soccer players (under 16 years) (age 15.6 ± 0.8, height 1.70 ± 0.1 cm, weight 65.6 ± 10.2 kg) who belonged to the Torre Pacheco Soccer School (Spain). These players also participated in the cadet category of the Autonomous League of the region of Murcia during the 2018-2019 season. The team’s staff gave their consent for their participation in this study. A written consent was signed by their legal guardians and players gave their assent to participate. The study, which was conducted according to the Declaration of Helsinki (2013), was approved by the Bioethics Commission of the University (Reg. Code 67/2017).

Procedure
Two teams of eight players participated in the exercises on a field of 30x40 meters (Coutinho et al., 2018). They were asked to execute three different controlled tasks: i) players walked for 1 min along the line that described the perimeter of the area arranged (see Figure 1a) ii) players walked along the perimeter line and after the coach’s signal they ran to the centre of a smaller area placed in the middle of the total area and then scattered towards the perimeter line again continuously for one minute (see Figure 1b), and iii) players walked along the perimeter line and after the coach’s signal they ran to the corners of a smaller area placed in the middle of the total area and then scattered towards the perimeter line again continuously for one minute (see Figure 1c).

Figure 1. Representation of a) Task 1; b) Task 2; and c) Task 3. Total Area is represented as yellow and black polygons, players in same colours defining area as vertexes.

Data collection
Positional data were collected using a commercial EPTS (WIMU PRO™, RealTrack Systems, Almeria, Spain). Each device contains a 10 Hz GPS and an 18 Hz UWB (Ultra-wideband), as well as other sensors (three 3-axis gyroscopes, a 3-axis magnetometer, four 3-axis accelerometers, etc.). For data collection TA (m²) and CCP (m) were measured at 18 Hz for raw data by radio ultra-wide band (UWB) sensor. The UWB system is composed of two sub-systems: (1) the reference system and, (2) the devices tracked (carried by the players). The first is composed of six antennae that are transmitters and receivers of the radio-frequency signals. The antennae (mainly the master antenna) computerize the position of the devices that are in the play area, while the devices receive that calculation (Bastida-Castillo, Gómez-Carmona, De la Cruz-Sánchez, et al., 2019). The TDOA algorithm was used to estimate positioning. The UWB occupies a very large frequency band (i.e. at least 0.5 GHz), as opposed to more traditional radio communications that operate on much smaller frequency bands (Alarifi et al., 2016). On the other hand, since UWB is only allowed to transmit at very low power. Its signal emits little noise and can coexist with other services without influencing them (Bastida-Castillo, Gómez-Carmona, De la Cruz-Sánchez, et al., 2019). This UWB system has recently been validated for collective tactical behaviour variables (Bastida-Castillo, Gómez-Carmona, De La Cruz Sánchez, et al., 2019).

Ultra-wide band antennas were placed around the playing field. The auto-start process was carried out followed by their synchronisation prior to the placement of the tracking devices on the participants. The auto-start process followed a protocol that incorporated each device in the internal configuration at the start. For auto-start, three aspects were considered: (i) leaving the device immobile for 30 seconds, (ii) placing it on a flat area and (iii) without magnetic devices around it. All
WIMUs were attached to the players by a special vest placed inside a pocket placed between the scapulae at the T2-T4 level and prior to in-field exercises following previous study protocols (Reche-Soto et al., 2019).

Data processing

To investigate the accuracy of the UWB system for monitoring players’ positions on the court, the data were transformed into raw position data (x and y coordinates) using a software (S PRO, RealTrack Systems, Almeria, Spain). Four different sampling frequencies were considered (1 Hz, 2 Hz, 4 Hz, and 10 Hz). And then, the x and y coordinate data of the UWB system were introduced and compared. Subsequently, we assessed the impact of the sampling frequency on the measurement of the CCP and TA. For statistical analysis purposes, the datasets corresponding to each sampling frequency were balanced in order to perform intraclass correlation coefficient and Bland Altman agreement. The balancing was performed downsampling each dataset, calculating the mean of the data each 2 Hz, 4 Hz and 10 Hz values in order to have the same amount of data in each dataset.

Statistical analysis

The data are presented as means with standard deviation. The Shapiro Wilk test was applied to confirm the normality of the data, verifying the feasibility of using parametric inference. Following previous study principles (Kottner & Streiner, 2011; Zaki et al., 2012), we analyzed the agreement among the different sampling frequencies. We used these tests: 1) intraclass correlation coefficient (ICC) with a mixed two-way model and a 95% CI; 2) 1 sample t-test of the differences using the Bland and Altman (1987) method to assess bias and agreement, 3) r-Pearson to explore linear correlation between the different sampling frequencies; 4) t-test to explore significant differences between variable sampling frequency. Moreover, magnitude of the difference was assessed using Cohen’s d effect size (Cohen, 1988), qualitatively rated as follows: < 0.2 trivial, 0.2-0.6 small, 0.6-1.2 moderate, 1.2-2 large, and 2.0-4.0 very large (Hopkins et al., 2009). Statistical differences were considered significant if p< 0.05. Statistical analyses were developed using SPSS and Figures were drawn using Graph Prism software.

RESULTS AND DISCUSSION

Results

ICC and linear correlation values ranged from 0.07 to 0.79 and from 0.49 to 0.99, respectively, according to the sampling frequencies (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz) and the task. Significant (p<0.01) and substantial (ES = large) differences were found among the CCP values recorded at different sampling frequency in all tasks (Table 1).

Table 1. Intra-class Correlation Coefficient, Linear Correlation and mean comparison of change in centroid position (CCP) by sampling frequency

<table>
<thead>
<tr>
<th>Task</th>
<th>Variable</th>
<th>Sampling frequencies</th>
<th>N</th>
<th>ICC</th>
<th>95% IC</th>
<th>BIAS</th>
<th>95%MC</th>
<th>r (p value)</th>
<th>t (p value)</th>
<th>Cohen    d (rating)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>Change in Centroid (m)</td>
<td>10v4</td>
<td>300</td>
<td>0.33</td>
<td>0.22; 0.43</td>
<td>-0.18</td>
<td>-0.57; 0.21</td>
<td>0.49 (&lt; 0.01)</td>
<td>-15.07 (&lt; 0.01)</td>
<td>-1.27, large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10v2</td>
<td>300</td>
<td>0.18</td>
<td>0.07; 0.3</td>
<td>-0.49</td>
<td>-1.33; 0.84</td>
<td>0.49 (&lt; 0.01)</td>
<td>-19.2 (&lt; 0.01)</td>
<td>-1.63, large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10v1</td>
<td>300</td>
<td>0.1</td>
<td>-0.02; 0.21</td>
<td>-</td>
<td>-2.85; 0.63</td>
<td>0.5 (&lt; 0.01)</td>
<td>-20.77 (&lt; 0.01)</td>
<td>-1.76, large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4v2</td>
<td>300</td>
<td>0.79</td>
<td>0.75; 0.83</td>
<td>-0.31</td>
<td>-0.76; 0.14</td>
<td>0.99 (&lt; 0.01)</td>
<td>-21.89 (&lt; 0.01)</td>
<td>-1.85, large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4v1</td>
<td>300</td>
<td>0.46</td>
<td>0.36; 0.55</td>
<td>-0.93</td>
<td>-2.32; 0.46</td>
<td>0.98 (&lt; 0.01)</td>
<td>-21.82 (&lt; 0.01)</td>
<td>-1.85, large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2v1</td>
<td>300</td>
<td>0.79</td>
<td>0.74; 0.83</td>
<td>-0.62</td>
<td>-1.56; 0.32</td>
<td>0.99 (&lt; 0.01)</td>
<td>-21.57 (&lt; 0.01)</td>
<td>-1.83, large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10v4</td>
<td>300</td>
<td>0.69</td>
<td>0.62; 0.76</td>
<td>-0.09</td>
<td>-0.02; 0.02</td>
<td>0.96 (&lt; 0.01)</td>
<td>-19.58 (&lt; 0.01)</td>
<td>-1.66, large</td>
</tr>
</tbody>
</table>
Assessing the Amount of

High to perfect ICCs (0.91-1) and high to perfect linear correlations (r= 0.961-1; p < 0.01) were found among the TA values obtained through all sampling frequencies added (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz) derived from the software in the three tasks. No significant (p > 0.05) and substantial (ES = trivial) differences were found among the TA values obtained with all sampling frequency in all tasks (Table 2).

Table 2. Intra-class Correlation Coefficient, Linear Correlation and mean comparison of total area (TA) by sampling frequency

<table>
<thead>
<tr>
<th>Task</th>
<th>Variable</th>
<th>Sampling frequencies</th>
<th>N</th>
<th>ICC 95% CI</th>
<th>BIAS 95%CI</th>
<th>r (p value)</th>
<th>t (p value)</th>
<th>Cohen d (rating)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10vs2</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.96 0.95; 0.97</td>
<td>2.61</td>
<td>131.18; 125.97</td>
<td>0.96 (&lt; 0.01)</td>
<td>0.66 (=507)</td>
<td>0.06, trivial</td>
</tr>
<tr>
<td>10vs2</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.96 0.95; 0.97</td>
<td>2.7</td>
<td>130.35; 124.95</td>
<td>0.96 (&lt; 0.01)</td>
<td>0.69 (=489)</td>
<td>0.06, trivial</td>
</tr>
<tr>
<td>10vs2</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.96 0.95; 0.97</td>
<td>3.71</td>
<td>125.54; 119.18</td>
<td>0.96 (&lt; 0.01)</td>
<td>0.82 (=411)</td>
<td>0.07, trivial</td>
</tr>
<tr>
<td>10vs4</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.96 0.95; 0.97</td>
<td>0.09</td>
<td>6.4; 6.22</td>
<td>1 (&lt; 0.01)</td>
<td>0.48 (=635)</td>
<td>0.04, trivial</td>
</tr>
<tr>
<td>10vs1</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.99 0.99; 0.99</td>
<td>0.57</td>
<td>22.19; 21.05</td>
<td>0.99 (&lt; 0.01)</td>
<td>0.86 (=639)</td>
<td>0.07, trivial</td>
</tr>
<tr>
<td>10vs1</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.99 0.99; 0.99</td>
<td>0.47</td>
<td>15.25; 14.29</td>
<td>0.99 (&lt; 0.01)</td>
<td>0.91 (=632)</td>
<td>0.08, trivial</td>
</tr>
<tr>
<td>10vs1</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.99 0.99; 0.99</td>
<td>462.3</td>
<td>107.73; 1032</td>
<td>0.998 (&lt; 0.01)</td>
<td>0.12 (=907)</td>
<td>0.01, trivial</td>
</tr>
<tr>
<td>10vs1</td>
<td>Total Area (m²)</td>
<td>300</td>
<td>0.99 0.99; 0.99</td>
<td>453.7</td>
<td>120.4; 1028</td>
<td>0.993 (&lt; 0.01)</td>
<td>-1.11 (=267)</td>
<td>-0.09, trivial</td>
</tr>
</tbody>
</table>

Hz: Hertz; M: metres; p: p value; r: Pearson r; t: t test; %: percentage
As an example, Figure 2 shows the CCP and TA values for each sampling frequency (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz) in Task 1.

Figure 2. Mean representation of a) total area and b) change in centroid position (i.e. Task 1). 
TA= Total Area, CCP= Change in Centroid Position.
Discussion

To our knowledge, no study has assessed the influence of the sampling frequency on the measurement of tactical positioning variables in team sports. For this reason, the aim of the present study was to assess how the sampling frequency (i.e. 1 Hz, 2 Hz, 4 Hz, and 10 Hz) affected the outcomes of the CCP and TA during tactical analysis in sports. We found significant and large differences between the values of CCP measured at different sampling frequencies. However, we did not find significant and substantial differences between TA values measured at different sampling frequencies during three controlled tasks. These results suggested that the sampling frequency could indeed affect the outcomes of tactical positioning variables, requiring different sampling frequencies for each variable.

Significant and large differences, from 0.07 to 0.79 ICC values and 0.49 to 0.99 association values, respectively, were found between the CCP values measured at different sampling frequencies (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz) during the three controlled tasks (Table 1 and 2). This suggested that the CCP values in the three controlled tasks depended on the sampling frequency. Furthermore, at low frequencies obscured relevant data. A few years ago, Duarte et al. (2010) compared an original data set with different cut-off frequencies (3-Hz and 6-Hz) of an attacker’s locomotion in order to determine what sampling frequency is more adequate for their main analysis (1 vs. 1 football sub-phase). They found less variation using a higher sampling frequency (6-Hz vs 3 Hz). So, it seems that variable represented by a single point (a single pair of spatial coordinates), where magnitude and the minimum time to change substantially may be lower than other collective tactical variables such as total area, could be more sensitive to the sampling frequency.

On the contrary, we found only trivial differences between TA values that were measured at different sampling frequencies (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz) during the three controlled tasks (Table 2). The high to perfect ICC values and linear association suggested that the addition of just 1 Hz is sufficient to accurately measure TA in these types of tasks. The structural traits and training tasks of each team sport considerably affect the magnitude of TA (Clemente et al., 2013; Frencken et al., 2011; Timmerman et al., 2017) and the time to change substantially. Therefore, in further studies, the impact of the frequencies on TA values should be expanded to several team sports and training drills. If the results of these future studies were similar, less data would be helpful in the practical setting where a rapid evaluation of training/competition loads is necessary to assess performance and inform exercise prescription (Malone et al., 2017).

Positional data for collective analyses has become an important topic in team sports analysis (Low et al., 2020; Rico-González, Pino-Ortega, et al., 2020). Usually, researchers apply the same sampling frequency to measure all tactical variables in their studies (Rico-González, Los Arcos, et al., 2020). Moreover, coaches and technical staffs should not consider the same sampling frequencies to assess all tactical variables (Rico-González, Arcos, et al., 2020). This can result in a loss of relevant data for some tactical variables (e.g. CCP in this study) but, simultaneously, a superfluous amount of data to measure others (e.g. TA). In case of CCP, one would lose relevant data and in the TA case, the excessive data could delay the analysis of the report resulting in difficulty to respond to a complex calendar, which requires rapid performance analysis (Malone et al., 2017).

CONCLUSION

The sampling frequency (i.e. 1 Hz, 2 Hz, 4 Hz and 10 Hz) does not affect the measurement of total area during tactical behaviour analysis but does significantly affect the change in centroid position measurement. Thus, despite the fact that more studies are necessary, we recommend the use of different sampling frequency to measure each tactical variable (i.e. total area and distance of centroid) in team sports. The consideration of 1 Hz is enough to accurately measure total area, while 10 Hz is suggested for the change in centroid position.

REFERENCES


74 | Journal of Coaching and Sports Science


