

The Calculation of Player Speed from Tracking Data

Lucas Y. Wu and Tim B. Swartz *

Abstract

This short communication considers the calculation of player speed from tracking data. Whereas there are many player tracking systems, all rely on the collection of Cartesian coordinates corresponding to the players on the pitch. From these Cartesian coordinates, there are many ways that one could approximate player speed and acceleration. We introduce some simple principles from exploratory data analysis which help yield more reliable speed calculations. The general principles are illustrated on various player tracking systems.

Keywords : big data, player tracking data, spatio-temporal data, soccer analytics, exploratory data analysis.

*L. Wu is a PhD candidate and T. Swartz is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Swartz has been partially supported by the Natural Sciences and Engineering Research Council of Canada. Swartz is also a member of the Canadian Statistical Sciences Institute (CANSSI) Collaborative Research Team in Sports Analytics which has also partially supported the research. The authors thank Daniel Stenz, former Technical Director of Shandong Luneng Taishan FC who provided the data used in this paper. The authors also thank the reviewer László Csató whose helpful comments improved the paper.

1 INTRODUCTION

In the past decade, the advent of player tracking data has sparked a revolution in sports analytics (Morgulev, Azar and Lidor 2018). With player tracking data, analysts have access to the Cartesian coordinates of each player on the pitch where the observations are recorded frequently (e.g. 10 times per second). The availability of such detailed data provides opportunities to investigate sporting questions that were previously unimaginable. Gudmundsson and Horton (2017) provide a review paper on spatio-temporal analyses used in invasion sports where player tracking data are available.

Currently, player tracking systems are expensive, and consequently, tracking data are only collected in “big” sports such as basketball (the National Basketball Association), soccer (various leagues and competitions), football (the National Football League) and hockey (the National Hockey League). Tracking data are not only collected during matches but also during workout sessions where fitness, training and health considerations are main concerns.

Tracking data are typically proprietary and are supplied by service providers using various technologies (Torres-Ronda et al. 2022). There are four prominent technologies: (1) global positioning systems (GPS), (2) local positioning systems (LPS), (3) inertial measurement units (IMU) and (4) optical tracking (OT) systems. OT systems are fundamentally different as they do not require wearable devices and do not directly determine player coordinates. Instead, OT technology requires advanced camera systems and player recognition software to evaluate player coordinates. No matter which technology is utilized, tracking systems begin with the collection of the (x, y) coordinates of participants measured at frequent time intervals. With the coordinates, various statistics can be calculated or approximated (e.g. speed, acceleration, distance travelled, etc.).

In this paper, we are concerned with derivative calculations associated with tracking data coordinates. Specifically, we are interested in the approximation of player speed which is an important statistic in sports analytics and sports science. For example, Wu and Swartz (2022) require player speeds in soccer to assess off-the-ball activity. They introduce a measure which addresses defensive anticipation. Buchheit et al. (2014) use regression methodology to determine factors that are associated with player speed in soccer. For example, horizontal force and horizontal power were seen to be associated with speed.

Oliva-Lozano et al. (2020) characterize positional differences in soccer based on acceleration and sprint profiles. Related to speed, Shen, Santo and Akande (2022) analyze pace of play in soccer, and conclude that pace increases with decreasing team quality, which indicates the importance of playing with pace. From a training and performance perspective, Ferrari Bravo et al. (2008) demonstrate that sprint-training significantly increases both aerobic and anaerobic performances in soccer. Naturally, different applications require different levels of accuracy. For example, in sports science, critical velocity is an active research field which relies on highly accurate measurements of speed (Peng, Clarke and Swartz 2022).

Much has been written on the accuracy of various tracking data technologies. For example, Mara et al. (2017) considered the displacement accuracy of an OT system, Tan, Polglaze and Peeling (2021) investigated the validity and accuracy of a GPS system, and Pino-Ortega et al. (2022) provided a review of the validity and reliability of LPS systems against other devices. Massard, Eggars and Lovell (2017) questioned the need for sprint testing based on the comparison of GPS match and field-testing data. However, all of these investigations rely on some measure of the truth against which tracking measurements are compared. What should experimenters do if they do not have access to the truth and they are unsure of the accuracy of speed calculations obtained from tracking data? This paper introduces some simple principles from exploratory data analysis that assists experimenters to obtain more reliable estimates of speed.

In Section 2, we describe the datasets upon which our methods are illustrated, and we describe how player speed is calculated from tracking data coordinates. In Section 3, some simple exploratory plots are introduced that help the analyst obtain more reliable speed calculations. We conclude with a short discussion in Section 4.

2 DATA

We have access to tracking data from matches during the 2019 season of the Chinese Super League (CSL). The CSL uses OT technology (previously discussed) provided by Stats Perform where observations were recorded 10 times per second. The tracking data consist of roughly one million rows per match measured on 7 variables. Each row corresponds

to a particular player at a given instant in time. The soccer tracking data were initially provided as xml files, and were processed in R for further analysis. In Table 1, we present three rows of the soccer tracking data. Here we observe x-y coordinates and player identifiers at every 1/10th of a second. The entries are mostly intuitive except perhaps for the x-y coordinates which refer to the player location on a 105m by 68m soccer field. For example, $(x, y) = (-52.5, 0)$ corresponds to the middle of the goal line on the left hand side of the soccer field.

gameID	Time	x	y	IdActor	IsBall	IdHalf	JerseyNumber
WUHAN-BEIJI-01032019	30	-4	-9.6	345354	FALSE	1	25
WUHAN-BEIJI-01032019	30.1	-4	-9.5	345354	FALSE	1	25
WUHAN-BEIJI-01032019	30.2	-4	-9.4	345354	FALSE	1	25

Table 1: A sample of soccer tracking data from the CSL.

Our second dataset corresponds to tracking data from the National Football League (NFL). Unlike the OT soccer data, the NFL data were based on GPS technology, but were also collected using 10 hertz sampling frames. The data were used in the 2019 Big Data Bowl competition and are publicly available at <https://github.com/nfl-football-ops/Big-Data-Bowl>. Here we use data corresponding to a single deep pass play by the wide receiver Brandin Cooks of the New England Patriots taken from a 7-second interval during the September 7/2017 match against the Kansas City Chiefs. In Table 2, we present three rows of the football tracking data. Here we observe a similar structure to the tracking data in soccer. The football tracking data include the x-y coordinates for players measured in yards where x refers to the player position along the long axis of the field ranging from 0 to 120 yards, and y refers to the player position along the short axis of the field ranging from 0 to 53.3 yards. For instance, $(x, y) = (0, 0)$ corresponds to the bottom left of the football field. The remaining variables in Table 2 are mostly intuitive where dis corresponds to distance travelled from the previous frame (i.e. previous 1/10th second) and dir corresponds to the angle of player motion in degrees. The $frame.id$ is the frame identifier for each frame which resets to 1 for each play.

gameId	playId	frame.id	x	y	dis	dir	event	playerId	displayName	jerseyNumber
2017090700	160	40	53.78	10.82	0.77	239.36	pass_forward	2543498	Brandin Cooks	14
2017090700	160	41	53.11	10.45	0.76	238.66	NA	2543498	Brandin Cooks	14
2017090700	160	42	52.44	10.08	0.76	237.76	NA	2543498	Brandin Cooks	14

Table 2: A sample of football tracking data from the NFL.

2.1 Speed Calculations

We emphasize that the approach that we introduce is general and straightforward. It can be utilized using any tracking technology in any sport. However, knowledge of the sport dictates our interpretation of the exploratory plots.

Consider then a particular player where our interest concerns the calculation of their speed. If $(x(t), y(t))$ denotes the location of the player at time t , then the player’s speed at time t is defined by

$$s(t) = \lim_{\Delta \rightarrow 0} \frac{\sqrt{(x(t + \Delta) - x(t - \Delta))^2 + (y(t + \Delta) - y(t - \Delta))^2}}{2\Delta}. \quad (1)$$

In words, formula (1) is the limiting change in distance travelled with respect to time. Of course, (1) is a mathematical expression based on taking a limit, and is not a quantity that can be calculated from data. Instead, with tracking data, the player’s locations are obtained at regular times which are denoted by $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Here, the subscripts $i = 1, \dots, n$ of the Cartesian coordinates refer to the time increments. Therefore, assuming that t corresponds to an observed time increment from the tracking data, it is reasonable to approximate $s(t)$ in (1) by

$$\hat{s}(t) = \frac{\sqrt{(x_{t+\Delta} - x_{t-\Delta})^2 + (y_{t+\Delta} - y_{t-\Delta})^2}}{2\Delta} \quad (2)$$

where $\Delta = 1, 2, \dots$ is an increment that needs to be specified. In our illustration with 10 hertz data, the value $\Delta = 1$ corresponds to 1/10th of a second.

We have simplified the discussion by referring to speed. The approximation of velocity is also of interest where velocity has a directional component in addition to the scalar quantity speed. Note that acceleration calculations are also important, and are obtained

as derivatives of speed.

3 EXPLORATORY ANALYSES

Whereas the estimand $s(t)$ in (1) is an instantaneous speed, its estimate $\hat{s}(t)$ in (2) is an average speed taken over the time period 2Δ . It may therefore appear that smaller values of Δ will yield better estimates. However, this needs to be balanced against the fact that player coordinates (x_t, y_t) are subject to measurement error as is the time interval 2Δ . Therefore, inaccuracies in the speed estimates are propagated from inaccuracies in the raw data.

To theoretically investigate the magnitude of error in speed via measurement error in the numerator of (2), we consider the true speed $\Delta_l/(2\Delta)$ which denotes the change in location Δ_l by the change in time where Δ denotes the previously defined incremental step size in time. With measurement error present, we denote the observed speed $(\Delta_l \pm E)/(2\Delta)$ where E denotes a fixed error in the location measurement corresponding to the device. Then absolute error AE is given by

$$\begin{aligned} \text{AE} &= \frac{|\Delta_l/(2\Delta) - (\Delta_l \pm E)/(2\Delta)|}{\Delta_l/(2\Delta)} \\ &= |E|/\Delta_l. \end{aligned} \tag{3}$$

We note that the absolute error (3) is smaller for larger speeds (i.e. greater changes in location Δ_l). For example, when $\Delta = 1$, consider a true location displacement $\Delta_l = 8$ metres which is incorrectly measured as 9 metres. Then the actual speed is 8.0 metres/sec (fast), the observed speed is 9.0 metres/sec, and the measurement error is $E = 1$ metre. This results in absolute error $AE = 0.125$. For contrast, when $\Delta = 1$, consider a true location displacement $\Delta_l = 2$ metres which is incorrectly measured as 3 metres. Then the actual speed is 2.0 metres/sec (slow), the observed speed is 3.0 metres/sec, and the measurement error is $E = 1$ metre. This results in absolute error $AE = 0.50$.

3.1 Soccer Example

To begin our investigation, Figure 1 provides a plot of the locations of a player from the CSL dataset taken during a 29-second interval where he is known to be running fast during portions of the interval. When a player is running fast, it is physically impossible to make sharp turns, and therefore, the smoothness of the path suggests apparent accuracy in the location measurements.

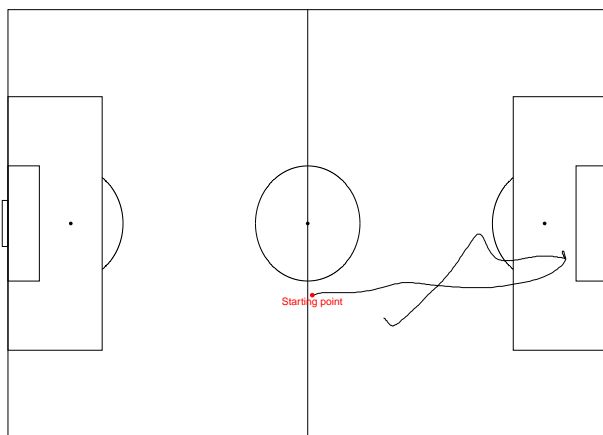


Figure 1: Path of a player over a 29-second interval based on location data recorded at 10 hertz.

However, when we take the path locations in Figure 1, and estimate speeds (2) using $\Delta = 1$, there seems to be a significant accuracy problem. Figure 2 provides a plot of estimated speed versus time for the selected path. In Figure 2, we observe that there are many instances where a player has a recorded speed which increases (or decreases) by roughly 1.0 metre per second in the subsequent 1/10th second, and then returns to the baseline speed 1/10th of a second later. When speeds are recorded in the (0,8) metres per second range, frequent fluctuations of this magnitude do not seem plausible. The problem here is that the location measurements were recorded to one decimal point on the metres scale, and therefore, there is inaccuracy in (2) when dividing by 2Δ which corresponds to 0.2 seconds.

A remedy to the estimation of the instantaneous speed $s(t)$ is to increase the time

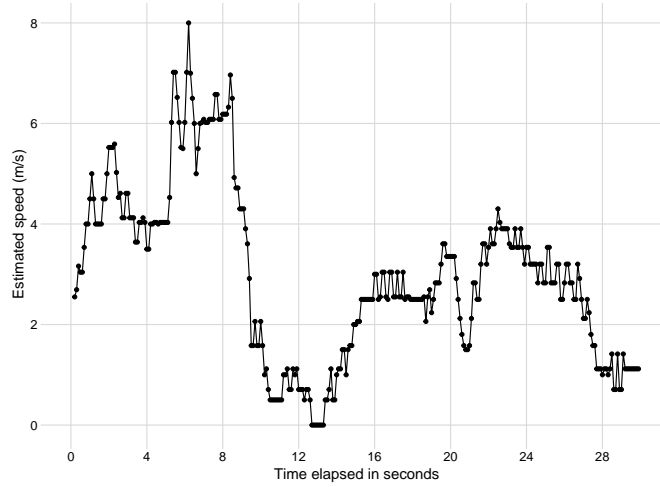


Figure 2: Estimated speed ($\Delta = 1$) of the player corresponding to the path in Figure 1 over a 29-second interval.

increment Δ surrounding t . Increasing the length of the time interval 2Δ results in less fluctuation in the estimated speeds which is desirable. However, this is done at the expense of moving in the direction from instantaneous speeds to average speeds. We have found that the approximation $\Delta = 4$ works well in this application. Figure 3 provides the analogous plot to Figure 2 where the time intervals have been widened to intervals of length 0.8 seconds. In Figure 3, we observe that the fluctuations are less pronounced, and that the plot of estimated speed versus time is smoother. For example, the fluctuations during the interval 16-18 seconds in Figure 2 are less believable than what is observed in Figure 3.

We refer back to the theoretical analysis of absolute error at the beginning of Section 3. In this example, we have seen that we prefer the time increment $\Delta = 4$ over $\Delta = 1$. With $\Delta = 4$, speed $\Delta_l/2\Delta = 8$ metres/sec and location measurement error $E = 1$ metre, this implies $\Delta_l = 64$ metres and absolute error $AE = E/64 = 0.015625$. With $\Delta = 1$, speed $\Delta_l/2\Delta = 8$ metres/sec and location measurement error $E = 1$ metre, this implies $\Delta_l = 16$ metres and absolute error $AE = E/16 = 0.0625$. Therefore, $\Delta = 4$ is preferred over $\Delta = 1$ in reducing absolute error. This exercise can be repeated for any speed.

Issues which arise in speed measurements are a consequence of the fact that speed

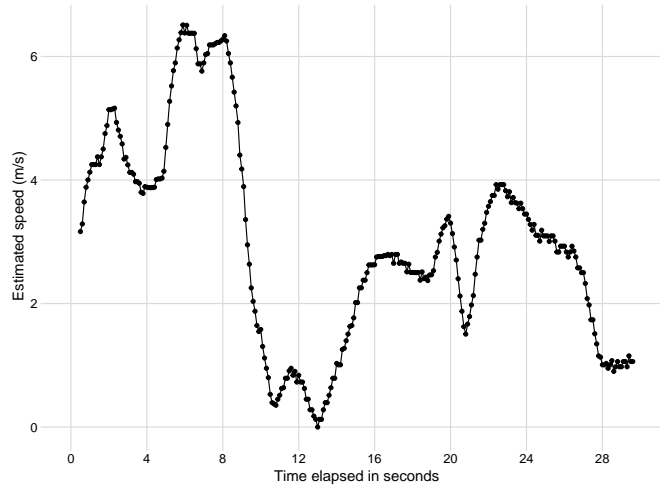


Figure 3: Estimated speed ($\Delta = 4$) of the player corresponding to the path in Figure 1 over a 29-second interval.

is the derivative of position, and that position is not measured with sufficient accuracy. In applications where acceleration measurements are important, one can imagine even greater challenges since acceleration is the derivative of speed. This is illustrated in the following example.

3.2 NFL Football Example

In the second example, we first note that the running patterns of a NFL wide receiver differ from those of a soccer player. Typically, the wide receiver sprints over a short time interval and does not make many changes of direction. This has implications for the estimation of speed.

In Figure 4, we provide the estimated speed and acceleration estimates for Brandin Cooks based on a 7-second pass route. The red-lined plots correspond to estimates based on $\Delta = 1$ (i.e. intervals of 0.2 seconds), and the blue-lined plots correspond to estimates based on $\Delta = 2$ (i.e. intervals of 0.4 seconds). Using $\Delta = 1$, the speed estimates appear satisfactory as there are no unrealistic fluctuations between successive estimates. When we compare the speed estimates using $\Delta = 1$ to $\Delta = 2$, there is no apparent improvement in the speed estimates. This suggests that $\Delta = 1$ may be adequate for this application which

is a different conclusion than with the soccer data. This may point to either a difference between the OT technology versus the GPS technology, or the intrinsic differences between the motions of soccer players compared to wide receivers in football.

When we look at the acceleration plots in Figure 4, it appears that $\Delta = 1$ may exhibit untenable fluctuations in acceleration, especially around the 5.5 second mark. For example, from the 5.2-second to 5.9-second mark, there is a change in acceleration in each successive time step, and the acceleration follows an unlikely fluctuating pattern of up, down, up, down, down, up and down (i.e. five changes in direction). From the 5.2-second to 5.9-second mark with $\Delta = 2$, we observe the more believable pattern of up, up, down, same, same, down and same (i.e. only one change in direction). With respect to the estimation of acceleration, $\Delta = 2$ is preferred over $\Delta = 1$.

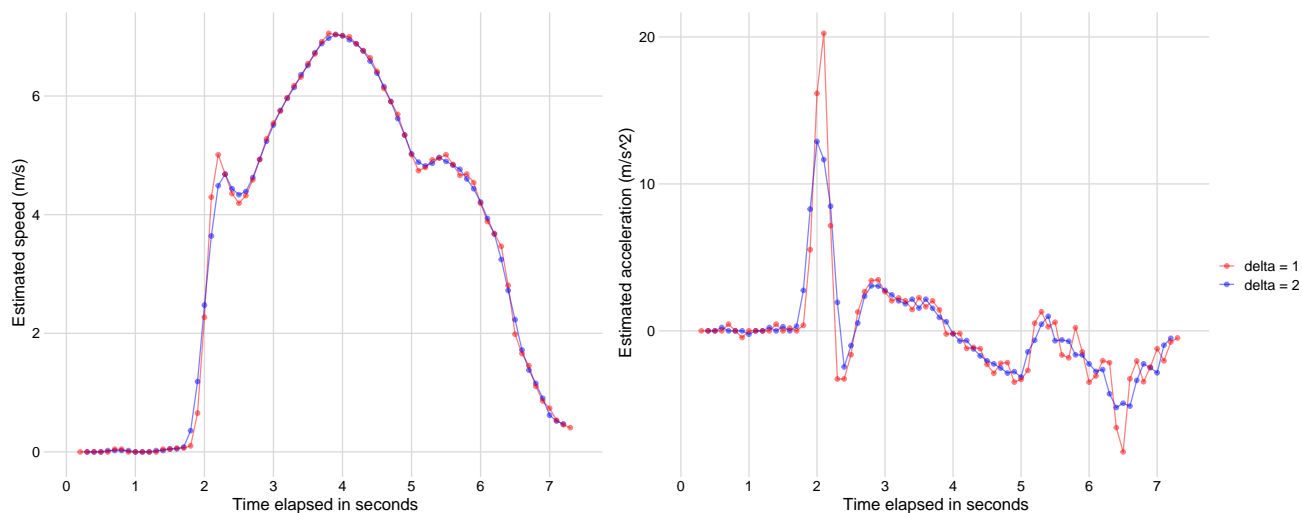


Figure 4: The red-lined plots correspond to speed and acceleration estimates ($\Delta = 1$) for Brandin Cooks of the NFL during a 7-second time interval. The analogous blue-lined plots correspond to $\Delta = 2$.

4 DISCUSSION

Tracking data have provided opportunities to study problems in sports analytics which were once unimaginable. However, sound tracking data analyses require data that are reli-

able, and the reliability of tracking data statistics often degrade with increasingly complex statistics. We have provided some simple principles from exploratory data analysis to help experimenters derive more reliable estimates of player speed. The same principles can be utilized in the calculation of velocities and accelerations.

The principles developed here are general and can be used with any type of player tracking system in any sport. The experimenter needs to consider the estimands of interest. The experimenter also requires domain knowledge of the sport to assess whether the resultant variations in the estimates are reasonable.

An avenue of future research may involve the implementation of statistical methods to smooth estimates of speed and acceleration. For example, one might consider the Hodrick-Prescott filter to smooth estimates of speed (Hodrick and Prescott 1997).

Instead of having experimenters manually estimate speed from (x, y) coordinates, some tracking data providers automatically provide speed statistics. Coleman (2018) describes the procedure that the data provider Opta uses in calculating top speeds for players in soccer: “The speed in kilometers per hour for a given frame is based on the previous 15 frame-to-frame speeds. Out of the 15 frame-to-frame speeds, the four highest and the four lowest values are discarded and the result is an average of the remaining seven values.” Given that speed is of great importance in sports analytics, we suggest that it would be good practice for the providers to be explicit about the the derivation and justification of their speed calculations.

5 REFERENCES

- Buchheit, M., Samozino, P., Glynn, J.A., Michael, B.S., Al Haddad, H., Mendez-Villanueva, A. and Morin, J.B. (2014). Mechanical determinants of acceleration and maximal sprinting speed in highly trained young soccer players. *Journal of Sports Sciences*, 32(20), 1906-1913.
- Coleman, J. (2018). Revealed: The 20 fastest players in the Premier League, including Kyle Walker and Raheem Sterling. *talkSPORT Football*, accessed August 7, 2022 at <https://talksport.com/football/348058/20-fastest-players-premier-league-walker/>
- Ferrari Bravo, D., Impellizzeri, F.M., Rampinini, E., Castagna, C., Bishop, D. and Wisloff, U.

- (2008). Sprint vs. interval training in football. *International Journal of Sports Medicine*, 29(8), 668-674.
- Gudmundsson, J. and Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, 50(2), Article 22.
- Hodrick, R. and Prescott, E.C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1), 1-16.
- Mara, J., Morgan, S., Pumpa, K. and Thompson, K.G. (2017). The accuracy and reliability of a new optical player tracking system for measuring displacement of soccer players. *International Journal of Computer Science in Sport*, 16(3), 175-184.
- Massard, T., Eggers, T. and Lovell, R. (2017). Peak speed determination in football: Is sprint testing necessary? *Science and Medicine in Football*, 2(2), 1-4.
- Morgulev, E., Azar, O.H. and Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5(4), 213-222.
- Oliva-Lozano, J.M., Fortes, V., Krustap, P. and Muyor J.M. (2020). Acceleration and sprint profiles of professional male football players in relation to playing position. *PLOS ONE*, 15(8): e0236959. <https://doi.org/10.1371/journal.pone.0236959>
- Peng, K., Clarke, D.C. and Swartz, T.B. (2022). Bayesian approaches for critical velocity modelling of data from intermittent efforts. *International Journal of Sports Science and Coaching*, 17(4), 868-879.
- Pino-Ortega, J., Oliva-Lozano, J.M., Gantois, P., Nakamura, F.Y. and Rico-González, M. (2022). Comparison of the validity and reliability of local positioning systems against other tracking technologies in team sport: A systematic review. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 236(2), 73-82.
- Shen, E., Santo, S. and Akande, O. (2022). Analyzing pace-of-play in soccer using spatio-temporal event data. *Journal of Sports Analytics*, 8(2), 127-139.

- Tan, J.H.Y., Polglaze, T. and Peeling, P. (2021). Validity and reliability of a player-tracking device to identify movement orientation in team sports. *International Journal of Performance Analysis in Sports*, 21(5), 790-803.
- Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A. and Clubb, J. (2022). Tracking systems in team sports: A narrative review of applications of the data and sport specific analyses. *Sport Medicine Open*, 8, Article 15.
- Wu, L. and Swartz, T.B. (2022). Evaluation of off-the-ball actions in soccer. Under review, accessed August 6, 2022 at <https://www.sfu.ca/tswartz/>