Journal of Sports Sciences

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rjsp20

Sequential movement pattern-mining (SMP) in field-based team-sport: A framework for quantifying spatiotemporal data and improve training specificity?

Ryan White, Anna Palczewska, Dan Weaving, Neil Collins & Ben Jones

To cite this article: Ryan White, Anna Palczewska, Dan Weaving, Neil Collins & Ben Jones (2022) Sequential movement pattern-mining (SMP) in field-based team-sport: A framework for quantifying spatiotemporal data and improve training specificity?, Journal of Sports Sciences, 40:2, 164-174, DOI: 10.1080/02640414.2021.1982484

To link to this article: https://doi.org/10.1080/02640414.2021.1982484

Published online: 27 Sep 2021.

JOURNAL OF SPORTS SCIENCES
2022, VOL. 40, NO. 2, 164–174
https://doi.org/10.1080/02640414.2021.1982484

Routledge
Taylor & Francis Group

SPORTS PERFORMANCE

OPEN ACCESS Check for updates

Sequential movement pattern-mining (SMP) in field-based team-sport: A framework for quantifying spatiotemporal data and improve training specificity?

Ryan White[a,b], Anna Palczewska[a,c], Dan Weaving[a,b], Neil Collins[a,d] and Ben Jones[a,b,d,e,f]

[a]Carnegie Applied Rugby Research (Carr) Centre, Carnegie School of Sport, Leeds Beckett University, Leeds, UK; [b]Leeds Rhinos Rugby League Club, Leeds, UK; [c]School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds, UK; [d]England Performance Unit, Rugby Football League, Leeds, UK; [e]School of Science and Technology, University of New England, Armidale, New South Wales, Australia; [f]Division of Exercise Science and Sports Medicine, Department of Human Biology, Faculty of Health Sciences, The University of Cape Town and the Sports Science Institute of South Africa, Cape Town, South Africa

ABSTRACT

Athlete external load is typically quantified as volumes or discretised threshold values using distance, speed and time. A framework accounting for the movement sequences of athletes has previously been proposed using radio frequency data. This study developed a framework to identify sequential movement sequences using GPS-derived spatiotemporal data in team-sports and establish its stability. Thirteen rugby league players during one match were analysed to demonstrate the application of the framework. The framework (Sequential Movement Pattern-mining [SMP]) applies techniques to analyse i) geospatial data (i.e., decimal degree latitude and longitude), ii) determine players turning angles, iii) improve movement descriptor assignment, thus improving movement unit formation and iv) improve the classification and identification of players' frequent SMP. The SMP framework allows for sub-sequences of movement units to be condensed, removing repeated elements, which offers a novel technique for the quantification of similarities or dis-similarities between players and playing positions. The SMP framework provides a robust and stable method that allows, for the first time the analysis of GPS-derived data and identifies the frequent SMP of field-based team-sport athletes. The application of the SMP framework in practice could optimise the outcomes of training of field-based team-sport athletes by improving training specificity.

ARTICLE HISTORY
Accepted 15 September 2021

KEYWORDS
Global positioning systems; performance analysis; sport analytics; team sports; time-motion analysis

INTRODUCTION

Quantifying the activities completed (i.e., external load) by team sport players during training and competition is a fundamental role of the applied sport scientist (Bourdon et al., 2017). Historically, these activities have been calculated using semi-automated camera, radio frequency (RF) and more recently, global positioning systems (GPS) and associated microtechnology (e.g., accelerometers, gyroscopes and magnetometers). Such systems can estimate a player's position with respect to the coordinates of a playing area and allow for the computation of displacement over time (i.e., time motion analysis) (Aughey, 2011; A. J. Sweeting et al., 2017a).

There has been considerable research using time motion methods to describe the characteristics of team sport training and competition (e.g., Cummins et al., 2013; Whitehead et al., 2019a). To date, most research reports, variables such as total distance, distance covered at varying speed thresholds, accelerations, decelerations and collision frequency to represent the external load construct, either using GPS or video time motion analysis (Buchheit et al., 2010; Glassbrook et al., 2019). Often significant differences exist between positional groups (Glassbrook et al., 2019) and standards of competition (Weaving et al., 2019; Whitehead et al., 2019a, 2019b) during match play. These data are important for practitioners to ensure aspects of training prescription (e.g., technical-tactical or small-sided-games) replicate the general characteristics required during competition, but other important characteristics (e.g., movement strategies and patterns) may exist, which are yet to be captured.

External load is typically reported as volumes (e.g., total distance) or as discretised threshold values (e.g., the distance covered at different speed thresholds) (Bourdon et al., 2017). These variables are often arbitrarily chosen (e.g., speed thresholds), and only provide a global aggregation of a series of physical actions (Dalton-Barron et al., 2020). While still important, these findings provide little context for how a player accumulates a given external load. For example, the same player can theoretically cover the same total, high-speed and sprint distance between two matches but achieve these in different ways. Therefore, quantifying the specific movement patterns of athletes (e.g., movement angles and running velocities) and their frequency would provide a more granular evaluation of the external load construct and could increase the specificity of training practices.

One method that demonstrates promise in this regard used data mining and pattern recognition techniques to discover the unique repetitive movement patterns performed by netball playing positions during match-play (A. J. Sweeting et al.,

CONTACT Ryan White ryan.white@leedsbeckett.ac.uk Headingley Campus, Leeds Beckett University, West Yorkshire, G12 Cavendish Hall, LS6 3QS, UK
This article has been republished with minor changes. These changes do not impact the academic content of the article.

2017a). A. J. Sweeting et al. (2017a) used spatiotemporal data collected via a RF tracking system and unsupervised learning techniques (i.e., k-means clustering) to identify the frequently recurring movement sequences of elite netball players, thus allowing for the movement strategies of each playing position to be directly compared to one another. A more granular approach such as this may offer a unique alternative when it comes to the quantification of the external loads of training and matches, as it will allow practitioners to understand how players accumulate them (i.e., the exact movement strategies utilised) and thus allow for improved training practice, monitoring methods and return-to-play protocols.

It is unclear whether the A. J. Sweeting et al. (2017a) framework can be directly applied to field-based team sports using GPS data, due to the difference in data types (i.e., x, y cartesian coordinates vs. decimal degree latitude and longitude). There has also been limited evaluation of the reliability and stability of the unsupervised learning techniques (i.e., k-means clustering) previously used to classify a player's movement descriptors (e.g., velocity and acceleration) both within a single training session or match and across multiple matches or training sessions. This is due to the nature of the k-means clustering algorithm, which has previously been deemed unsuitable when attempting to classify team sport athletes' instantaneous velocity thresholds (Park et al., 2019).

Therefore, the aims of this study were to i): formulate a new methodological framework to identify sequential movement sequences using GPS (example data from professional rugby league) and ii): compare the stability of each step for the new methodological framework to that previously proposed using RF tracking (A. J. Sweeting et al., 2017a).

METHODS

Study Design

This study developed a methodological framework (aim i), termed the Sequential Movement Pattern-mining (SMP) framework, to identify sequential movement sequences using GPS-derived data. The SMP framework was applied using an un-edited (i.e., movement sub-sequences remained identical length) and condensed (i.e., movement sub-sequences were shortened) method. A sample data set of 13 rugby league players during one match was used to determine the stability of each framework. GPS data from any team sport could have been used for this study, rugby league data were chosen due to convenience sampling. The data set contained 778,817 samples and were deemed sufficient to determine the stability of each framework (i.e., were the same results returned when the same data set was analysed twice using the same framework). The dataset was analysed twice using the SMP framework and twice using the A. J. Sweeting et al. (2017a) framework to determine the stability of each step of the respective frameworks (aim ii).

Standardisation steps were taken to allow for the direct comparison of the experimental results between the SMP framework and the A. J. Sweeting et al. (2017a) framework. Minor modifications were applied to the A. J. Sweeting et al. (2017a) framework to allow for GPS-derived data to be utilised. These served to standardise the initial data processing step. Minor modifications included the method for calculating a player's directional data, the use of GPS provided locomotive time-series data (velocity and acceleration) and the use of a standardised movement unit dictionary for both frameworks (Table 1).

Sample Data

Spatiotemporal data were collected via Catapult (Optimeye S5 10 Hz, Catapult Sports, Melbourne, Australia) GPS units from the activity profiles of 13 professional male rugby league players during one competitive Super League match. The participants wore a Catapult Optimeye S5 mobile GPS unit, measuring $96 \times 52 \times 14$ mm, positioned between the shoulder blades. Each player's 10 Hz spatiotemporal data (geographic coordinates and sample time) and locomotive data (velocity and acceleration) represented only active time on the field (i.e., the data were trimmed to not include time on bench) and was exported into R statistical software (R: A language and environment for statistical

Table 1. The standardised movement unit dictionary.

Movement Unit characters	Movement units
a	Walk Deceleration Straight
b	Walk Deceleration Acute-Change
c	Walk Deceleration Large-Change
d	Walk Deceleration Backwards
e	Walk Neutral Straight
f	Walk Neutral Acute-Change
g	Walk Neutral Large-Change
h	Walk Neutral Backwards
i	Walk Acceleration Straight
j	Walk Acceleration Acute-Change
k	Walk Acceleration Large-Change
l	Walk Acceleration Backwards
m	Jog Deceleration Straight
n	Jog Deceleration Acute-Change
o	Jog Deceleration Large-Change
p	Jog Deceleration Backwards
q	Jog Neutral Straight
r	Jog Neutral Acute-Change
s	Jog Neutral Large-Change
t	Jog Neutral Backwards
u	Jog Acceleration Straight
v	Jog Acceleration Acute-Change
w	Jog Acceleration Large-Change
x	Jog Acceleration Backwards
y	Run Deceleration Straight
z	Run Deceleration Acute-Change
A	Run Deceleration Large-Change
B	Run Deceleration Backwards
C	Run Neutral Straight
D	Run Neutral Acute-Change
E	Run Neutral Large-Change
F	Run Neutral Backwards
G	Run Acceleration Straight
H	Run Acceleration Acute-Change
I	Run Acceleration Large-Change
J	Run Acceleration Backwards
K	Sprint Deceleration Straight
L	Sprint Deceleration Acute-Change
M	Sprint Deceleration Large-Change
N	Sprint Deceleration Backwards
O	Sprint Neutral Straight
P	Sprint Neutral Acute-Change
Q	Sprint Neutral Large-Change
R	Sprint Neutral Backwards
S	Sprint Acceleration Straight
T	Sprint Acceleration Acute-Change
U	Sprint Acceleration Large-Change
V	Sprint Acceleration Backwards

computing, Vienna Austria) for further analysis. The study was approved by the University Ethics Committee and participants provided signed informed consent.

Sequential Movement Pattern-mining (SMP) Framework (aim i)

The SMP framework is a series of logical stepwise calculations that utilised data mining techniques to identify mathematical descriptions of patterns and regularities in a data set (Fayyad et al., 1996). The framework makes use of spatiotemporal data, in the form of geospatial coordinates (i.e., latitude and longitude) in combination with locomotive data (i.e., instantaneous velocity ($m{\cdot}s^{-1}$) and acceleration ($m{\cdot}s^{-2}$)). An overview of the SMP framework and the four steps are visualised in Figure 1.

Step 1. Formation of movement descriptors

Step 1 of the SMP framework computes the required data that defined a player's state in time and is known as the players' movement descriptors (Figure 1, Step 1). The movement descriptors are velocity ($m{\cdot}s^{-1}$), acceleration ($m{\cdot}s^{-2}$) and turning angles (Ψ). Doppler-derived instantaneous velocity ($m{\cdot}s^{-1}$) and acceleration ($m{\cdot}s^{-2}$) were utilised because the Doppler shift method has been reported to demonstrate a higher level of precision and reduced error compared with positional differentiation techniques (Townshend et al., 2008 & Malone et al., 2017). The framework borrowed from research in animal science (Dodge et al., 2008; Gurarie et al., 2009 & Zhang et al., 2015) and human transport behaviour (Zheng et al., 2008) to compute a time-series of turning angles, which represented a player's change in direction between consecutive samples within the GPS-derived coordinate system.

Turning angle calculation

The calculation of turning angles between consecutive samples within the geospatial time series is a two-phase process. This process involved the calculation of the bearing for each time point and then computed turning angle between successive samples. In phase 1 the bearing or heading angle between consecutive sampling points was calculated. The bearing is defined as a direction or angle, between the north–south line of the Earth or the meridian and the line connecting the target and reference point (Figure 2a).

The bearing between consecutive geospatial samples is denoted as positive (0° to 180°) or negative (0° to −180°). A positive bearing value is situated within the East and South quadrants and a negative bearing value is situated within the West and South quadrants. Thus, a bearing of 0° indicates due North, +90° indicates due East, +180° or −180° indicates due South and −90° indicates due West (Figure 2b).

The bearing (β) was calculated as:

$$\beta = \mathrm{atan2}(X, Y) \tag{1}$$

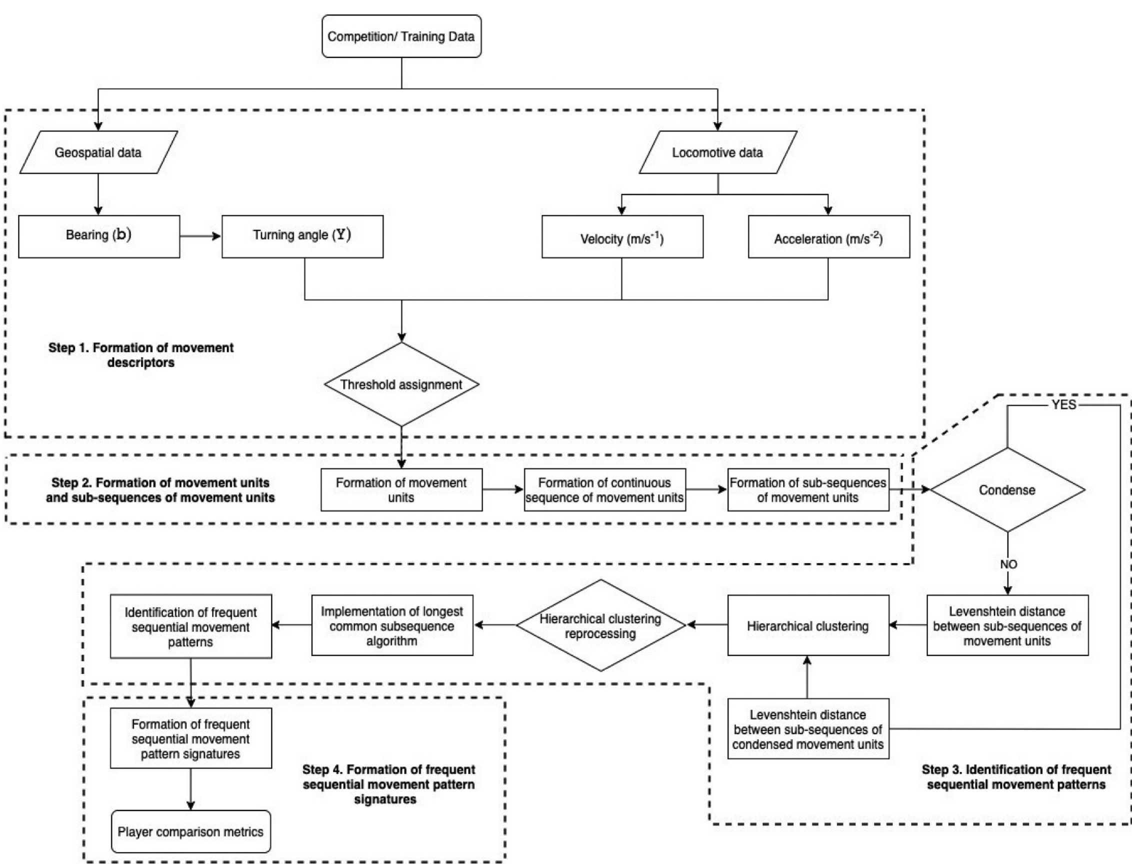

Figure 1. The 4-step methodological framework proposed for field-based team sports using GPS data – the SMP framework.

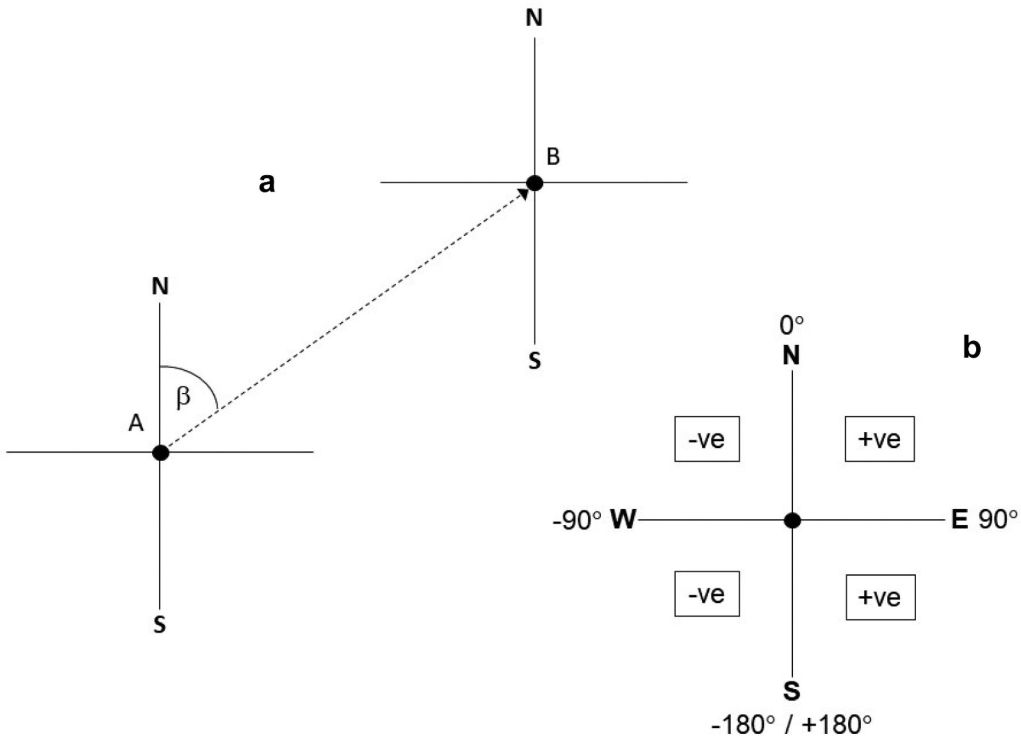

Figure 2. The bearing (β) relative to the North-South line from point A to point B (a) and the compass quadrants (b).

Where, X and Y were calculated as:

$$X = \cos(\theta B)\sin(\Delta L) \quad (2)$$

$$Y = \cos(\theta A)\sin(\theta B) - \sin(\theta A)\cos(\theta B)\cos(\Delta L) \quad (3)$$

"A" and "B" denote two different samples that consist of a pair of decimal coordinates, and "Θ" and "L" denote latitude and longitude. The change in longitude between A and B is represented with ΔL. Once the bearing values were known, the turning angles (Ψ) were computed.

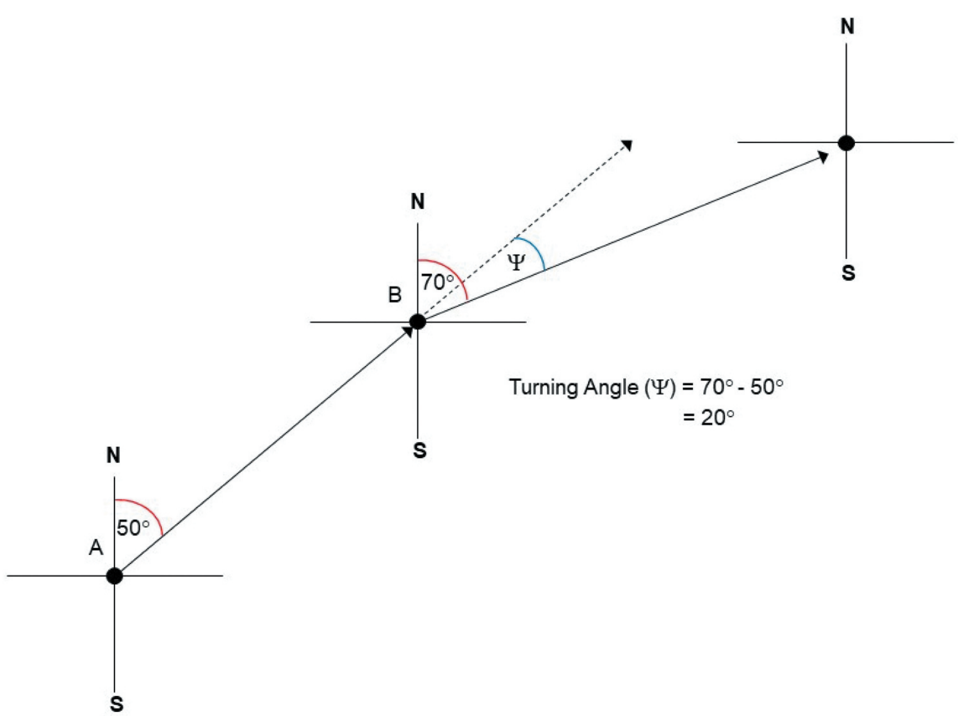

Figure 3. Example for computing the turning angle from consecutive GPS sample bearing values.

A turning angle (Ψ) is defined as the change in the heading between consecutive GPS samples (Figure 3) and is calculated as follows:

$$\Upsilon_i = \beta_i - \beta_{i-1} \quad (4)$$

The turning angle (Θ_i) at any sample was the change in absolute bearing (β) between consecutive GPS samples.

A combination of the above movement descriptors creates a movement unit that provides granular information, sample-by-sample, regarding the speed a player was travelling, the players state of acceleration or deceleration and the players' direction of travel at each temporal point.

Step 2. Formation of movement units and sub-sequences of movement units

Step 2 of the SMP framework forms the movement units, continuous sequences of movement units and identifies the sub-sequences of movement units (Figure 1, Step 2). To form a movement unit, velocity, acceleration and turning angle were discretised by applying threshold assignment values to each movement descriptor. The threshold values used represent the general movement characteristics of players and were derived from match and training literature within rugby league (Cummins et al., 2013). The thresholds were assigned a qualitative label and represent arbitrary descriptors rather than specific quantities (Table 2).

From the thresholds presented in Table 2, each unique concatenated movement descriptor combination (e.g., Jog-Acceleration-Straight), was assigned an identification code consisting of an upper- or lower-case alphabetic letter forming a movement unit dictionary (for example, see Table 1). The concatenated movement descriptors at each temporal point create a time-series of continuous movement units that are subsequently represented as a string of lower- or upper-case alphabetic letters. To isolate discrete sub-sequences of movement units, any period during which the athlete moved at a rate lower than 1.20 $m{\cdot}s^{-1}$ was classified as "inactivity" and delineated the continuous sequence of movement units to create sub-sequences of movement units. Additionally, any sub-sequences of movement units must exceed the movement threshold for at least 5 seconds in order for the sub-sequence to be considered further.

Step 3. Identification of frequent sequential movement patterns

Step 3 of the SMP framework identifies player-specific frequent SMP and provides player comparison metrics. Within this step, the sub-sequences of movement units are either condensed or left un-edited (i.e., a sub-sequence of actual length) (Figure 1, Step 3). Condensing sub-sequences of movement units removed consecutive repeated elements, leaving only the unique aspects of each sub-sequence of movement units (e.g., un-edited = i-i-G-G-i *vs.* condensed = i-G-i). The removal of continuous repeated elements results in shorter sub-sequences of movement units and may provide a novel method for identifying important similarities or dis-similarities, between sub-sequences and athletes. The similarity between sub-sequences of movement units, either condensed or un-edited, was quantified using the Levenshtein distance implementation in the R *stringdist* package (A. J. Sweeting et al., 2017a; van der Loo, 2014). The Levenshtein distance provides a numeric value representing the number of insertions, deletions or substitutions that are required to convert one movement unit sub-sequence into another (van der Loo, 2014). Thus, providing a "similarity" matrix allowing one string to be compared to another. Similar sub-sequences of movement units were then grouped into 25 clusters using a hierarchical cluster analysis (Ward, 1963).

The hierarchical cluster analysis results were then reprocessed, in which all single element clusters were identified and reassigned to their nearest cluster. The reassignment process involved computing the average Levenshtein distance between each single element and all other elements within each cluster. The single element was then reassigned to the cluster it was most similar to (e.g., the minimum average distance). This method ensured a dynamic data classification process, preventing the formation and subsequent exclusion of single elements and allowing each player's movement unit sub-sequence profile to dictate the number of clusters returned.

To discover the most common frequent SMP within each of the identified clusters for each player, the longest common subsequence algorithm, using the R *PTXQC* package (Bielow et al., 2016) was used to discover all of the common elements within sub-sequences of movement units whilst retaining the sequential order. Therefore, the longest common subsequence or frequent SMP of each player performed across the dataset could be identified.

Table 2. The movement descriptors and threshold assignment values.

Velocity Descriptor	Velocity Threshold ($m{\cdot}s^{-1}$)	Acceleration Descriptor	Acceleration Threshold ($m{\cdot}s^{-2}$)	Turning Angle Descriptor	Turning Angle Threshold (Θ)
Walk	0.00 to <1.70	Deceleration	Min accel to ≤-0.20	Straight	0.00 to <10.00
Jog	≥1.70 to ≤3.90	Neutral	>-0.20 to <0.20	Acute-change	≥10.00 to <45.00
Run	>3.90 to <5.00	Acceleration	≥0.20 to max accel	Large-change	≥45.00 to <90.00
Sprint	≥5.00	NA	NA	Backwards	≥90.00 to 180.00

Step 4. Formation of frequent sequential movement pattern (SMP) signatures

Step 4 of the SMP framework created player-specific frequent SMP signatures (Figure 1, Step 4). The frequent SMP signatures were created by 1) identifying a union of frequent SMP across all players, 2) cross-matching these frequent SMP to each player's individual dataset and 3) filtering out frequent SMP which were not present in all of the players' datasets. A percentage presence vector for each player was subsequently calculated and represents the relative presence (%) of a frequent SMP within each player's movement unit sub-sequence profile. The Minkowski distance implemented in the R *stats* package was used to quantify the distance between players using the frequent SMP relative presence results (R Core Team, 2020). The lower the Minkowski distance value between two players the more similar their frequent sequential movement patterns are and vice versa.

Statistical Analysis and Framework Comparison (aim ii)

Identical datasets, containing the activity profiles of 13 individual players were analysed twice using the SMP framework (both un-edited and condensed) and twice using the A. J. Sweeting et al. (2017a) framework. The stability of the SMP framework and the A. J. Sweeting et al. (2017a) framework was compared at steps 2–4.

Movement unit formation stability (SMP framework: Step 2)

Step 2 was completed twice using both frameworks, to determine the total count of movement units. An equal number of movement units should be formed when analysing an identical dataset twice. The difference between the total count of movement units for each framework highlights the respective stability of each framework at step 2.

Clustering stability of the sub-sequences of movement units (SMP framework: Step 3)

Step 3 was completed twice using both frameworks, and framework stability at this step was assessed using the count of null returns and the count of single element clusters within each trial. A null return occurred when the longest common subsequence algorithm failed to identify a result within a cluster group, and single element clusters prevented the implementation of the longest common subsequence algorithm, which both highlight the framework instability.

Frequent SMP and frequent SMP signature stability (SMP framework: Step 4)

Step 4 was completed twice using both frameworks, to determine the stability of the respective frameworks at this step. Stability was determined by evaluating whether the same results were found for trial 1 and trial 2 within each framework. These included the total count of i) frequent SMP identified, ii) unique frequent SMP identified and iii) the differences observed within the frequent SMP components. Frequent SMP signature stability was assessed between trial 1 and trial 2 by matrix subtraction. A difference in frequent SMP signature post-matrix subtraction represented a framework's inability to identify the exact same frequent SMP signature between trial 1 and trial 2 of the identical datasets.

RESULTS

Movement unit formation stability

No differences were observed in the count of movement units for the SMP framework (both un-edited and condensed) between trial 1 and trial 2 (Table 3). A difference was observed between trial 1 and trial 2 for the A. J. Sweeting et al. (2017a) framework at each movement unit except for Run Neutral Backwards and Sprint Acceleration Backwards (Table 3).

Clustering stability of the sub-sequences of movement units

No null returns were identified nor any single element clusters formed in either trial of the SMP framework (both unedited and condensed) (Table 4). Five null returns and 118 single element clusters were identified in trial 1, and six null returns and 117 single element clusters were identified in trial 2 (Table 4) for the A. J. Sweeting et al. (2017a) framework.

Frequent SMP identification stability

An identical count of frequent SMP and unique frequent SMP was identified between trial 1 and trial 2 using the SMP framework (both unedited and condensed) (Table 5). There were also no observable differences in the actual components of the identified frequent SMP in either trial using the SMP framework (both unedited and condensed) (Table 5). Between trial 1 and trial 2, 320 and 319 frequent SMPs were observed and 275 and 262 unique frequent SMPs were observed using the A. J. Sweeting et al. (2017a) framework (Table 5). Additionally, 126 differences were identified within the actual frequent SMP components (Table 5).

Frequent SMP signature stability

There were no differences in frequent SMP signatures between trial 1 and trial 2 for the SMP framework (both unedited and condensed) (Figure 4). Positive (i.e., coloured dark blue) and negative (i.e., coloured yellow) differences in frequent SMP signatures were observed between trial 1 and trial 2 using the A. J. Sweeting et al. (2017a) framework (Figure 5).

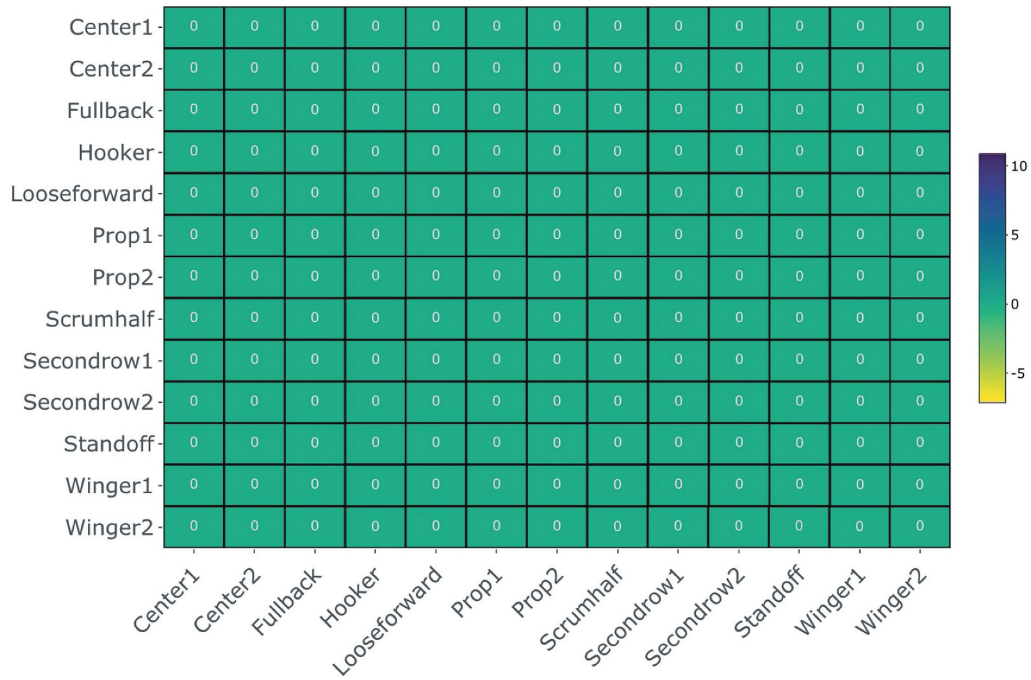

Figure 4. The difference in the SMP framework derived frequent SMP signatures between trials.

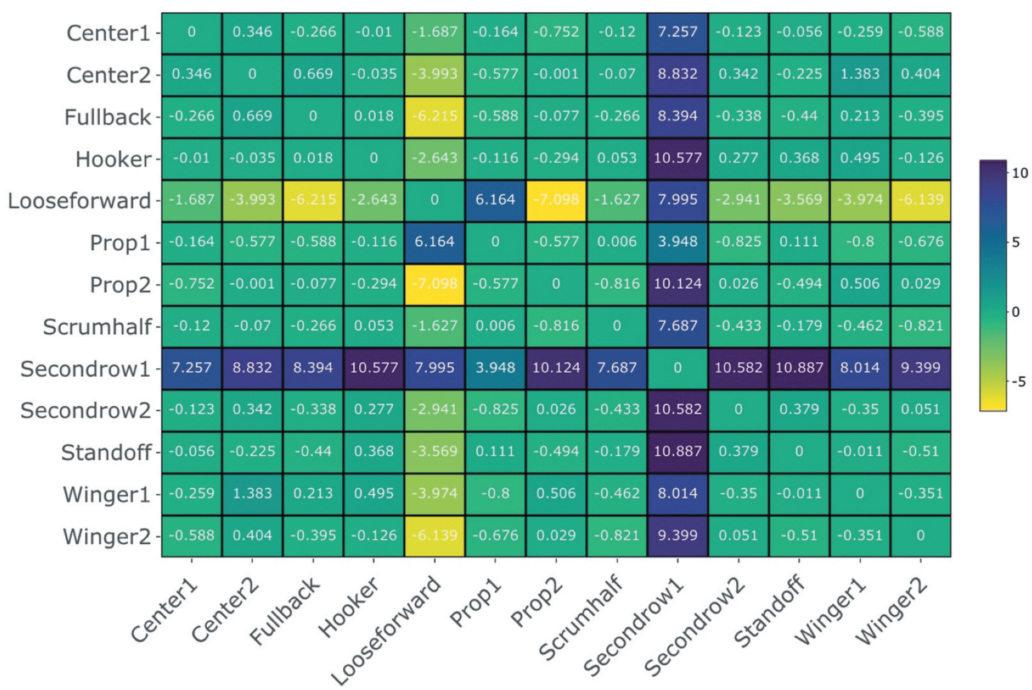

Figure 5. The difference in A. J. Sweeting et al. (2017a) framework derived frequent SMP signatures between trials.

DISCUSSION

The purpose of this study was to formulate a new methodological (SMP) framework to identify sequential movement sequences using GPS and compare the stability of each step to a previously proposed framework that used RF tracking (A. J. Sweeting et al., 2017a). When the SMP framework (both un-edited and condensed) was applied to an identical dataset twice (i.e., trial 1 and trial 2) a greater framework stability was demonstrated at each step compared to the previous framework (A. J. Sweeting et al., 2017a). The SMP framework provides a stable movement unit formation (Table 3), an improved clustering stability of sub-sequences of movement units (Table 4), a reliable frequent SMP identification process and a robust frequent SMP signature stability, which displayed no variation in the SMP patterns identified (Figure 4).

Table 3. The distribution of data points for each movement unit variable and the difference in distributions between trials.

		SMP Framework			A. J. Sweeting et al. (2017a)		
Movement Unit characters	Movement Units	T1	T2	Diff	T1	T2	Diff
a	Walk Deceleration Straight	44,395	44,395	0	22,965	31,394	8429 **
b	Walk Deceleration Acute-Change	50,267	50,267	0	10,476	13,089	2613 **
c	Walk Deceleration Large-Change	11,241	11,241	0	5042	5631	589 **
d	Walk Deceleration Backwards	3771	3771	0	2160	2520	360 **
e	Walk Neutral Straight	242,603	242,603	0	169,702	172,028	2326 **
f	Walk Neutral Acute-Change	116,969	116,969	0	53,973	54,658	685 **
g	Walk Neutral Large-Change	27,294	27,294	0	25,347	26,160	813 **
h	Walk Neutral Backwards	8172	8172	0	14,178	13,975	203 **
i	Walk Acceleration Straight	39,594	39,594	0	30,002	20,148	9854 **
j	Walk Acceleration Acute-Change	33,781	33,781	0	10,617	7235	3382 **
k	Walk Acceleration Large-Change	4095	4095	0	4492	2892	1600 **
l	Walk Acceleration Backwards	1160	1160	0	2094	1292	802 **
m	Jog Deceleration Straight	22,292	22,292	0	19,937	19,355	582 **
n	Jog Deceleration Acute-Change	21,098	21,098	0	13,953	13,786	167 **
o	Jog Deceleration Large-Change	2465	2465	0	5232	5223	9 **
p	Jog Deceleration Backwards	680	680	0	1882	1802	80 **
q	Jog Neutral Straight	17,603	17,603	0	115,211	113,724	1487 **
r	Jog Neutral Acute-Change	10,665	10,665	0	53,785	52,185	1600 **
s	Jog Neutral Large-Change	498	498	0	10,003	9591	412 **
t	Jog Neutral Backwards	91	91	0	2005	1971	34 **
u	Jog Acceleration Straight	41,439	41,439	0	22,570	26,298	3728 **
v	Jog Acceleration Acute-Change	22,409	22,409	0	10,132	10,854	722 **
w	Jog Acceleration Large-Change	690	690	0	1780	1686	94 **
x	Jog Acceleration Backwards	149	149	0	434	372	62 **
y	Run Deceleration Straight	2842	2842	0	11,826	11,226	600 **
z	Run Deceleration Acute-Change	1691	1691	0	7706	7305	401 **
A	Run Deceleration Large-Change	73	73	0	2022	1922	100 **
B	Run Deceleration Backwards	15	15	0	488	464	24 **
C	Run Neutral Straight	2709	2709	0	34,333	33,070	1263 **
D	Run Neutral Acute-Change	952	952	0	12,822	12,580	242 **
E	Run Neutral Large-Change	18	18	0	1420	1370	50 **
F	Run Neutral Backwards	0	0	0	178	178	0
G	Run Acceleration Straight	6978	6978	0	25,658	28,026	2368 **
H	Run Acceleration Acute-Change	2662	2662	0	8213	8313	100 **
I	Run Acceleration Large-Change	55	55	0	620	605	15 **
J	Run Acceleration Backwards	5	5	0	88	86	2 **
K	Sprint Deceleration Straight	1711	1711	0	4253	3856	397 **
L	Sprint Deceleration Acute-Change	554	554	0	1547	1489	58 **
M	Sprint Deceleration Large-Change	10	10	0	238	234	4 **
N	Sprint Deceleration Backwards	2	2	0	28	25	3 **
O	Sprint Neutral Straight	1481	1481	0	13,778	14,052	274 **
P	Sprint Neutral Acute-Change	318	318	0	2782	2824	42 **
Q	Sprint Neutral Large-Change	2	2	0	141	151	10 **
R	Sprint Neutral Backwards	0	0	0	7	9	2 **
S	Sprint Acceleration Straight	3870	3870	0	11,578	12,058	480 **
T	Sprint Acceleration Acute-Change	1115	1115	0	2655	2644	11 **
U	Sprint Acceleration Large-Change	14	14	0	139	136	3 **
V	Sprint Acceleration Backwards	2	2	0	8	8	0

T1 = trial one, T2 = trial two, Diff = difference, ** = indicate a difference between trials > 0.00

Table 4. The total count of single element clusters and the longest common subsequence algorithm null returns.

Framework	Trial	Null returns	Single element clusters
SMP (un-edited)	1	0	0
	2	0	0
SMP (condensed)	1	0	0
	2	0	0
A. J. Sweeting et al. (2017a)	1	5	118
	2	6	117

Table 5. The total count of frequent SMP identified, unique frequent SMP identified and frequent SMP differences between methods and trials.

Framework	Trial	Frequent SMP identified	Unique Frequent SMP identified	Frequent SMP differences
SMP (un-edited)	1	196	131	0
	2	196	131	
SMP (condensed)	1	192	87	0
	2	192	87	
A. J. Sweeting et al. (2017a)	1	320	275	126
	2	319	262	

Movement unit formation stability (SMP Framework: Step 2)

The SMP framework proposed in this study, utilised global threshold values for the formation of each player's movement units. The formation of movement units in both trials remained stable and resulted in identical counts of movement units formed at each movement unit feature (Table 3). The use of global threshold values provides a stable and robust movement unit formation and improved framework stability. Equivalent stability was not observed during movement unit formation using the previous framework (A. J. Sweeting et al., 2017a). The sampling frequency for data included within the SMP and the A. J. Sweeting et al. (2017a) frameworks were different (10 vs. 100 Hz). These differences would not affect the stability of the frameworks due to the relative nature of the methods used, and the specifics are discussed below.

The global threshold values utilised within the SMP framework are an alternative method to the k-means clustering technique used in the previous framework (A. J. Sweeting et al., 2017a). Global threshold values can be used to provide benchmarks for the examination of positional and individual differences both within and between teams over time (A.J. Sweeting et al., 2017b; Park et al., 2019). The k-means algorithm operates iteratively and assigns each observation to one of the k-specified clusters based on the closest centroid (Park et al., 2019). The algorithm may be inappropriate given that the data-mining technique assumes that each movement descriptor's data are independent and uncorrelated in nature (i.e., not related between successive samples) (Park et al., 2019). Movement descriptors are not independent, and it has been suggested that there is limited foundation in using this data-mining technique with these data types, other than that it will provide thresholds within the data (Park et al., 2019). The algorithm lacks any underlying assumptions regarding the distribution of the data, meaning that identical data can be analysed, and the algorithm can return different results introducing instability into the framework. This was observed within the study, when the A. J. Sweeting et al. (2017a) framework was applied twice to an identical data set, and different movement unit formations were observed between trials (Table 3). For example, "i" (i.e., "Walk-Acceleration-Straight") was formed 30,002 in trial 1 and 20,148 times in trial 2, with a difference of 9854 observed (Table 3). The observed difference indicated that different thresholds were identified each time the dataset was analysed and represented an unacceptable level of instability within step 2 of the framework. Alternatively, by making use of global threshold values, the same thresholds were applied in both trials, which resulted in the identification of the same movement units in trials 1 and 2 and thus improved the SMP frameworks' stability at this step.

Clustering stability of the sub-sequences of movement units and Frequent SMP signature stability (SMP framework: Steps 3–4)

The SMP framework reprocessed the hierarchical cluster analysis to provide stability within steps 3 and 4. Single element clusters were identified prior to the application of the longest common subsequence algorithm and reassigned to the next nearest cluster. This method ensured a dynamic data classification process, preventing the formation and subsequent exclusion of single elements and allowing each player's movement unit sub-sequence profile to dictate the number of clusters returned. Additionally, this method prevented "null" returns of the longest common subsequence algorithm (e.g., when the algorithm failed to find a frequent SMP) as the clusters had a more even distribution. Reprocessing of the hierarchical cluster analysis resulted in the same data being returned when the SMP framework was applied twice (i.e., trial 1 and trial 2) to the same data set, and no "null" returns nor single element clusters were observed (Table 4). This stability was further demonstrated by the equivalent count of frequent SMP and unique frequent SMP identified between trial 1 and trial 2 (Table 5). Subsequently, identical SMP signatures were calculated for each player, each time the dataset was analysed (Figure 4), highlighting a stable and robust framework stability.

This study shows that equivalent stability was not observed within steps 3 and 4 using the A. J. Sweeting et al. (2017a) framework. This may be due to single element clusters or null returns (Table 4), which were observed when the A. J. Sweeting et al. (2017a) framework was applied to the GPS data, which informed the reprocessing of the hierarchical clusters in the SMP framework. A. J. Sweeting et al. (2017a) did not report how these were dealt with within their study using RF data; therefore, it is unknown if these were observed in their respective data.

Practically if "null" returns and single element clusters are not prevented (i.e., as observed with the A. J. Sweeting et al. (2017a) framework), then the associated sub-sequences of movement units are no longer featured in further analysis, thus promoting a false frequent SMP profile for each player. This is more likely to occur in a player who demonstrates a large degree of movement variability. The exclusion of these sub-sequences of movement units may therefore identify an incorrect frequent SMP signature. This instability is demonstrated by the difference in the total count of frequent and unique frequent SMP observed, when the A. J. Sweeting et al. (2017a) framework was applied to the same data twice (Table 5). A different SMP signature was calculated for each player, each time the dataset was analysed (Figure 5). The accurate quantification of an athlete's frequent SMP signature will support practitioners in both improving training specificity and also return to training and play protocols.

Potential practical applications

The application of the SMP framework can facilitate a deeper evaluation of the training prescription by allowing practitioners to monitor and evaluate their athletes with respect to specific movement sequences rather than global quantification (e.g., total distance and high-speed) of the external load. This can allow the differences in exposure of specific movement sequences between training drills to be compared or an athlete's exposure (or lack of) to movement sequences evaluated across longitudinal periods of time (e.g., 6 weeks of training programme). For example, in a rehabilitation context, if a player's position (e.g., centre) required frequent exposure to a movement sequence involving "sprint, acceleration with a large change of direction" during competition, practitioners could use the SMP framework to establish which training drills can expose players to this movement sequence and also monitor the athletes' progression of exposure to that movement sequence through the return to play process. Applied together, this could aid the specificity of training drill and programme design and additionally, potentially assist in talent identification by allowing for the discovery of players who utilise similar movement strategies to already established or desirable players.

CONCLUSION

The SMP framework provides a novel method that for the first time allows GPS-derived data to be reliably analysed to discover previously unknown frequent SMP signatures of field-based team-sport athletes. The SMP framework demonstrated excellent stability, advancing previous frameworks (A. J. Sweeting et al., 2017a). The SMP framework deployed new data pre-processing steps, used more accurate Doppler-derived instantaneous velocity and acceleration data and unique data-mining techniques to provide robust and reliable movement signatures for athletes. The SMP framework provides a clear and concise methodology, with straightforward rationale and a step-by-step approach that may serve as a foundation for future research. Future research should focus on improving different elements of the framework, for example, more clearly defining threshold values for the movement descriptors, exploring alternatives to the LCS algorithm to extract athletes' frequent SMPs and extending the framework to provide more practical outcomes to improve training specificity.

Acknowledgments

The authors would like to acknowledge Dr Sean Williams from the University of Bath for his time and valuable feedback regarding the analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

ORCID

Dan Weaving http://orcid.org/0000-0002-4348-9681
Ben Jones http://orcid.org/0000-0002-4274-6236

References

Aughey, R. J. (2011, September). Applications of GPS Technologies to Field Sports. *International Journal of Sports Physiology and Performance*, *6*(3), 295–310. https://doi.org/10.1123/ijspp.6.3.295

Bielow, C., Mastrobuoni, G., & Kempa, S. (2016, March). Proteomics Quality Control: Quality Control Software for MaxQuant Results. *Journal of Proteome Research*, *15*(3), 777–787. https://doi.org/10.1021/acs.jproteome.5b00780

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., & Cable, N. T. (2017, April). Monitoring Athlete Training Loads: Consensus Statement. *International Journal of Sports Physiology and Performance*, *12*(s2), S2-161-S2-170. https://doi.org/10.1123/IJSPP.2017-0208

Buchheit, M., Mendez-Villanueva, A., Simpson, B. M., & Bourdon, P. C. (2010, November). Match Running Performance and Fitness in Youth Soccer. *International Journal of Sports Medicine*, *31*(11), 818–825. https://doi.org/10.1055/s-0030-1262838

Cummins, C., Orr, R., O'Connor, H., & West, C. (2013, October). Global Positioning Systems (GPS) and Microtechnology Sensors in Team Sports: A Systematic Review. *Sports Medicine*, *43*(10), 1025–1042. https://doi.org/10.1007/s40279-013-0069-2

Dalton-Barron, N., Whitehead, S., Roe, G., Cummins, C., Beggs, C., & Jones, B. (2020, May). Time to Embrace the Complexity When Analysing GPS Data? A Systematic Review of Contextual Factors on Match Running in Rugby League. *Journal of Sports Sciences*, *38*(10), 1161–1180. https://doi.org/10.1080/02640414.2020.1745446

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008, September). Towards a Taxonomy of Movement Patterns. *Information Visualization*, *7*(3–4), 240–252. https://doi.org/10.1057/PALGRAVE.IVS.9500182

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996, March). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37. https://doi.org/10.1609/aimag.v17i3.1230

Glassbrook, D. J., Doyle, T. L. A., Alderson, J. A., & Fuller, J. T. (2019, June). The Demands of Professional Rugby League Match-Play: A Meta-Analysis. *Sports Medicine - Open*, *5*(1), 24. https://doi.org/10.1186/s40798-019-0197-9

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A Novel Method for Identifying Behavioural Changes in Animal Movement Data. *Ecology Letters*, *12*(5), 395–408. https://doi.org/10.1111/j.1461-0248.2009.01293.x

Loo, M. (2014). van der (2014) The Stringdist Package for Approximate String Matching. *The R Journal*, *6*(1), 111–112. https://doi.org/10.32614/RJ-2014-011

Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017, April). Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *International Journal of Sports Physiology and Performance*, *12*(s2), S2-18–S2-26. https://doi.org/10.1123/ijspp.2016-0236

Park, L. A., Scott, D., & Lovell, R. (2019). Velocity zone classification in elite women's football: Where do we draw the lines? *Science and Medicine in Football*, *3*(1), 21–28. https://doi.org/10.1080/24733938.2018.1517947

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. URL https://www.R-project.org/

Sweeting, A. J., Aughey, R. J., Cormack, S. J., & Morgan, S. (2017a, December). Discovering Frequently Recurring Movement Sequences in Team-Sport Athlete Spatiotemporal Data. *Journal of Sports Sciences*, *35*(24), 2439–2445. https://doi.org/10.1080/02640414.2016.1273536

Sweeting, A. J., Cormack, S. J., Morgan, S., & Aughey, R. J. (2017b). When is a sprint a sprint? A review of the analysis of team-sport athlete activity profile. *Frontiers in Physiology*, *8*, 432. https://doi.org/10.3389/fphys.2017.00432

Townshend, A., Worringham, C., & Stewart, I. (2008, January). Assessment of Speed and Position during Human Locomotion Using Nondifferential GPS. *Medicine and Science in Sports and Exercise*, *40*(1), 124–132. https://doi.org/10.1249/mss.0b013e3181590bc2
https://doi.org/10.32614/RJ-2011-015

Ward, J. H. (1963, March). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Weaving, D., Sawczuk, T., Williams, S., Scott, T., Till, K., Beggs, C., Johnston, R. D., & Jones, B. (2019, February). The Peak Duration-Specific Locomotor Demands and Concurrent Collision

Frequencies of European Super League Rugby. *Journal of Sports Sciences*, *37*(3), 322–330. https://doi.org/10.1080/02640414.2018.1500425

Whitehead, S., Till, K., Weaving, D., Dalton-Barron, N., Ireton, M., & Jones, B. (2021, July). The Duration-Specific Peak Average Running Speeds of European Super League Academy Rugby League Match Play. *Journal of Strength and Conditioning Research*, *35*(7), 1964-1971. doi: 10.1519/JSC.0000000000003016. https://doi.org/10.1519/JSC.0000000000003016

Whitehead, S., Till, K., Weaving, D., Hunwicks, R., Pacey, R., & Jones, B. (2019b, January). Whole, Half and Peak Running Demands during Club and International Youth Rugby League Match-Play. *Science and Medicine in Football*, *3*(1), 63–69. https://doi.org/10.1080/24733938.2018.1480058

Zhang, J., O'Reilly, K. M., Perry, G. L. W., Taylor, G. A., & Dennis, T. E. (2015, April). Extending the Functionality of Behavioural Change-Point Analysis with k-Means Clustering: A Case Study with the Little Penguin (Eudyptula Minor). *PLOS ONE*, *10*(4), e0122811. https://doi.org/10.1371/journal.pone.0122811

Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W.-Y. (2008) Understanding Mobility Based on GPS Data [Online]. In: *Proceedings of the 10th international conference on Ubiquitous computing. New York, NY, USA:* Association for Computing Machinery, pp. 312–321. Available from: https://doi.org/10.1145/1409635.1409677 [Accessed 26 September 2020].