Determining spatio-temporal metrics that distinguish play outcomes in field hockey

MCINERNEY, Ciaran <http://orcid.org/0000-0001-7620-7110>

Determining spatio-temporal metrics that distinguish play outcomes in field hockey

Ciarán McInerney

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of Doctor of Philosophy

April 2017

Collaborating organisation: England Hockey

Abstract

Tactical behaviour in field sports can be examined using spatio-temporal metrics, which are descriptions of player behaviour derived from data of player positions over time. Many metrics can be computed that describe the cooperative and adversarial interactions between players. The methods typically used by sports performance analysts cannot appropriately analyse the many possible spatio-temporal metrics and their interactions. Tantalisingly, the interactions between these descriptions of player behaviour could potentially describe tactical differences in performance.

This thesis describes a programme of research that determined some spatio-temporal metrics that distinguish play outcomes in field hockey. Methods inspired by genetic analysts were used to estimate the influence of combinations of spatio-temporal metrics on the outcome of field hockey plays. The novel application of the genetic methods to sports performance data raised some practical difficulties. Adjustments to the method facilitated the selection of distinguishing metric combinations from an initially large list of over 3,600 metrics.

The adjustments made to the genetic methods represent one of several contributions to knowledge made by this programme of research. These contributions will help performance analysts with the increasingly common task of analysing high-dimensional data. Other contributions to knowledge are a suite of metric combinations that distinguish play outcomes in field hockey and empirical support for some tactical preconceptions.

The key finding of interest for players and coaches is that play outcomes in field hockey are distinguished by proximity to the goal and passing execution. The metrics that distinguish the several outcomes differ depending on the outcomes being compared. Coaches and athletes should therefore recognise the variety of tactics required to minimise negative outcomes and maximise positive ones.

Keywords: Backward Dropping Algorithm, Genomic Selection, Tactics, Team sports

Acknowledgments

I would like to thank the suite of supervisors that have helped me through the PhD programme, and everyone else at CSER. This study would also not have been possible if not for the funding provided by the Environmental and Physical Sciences Research Council, and the support and opportunity provided by England Hockey. A special thanks to everyone in Chestnut Court, too. Finally, I wouldn't be here if it weren't for the immeasurable support of my family, friends and loved ones.

Data availability statement

All data relevant to this thesis, including MATLAB scripts, are available, under embargo, at http://doi.org/10.17032/shu-170005.

Table of Contents

Chapter 1 Introduction

The thesis herein explains and justifies a method to distinguish outcomes of plays in field hockey using information about the location of players. This first chapter explains the motivation for the research before concluding with the aim of the thesis.

1.1 Motivation for research

Field hockey is a fast paced field sport with worldwide popularity, boasting 88 national Men's teams and 72 national Women's teams (FIH 2016). Although evidence of hockey-like games is found in many cultures worldwide, the modern game is said to have its origins in medieval Western Europe (Crego 2003). It has been an event at the Summer Olympic Games since 1908 for Men and 1980 for Women (International Olympic Committee 2011).The purpose of the game is to score in the opposition's goal at the far end of the pitch (*A* in Figure 1.1) by controlling the ball with a stick.

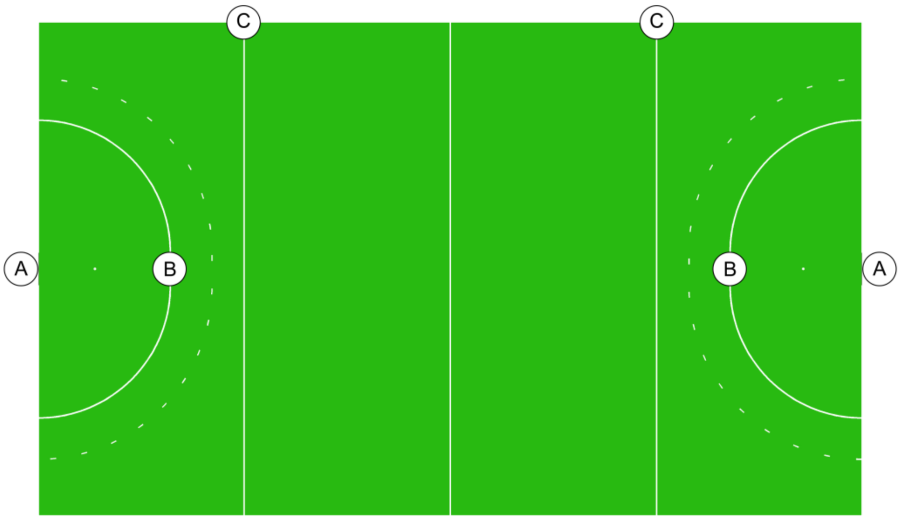


Figure 1.1 A field hockey pitch. *A*: the goals. *B*: the Circle line. *C*: the 23 m line.

The goals at the end of the pitch lie within an approximate semi-circle, known as the Circle (line *B* in Figure 1.1). The rules of field hockey dictate that for a scoring shot to be declared a goal, the final touch of the ball must have been within the Circle (FIH 2015). The Circle is therefore an important pitch marking when considering the outcomes of plays. Another important pitch marking is the 23 m line, which encloses the Circle in its quarter of the pitch (line *C* in Figure 1.1). There are different rules associated with this end region of the pitch defined by the 23 m line. In particular, fouls and out-of-bounds events can result in play being restarted from the 23 m line (FIH 2015) and penalty corners being awarded.

Both teams are constrained by these rules as they attempt to score and stop the opposition scoring. Teams' pre-game strategy and in-game tactics dictate their behaviour to this end. The current work's collaborators, England Hockey and Great Britain Hockey, are interested in these tactical behaviours that are associated with good and bad outcomes.

The most notable investigations into the tactics of field hockey has been led by Elferink-Gemser and colleagues (Elferink-Gemser et al. 2011; Elferink-Gemser et al. 2007; Elferink-Gemser 2005). Elferink-Gemser et al. (2004) developed the Tactical Skills Inventory for Sports. The inventory was informed by discussions with national and district level coaches and validated by elite and sub-elite players. The inventory addresses cognitive evaluations of gameplay scenarios using a self-assessed questionnaire. In contrast, work by Sunderland et al. (2006) investigated the spatial distribution of gameplay actions to infer tactical tendencies of elite players. These two paradigms exemplify the two methods commonly used to investigate tactics in team field sports. The first takes advantage of knowledge from domain experts under the assumption that athletes and coaches are performance experts in their domain. The second measures player behaviour on the pitch.

Domain experts have been used to create performance metrics (James et al. 2005), to weight the importance of metrics (Hraste et al. 2008; Bremner et al. 2013), and to evaluate performance models (Foretić et al. 2013). It can be difficult to get access to domain experts because of their busy performance schedules, and access to the opposition's domain experts is unlikely for competitive reasons, which limits and potentially biases any findings. Initial attempts to engage the current work's collaborators failed for these very reasons.

The alternative is to investigate players' behaviour non-invasively and infer tactical intentions and methods (Sampaio and Maçãs, 2012). The rationale is that the cooperative and adversarial interactions between players are the consequence of tactical intentions (Gudmundsson and Horton, 2016). This can be done by obtaining information about player locations over time, which has only been practically possible with recent advances in technology (Kelley et al. 2016). The number of possible metrics that makes use of this data rises sharply when teams, player groups and events are specified (Memmert et al. 2017). Different analyses also provide new metrics that add to the growing number of ways to describe the spatio-temporal behaviour of players (Duarte et al. 2013; Moura et al. 2013; Correia et al. 2013; Bourbousson et al. 2010b; Bourbousson et al. 2010a). There are so many spatio-temporal metrics that teasing out the important ones can be difficult. It would be useful for coaches and athletes to know which metrics sufficiently distinguish better and worse performances.

1.2 Conclusion

The aim of the current work was, therefore, to determine which spatio-temporal metrics distinguish play outcomes in field hockey. A review of the relevant literature was required and is detailed in the next chapter. The objectives of the review were:

1. To determine the appropriate method to measure player location.
2. To compile a list of appropriate spatio-temporal metrics.
3. To determine the appropriate method to analyse spatio-temporal metrics.
4. To determine the spatio-temporal metrics that distinguish play outcomes.

Chapter 2 Literature review

2.1 Introduction

The aim of the current work is to determine the spatio-temporal metrics that distinguish play outcomes in field hockey. Before tackling this aim, several questions need answering:

1. What are tactics?
2. What is 'a play'?
3. What is the most appropriate method to measure player location?
4. What spatio-temporal metrics are typically used?
5. What is the most appropriate method to analyse the metrics?

The sections of the chapter address these questions before concluding with an informed set of updated objectives for achieving the work's aim.

2.2 Tactics

There are many definitions of tactics in relation to team sports but common themes exist. Firstly, tactics are defined in the context of strategy and should involve the consideration of coaches and athletes (Cordes et al. 2012). Gréhaigne et al.'s (1999) frequently cited work on strategy and tactics is summarised by Turner et al. (2001, p. 42):

> "*Strategy is concerned with issues in advance of game play without its time constraints...Tactics, on the other hand, require immediate adaptation to the opposing team's players as well as team-mates, and operate under the constraints of time*"

From this, we gather that tactics pertain to in-the-moment decisions about player behaviour. Gréhaigne et al.'s (1999) definition also exemplifies another common theme in definitions of tactics, that is to say, that tactics are a "*reaction* *to an adversary in a game situation*" (Gréhaigne and Godbout, 1995). It is therefore important to consider offensive and defensive perspectives when investigating tactics (Vogelbein et al. 2014) because sports like field hockey require teams to score while simultaneously preventing the opponent from scoring (Lames 1991).

Additionally, tactics are often discussed with respect to player locations and time. Turner et al. (2001) focuses on the "*constraints of time*" but also cites the work of Gréhaigne et al. (1999), which references player positions and team formations. Earlier work of Gréhaigne and Godbout (1995) mentions player configurations and much of the research into tactics of team sports discuss players' spatial distributions (Moura et al. 2016) and uses metrics that describe the spatial behaviour of players (Memmert et al. 2017). Carling et al. (2005, p. 129) provides a summarising definition:

> "*Tactics are the individual and collective ways used by a team to try to best employ player skills to make the overall strategy work either by scoring or preventing goals*"

2.2.1 Tactics in field hockey

Despite these philosophical considerations of tactics, little has been done to assess tactical behaviour in team sports until recently (Memmert et al. 2017). Podgórski and Pawlak (2011) reviewed sports performance literature relating to field hockey from 1960 to 2010 and found only 36 articles investigating tactics. Studies that investigated tactics looked at "*improving players' training and effectiveness in matches*" and "*physical performance results obtained by field hockey players*". These criteria might therefore have covered notational analyses of player movements, measurements of playing intensity, investigations into group dynamics, technical execution of skills, and other biomechanics analyses. If Carling et al.'s (2005) definition were used, then tactics would likely be further underrepresented in field hockey.

Elferink-Gemser et al. (2004) developed the Tactical Skills Inventory for Sports with the support of elite field hockey coaches. The inventory addresses cognitive evaluations of gameplay scenarios using a self-assessed questionnaire. Elferink-Gemser et al. (2004) found that elite and non-elite players could be distinguished based on the four factors of the inventory: 'Positioning and Deciding', 'Knowing about Ball Actions', 'Knowing about Others', and 'Acting in Changing Situations'. Unfortunately, these insights were only obtained using questionnaires. The drawbacks of such methods are that they rely on participants’ self-beliefs, and do not lend themselves well to on-going tactical evaluation because they require players to take the time to complete the questionnaires. Furthermore, the opposition are unlikely to participate in the questionnaire so comparisons between opponent’s beliefs cannot be made.

In contrast, work by Sunderland et al. (2006) investigated the spatial distribution of gameplay actions to infer tactical tendencies of elite field hockey players. Sunderland et al. (2006) concluded that there was a preference for dribbled Circle entries from the right side and that possession was predominantly regained via free-hits. The advantage of discussing tactics in this way is that the resultant behaviour of players is analysed rather than their tactical intentions. Whilst tactical principles are important for guiding tactical behaviours (da Costa et al. 2009), spatio-temporal metrics provide a way to objectively measure behaviours that arise from their application. Spatio-temporal metrics might therefore be a more pragmatic tool to investigate field hockey tactics.

2.3 Defining 'a play'

To investigate tactical performance in team sports, it is necessary to define a unit of observation on which to focus. This unit of observation should describe tactical performance and define the boundary of interest that permits the work of researchers to be compared and contrasted. The term 'a play' might be a good candidate but is ambiguous in the performance analysis literature.

Generally speaking, 'play', 'gameplay', or 'match-play' describe the period that a sporting contest is underway. Some authors, however, will refer to 'a play' rather than 'play' without definition (Sainz De Baranda et al. 2008; Escalante et al. 2011; Ito et al. 2004). The term 'a play' often implies a specific sequence of purposefully linked movements and actions. This term is also synonymous with 'a possession', which often implies a portion of gameplay that begins when a team has control of the ball and ends when control is lost (James et al. 2004). To confuse matters further, some analysts mathematically define variables called 'Possession' to measure team quality (Kubatko et al. 2007; Crum 2013).

Turnovers are changes in possession, conceded by the team losing possession and earned by the team gaining possession. The concept must be considered to provide a complete definition of a play because it defines an end to possession. Although the term ‘Turnover’ is commonly used (J. Bradley and O’Donoghue, 2011; Michael David Hughes and Bartlett, 2002; Lupo et al., 2009), there are other terms like loss of possession (Amjad et al. 2013), dispossession (Spearritt 2013), ball recovery (Hewitt et al. 2016; Barreira, Julio Garganta, et al. 2013; Almeida et al. 2013) and (re)gain in/of possession (Bradley et al. 2013; Lupo, Condello, et al. 2012) that describe similar events. The choice of term depends on the perspective of the performer and turnover often implies a loss of possession (Carroll 2013; Lupo et al. 2011). Turnovers have the potential to define the start and end of possession and are therefore important for defining a play.

Field hockey has some unique gameplay events that influence possession, e.g. Free Hits, Penalty Corners and restarts following an out-of-bounds event caused by a defender within the 23 m region. Free Hits are sometimes awarded as compensation for a foul (FIH, 2015, Section 12.2). Play is paused, the ball must be stationary before play is restarted and opponents cannot interfere within 5 m of the restart location (FIH, 2015, Section 13.2). Free Hits affect possession when the offensive team commit the foul because the defensive team are awarded the Free Hit. Gameplay is also affected when the offensive team are awarded the Free Hit because the ball must be stationary and the defence must accommodate the 5 m rule into their tactical behaviour. The restarting player may pass to herself or himself to restart play and the ball need only be stationary "briefly" (FIH, 2015, p 7). The duration for which gameplay is paused varies depending on the tactical intentions of the restarting player. A player might continue play immediately with little effect on the trajectories and intentions of player. Alternatively, a player might pause and allow for more favourable conditions to arise before restarting play. Defining a play that is suitable for field hockey must incorporate the potential for Free Hits to allow significant changes in player arrangements and tactical intentions.

Penalty Corners are a method of restarting play from the backline that are only awarded for fouls committed within the 23 m region (FIH, 2015, Section 12.3) (Figure 2.1).

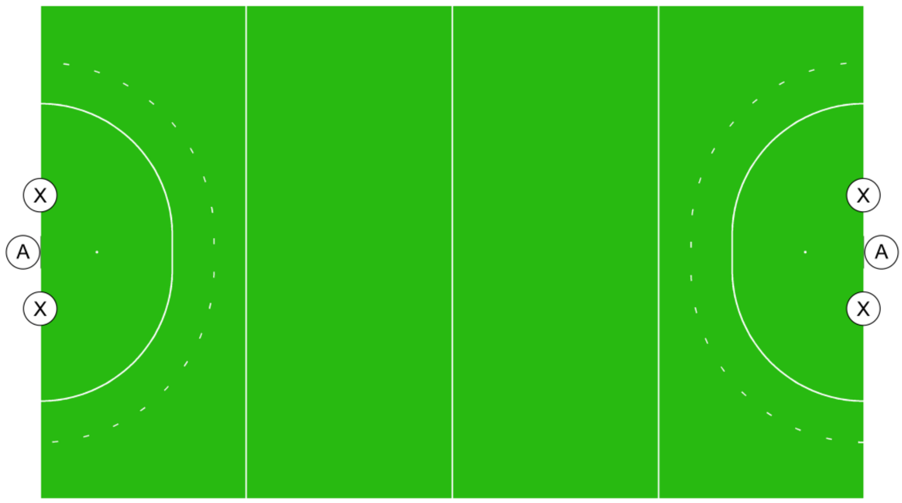


Figure 2.1 A field hockey pitch. *A*: the goals. *X*: the locations from which a penalty corner can be taken.

Penalty Corner restarts in field hockey have parallels with the corner kick of association football. That is to say, teams adopt specialised strategies and spatial arrangements to take advantage of the situation (Casal et al. 2015). Unlike association football, the Penalty Corner constrains the defensive team by only allowing four outfield players to be within their half of the pitch until the ball is played (FIH, 2015, Section 13.3). The rules also forbid the offensive team from entering the Circle until the ball is played and require the ball to be played out of the Circle before it can be returned for a shot. These rules create a gameplay scenario for which teams adopt distinct strategies.

Laird and Sutherland (2003) found that Penalty Corners usually involve few touches or are an immediate shot at goal. Although less common, the ball is sometimes returned to normal gameplay as the defensive team return to full strength within the 23 m region. Studies have yet to investigate the transition of perceived offensive advantage in scenarios where the first shot is not successful. Defining a play needs to incorporate this undefined transition between this set-piece and open gameplay for it to be useful for field hockey analysts.

The importance of the 23 m line must also be considered when defining a play in field hockey. As presented earlier, the rules governing fouls, and thus the pausing and restarting of gameplay, change when the offensive team progress into the 23 m region. If the ball unintentionally travels over the backline of the pitch from a defender, then the game is restarted by the offensive team on the 23 m line (FIH, 2015, Section 7.4b). The rules governing this restart are the same as those for a Free Hit, for example, defenders cannot interfere within 5 m of the restart location. The umpire formally restarts gameplay once they are satisfied with the restarting location. The effect of these rules is that both teams are given time to rearrange themselves. Any offensive advantage from a long pass or a breakaway player might be nullified as the defence set up between the offence and the goal. The rearrangement associated with a 23 m restart defines a scenario distinct from a 23 m intrusion from open play. Any definition of a play must handle the different scenarios associated with these two origins of a 23 m intrusion.

2.3.1 Theoretical and operational definitions

The preceding discussion informs a theoretical definition of a play (Table 2.1). Associating a play with "*purposefully linked movements and actions*" reflects

Carling et al.'s (2005) definition of tactics that mentions of players' individual and collective ways of achieving their overall strategy (Section 2.2).

Table 2.1 Operational definitions of *gameplay*, *possession*, and a *play*.

Term	**Definition**
Gameplay	The period that a sporting contest is underway. It is the duration of a game less stoppage time.
Possession	The portion of gameplay that begins when a team has control of the ball and ends when control is lost.
A play	A portion of a possession defined by a sequence of purposefully linked movements and actions.

This theoretical definition still requires an associated operational definition for performance analysts to apply it appropriately. This section discusses some frameworks that could help translate the definition provided in Table 2.1 into a practical and applicable definition for performance analysts. Hughes and Franks (2005) examined styles of play by defining possessions by the number of completed passes in a passing sequence before a turnover. It introduces the idea of defining plays based on notational data and addresses some aspects of the theoretical definition, i.e. connection with possession and based on a sequence of actions. Time is not factored into the framework nor is any other gameplay action, which limits investigations to the effects of passing sequences of different length.

James et al. (2004) concluded that the duration of possessions is an important factor for distinguishing game outcome and explicitly defined the start and end of possession using gameplay actions. This framework (Pollard and Reep, 1997b) is very practical and allows performance analysts to define a play by identifying appropriate gameplay actions. The identifiable action for the start of a possession was "*a player [gaining] sufficient control over the ball to effect [sic] a deliberate influence on its subsequent path*" and there was a choice of ending actions. Although Pollard and Reep's (2013) framework is practical and definitive, it lacks objectivity - what is "*sufficient*" control? An alternative to defining "*sufficient*" control is to use indicators with more objectivity. Ito et al. (2004) defined four football goalkeeper plays based on the gameplay actions that were involved. The kicks, dribbles and catches used by Ito et al. (2004) are easily identifiable. Determining easily identifiable gameplay actions might therefore be useful for defining a play. Based on the discussions so far, Figure 2.2 shows a suggested framework for an operational definition of a field hockey play, from the perspective of the offensive team.

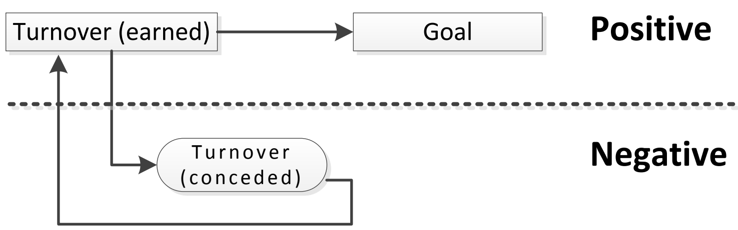


Figure 2.2 Schematic of a suggested framework to describe a play.

Presenting plays in a directed graph like Figure 2.2 encourages a reductionist but interaction-based approach to defining gameplay performance. Using a similar design, Barreira, Garganta, Guimaraes, et al. (2013) used their 'organizational model of soccer' as a foundation for investigations into ball recovery patterns. Barreira, Garganta, Guimaraes, et al.'s model puts a ball recovery event in the bigger picture of offensive gameplay for readers to understand its relevance. Similarly, Figure 2.2 succinctly shows the sequences between gameplay events in a way that captures the variety of possible plays:

1. Turnover (earned) - Goal
2. Turnover (earned) - Turnover (conceded)
3. Turnover (conceded) - Turnover (earned)

The framework in Figure 2.2 is basic and provides only one positive outcome per starting event. Using a goal as the only positive outcome is not advised because field hockey is a low scoring game (Van Calster et al. 2008). Pollard and Reep (2013) included multiple outcomes in their definition of a play because of the extreme inequality between plays that end in goals and those that do not. Such inequality between outcomes is not desired because it provides inadequate information for comparison (Atkinson and Nevill, 2001). Pollard and Reep (2013) also considered shots as an outcome event because they precede goals but noted that they are also very infrequent. A shot event would also require further definition, being described as "successful", "on-target", "on-goal", "at goal" and "led to goal" in the literature (Abraldes et al. 2012; Abreu et al. 2012; Alcaraz et al. 2011; Collet 2013; Vila et al. 2011). In most cases, insufficient definitions are provided and the reader is left to wonder at the difference between them. Furthermore, defining plays based only on gameplay events is not ideal because of possible ambiguity about when an event happened during a performance. This is especially relevant given James et al.'s (2004) conclusion that the duration of possessions is an important factor for distinguishing game outcome. An alternative to only using gameplay actions is to use spatial distinctions.

The frameworks discussed thus far have been from the offence's perspective but it has already been identified that both teams must be considered when investigating tactical performance (Section 2.2). Suzuki and Nishijima (2004, 2005) proposed a framework that defined defensive plays in football that made use of player locations. The framework has three phases:

1. Delaying attack

2. Forcing play in one direction

3. Squeezing the working space of attackers

The three phases are defined by subjective identification of player behaviour and are the basis for the Soccer Defending Skills Scale (Suzuki and Nishijima, 2004, 2005, 2007). Identifying the moment at which play is forced in a direction might differ between observers, even if consensus could eventually be reached. A useful characteristic of the Soccer Defending Skills Scale is that it scores performance based on the spatio-temporal metrics of distances, angles and tallies of players in a region. For example, the pitch-width spread of players is a metric used to score the final phase called 'Squeezing the working space of attackers'. Some value of width could objectively define the beginning of this phase.

Depending on the sport, there might be other meaningful spatial measurements that could objectively indicate key moments of a play. For example, section 1.1 introduced and explained the pitch markings on a field hockey pitch. The 23 m line and the Circle line mark regions of the pitch that are meaningfully different for performance (Figure 1.1). Briefly, some rules differ once players are within the region defined by the 23 m line, and only scoring shots from within the Circle can be goals. Crossing the 23 m line also means that the offensive teams are within the attacking quarter of the pitch. Moving gameplay into these regions represents distinct and meaningful progressions of an offensive play. Sunderland et al. (2005) took advantage of this and segmented a play into three phases using the Circle as a defining location: repossession of the ball, passing into the Circle and play within the Circle. This provides a field hockey example of how objective measurements of player behaviour can be used to define key moments in a play.

Figure 2.3 suggests framework for an operational definition of a field hockey play based on the discussion so far from the perspective of the offensive team. The framework is based on identifiable gameplay actions and objective indications of player locations that are meaningful and relevant to field hockey. It is an improvement on the framework suggested in Figure 2.2 in part because it has a higher resolution of gameplay events.

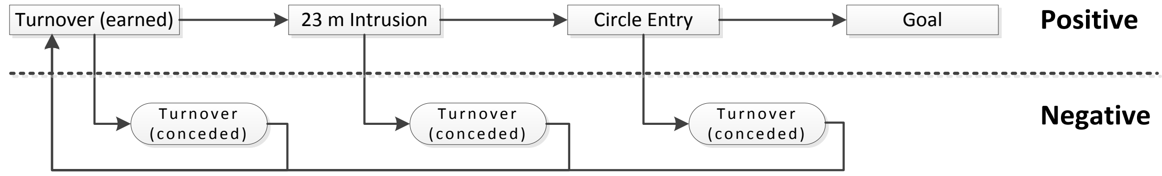


Figure 2.3 Schematic of a suggested framework to describe a play, based on easily identifiable gameplay actions and objective indications of player locations.

The most positive complete sequence starts with gaining possession outside the 23 m region ('Turnover (earned)'), followed by an intrusion into the opponent's 23 m region ('23 m Intrusion'), followed by an entry into the opponent's scoring area ('Circle Entry'), and finishes with 'Goal'. These positive events are indicated by rectangles in Figure 2.3. The '23 m Intrusion' and 'Circle Entry' events can be defined objectively by the moment the ball crosses the 23 m line and Circle line, respectively. Additionally, a ’23 m Intrusion’ includes gains of possession within the 23 m region, for example, by intercepting a pass or stealing the ball.

The framework in Figure 2.3 is more complicated than the framework in Figure 2.2 but still does not consider the variety of possible outcomes at each phase of a play. These other outcomes include the gameplay actions alluded to earlier like Free Hits, Penalty Corners and restarts of play after the ball crosses the backline from a defender. These gameplay events indicate a disruption in play but are not necessarily negative outcomes. These other outcomes are partially positive outcomes (Figure 2.4). Diedrick and van Rooyen (2011) used a similar paradigm when they divided rugby line-breaks into those that resulted in a try (positive), those where possession was lost (negative), and those where there was no try but possession was maintained (partially positive).

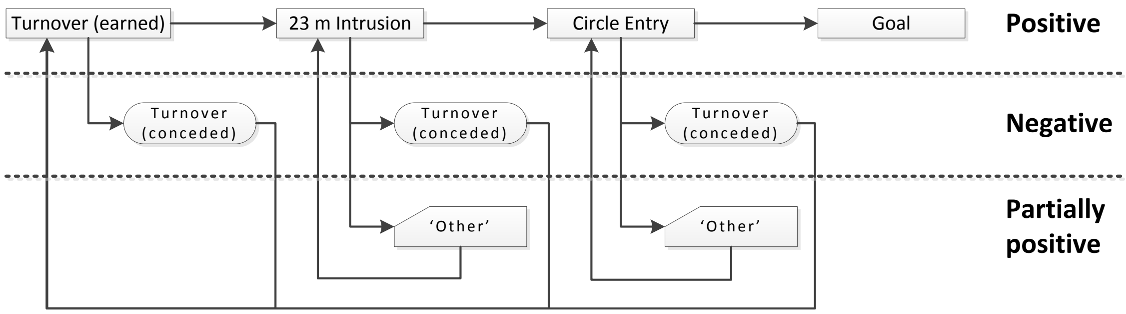


Figure 2.4 Schematics of a suggested framework to describe a play, which includes positive, negative and partially positive outcomes.

Figure 2.4 suggests a schematic that could operationally define both a possession and a play, as per the theoretical definitions in Table 2.1. Under this framework, possessions start with a 'Turnover (earned)' and continue until a 'Turnover (conceded)' or 'Goal' happen. Possessions can therefore be of four different lengths that distinguish the progression of the offensive team:

1. Turnover (earned) - Turnover (conceded)
2. Turnover (earned) - 23 m Intrusion - Turnover (conceded)
3. Turnover (earned) - 23 m Intrusion - Circle Entry - Turnover (conceded)
4. Turnover (earned) - 23 m Intrusion - Circle Entry - Goal

Within these possessions, a number of plays can be defined by following the paths between the nodes of the graph. These plays can be defined by starting at any node in the schematic and following the arrows until a 'Turnover (conceded)' or 'Goal' are reached. The most desirable play from the offence's perspective is 'Turnover (earned)' - '23 m Intrusion' - 'Circle Entry' - 'Goal' but a real-life play might involve some cycles in and out of the 23 m region.

Although derived from the behaviour of the offensive team, the framework allows defensive plays to be defined, which have had little attention in the literature to date (Wheeler et al. 2013). For example, analysts can investigate the behaviour of the defensive team during the portion of gameplay between an attacker intruding into the 23 m region and entering the Circle. This definition of a play could be represented by the second and third phases of Suzuki and Nishijima's (2004, 2005, 2007) defensive framework but with improved objectivity of the start and end events.

2.3.2 Conclusion

A play represents a unit of tactical investigation on which to focus but it is not well-defined in the literature. A definition appropriate for field hockey must account for general considerations of tactical performance and specific considerations relating to the rules of the sport. The definition should also be operationally defined so that it can be unambiguously applied.

Discussion informed a theoretical definition of a play as it related to field hockey. The theoretical definition informed further discussion about frameworks that help operationally define a play based on the expected events of the game (Figure 2.4). Error associated with the subjective identification of the events can be reduced by defining them using objective measurements of player locations.

2.4 Measuring player location

Kelley, Higham and Wheat (2016) reviewed methods used to determine player locations and grouped them under four headings:

1. Electromagnetic tracking
2. Inertial sensors
3. Signal propagation sensing
4. Vision-based systems

Electromagnetic tracking systems are not appropriate for measuring player locations in field sports because they are limited range of 1.5 - 9 m. Inertial sensors are not accurate for location measurements because they are designed to measure accelerations and orientations. Errors are exacerbated by the data transformation procedures required to obtain location data.

The most popular signal propagation sensing system used in sports performance analysis is Global Positioning System (GPS). GPS uses time-of-flight measurements from multiple orbiting satellites to determine location by trilateration. GPS is not useful for large stadia or indoor sports because a line-of-sight is required from satellite to GPS unit (Abidin 2010). Although GPS demonstrates good accuracy and repeatability up to moderate speeds (Coutts and Duffield, 2010; MacLeod et al., 2009; Wu et al., 2007), it requires players to wear sensors. This makes GPS and any other sensor-based systems unusable in competition environments where sensors are banned. The requirement for sensors also makes sensor-based systems unusable if the locations of the opposition are of interest, unless data-sharing arrangements are made between opponents. Kelley, Higham and Wheat (2016) conclude that vision-based systems are currently the only option for non-invasive tracking of player location.

2.4.1 Vision-based systems

Vision-based systems track objects using one or more cameras. The location of the player in an image is found automatically or with user input and a mapping function translates the image coordinates to real-world coordinates. Some vision-based systems like Vicon® (Vicon 2016) and MotionAnalysis® (MotionAnalysis 2016) track reflective markers. The need for additional markers makes these systems unusable for the same reason that sensor-based systems are not usable.

Both single- and multi-camera systems exist. Multi-camera systems allow for the three-dimensional location of objects to be measured provided the object can be seen in two cameras' views at the same time (Choppin et al. 2007). Standard single-camera systems can only provide two-dimensional, planar measurements of location (Walton 1981; Mauthner et al. 2008; Dunn 2014). The relatively new introduction of depth cameras has made it possible to measure three-dimensional location with one camera, but cameras have limited range and the method is still in its infancy (Choppin and Wheat, 2013).

Single cameras have limited fields of view so multiple cameras can be used to cover a large area of interest (Bialkowski et al. 2013). Companies like ProZone® (ProZone 2016) and SportUV® (SportUV 2016) provide data on player locations and gameplay events using 8- and 6-camera systems, respectively. These systems require permanent installations in large or covered stadia. Field hockey stadia are neither large nor covered and have not attracted sufficient commercial attention for investment in these kinds of permanent installations. The solution for single-camera systems is to use a wide-angle lens (see Konarski (2010) for example in field hockey). These lenses increase the angle of view but require the image to be undistorted through post-processing.

2.4.2 Measuring player location

Once images of player performance are collected, the next concern is appropriately selecting the location of a player in an image. Obtaining location data from vision-based systems can be automated, semi-automated or manual. Automated and semi-automated methods use image-processing algorithms to track an object of interest but current vision-based systems are not reliable enough to provide fully automated tracking in multi-person sporting environments (Bialkowski et al., 2013). Semi-automated systems, like SAGIT (Perš et al. 2002), require a user to initialise tracking and supervise the tracking process.

Manual systems require the user to specify the location of the player in every frame of the video. Low-resolution estimates 'tag' players in regions of the pitch (Sunderland et al. 2005; Barreira, Júlio Garganta, et al. 2013; Clemente et al. 2013). Higher-resolution methods reconstruct player locations in metres based on coordinates of players in an image (Clemente et al. 2012; Ricardo Duarte et al. 2010). To do this, a distribution of known locations in the field of view are chosen and their corresponding locations in a video image are selected (Figure 2.5). The known locations define a plane upon which future points of interest lie. A planar Direct Linear Transformation (Abdel-Aziz and Karara, 1971; Walton, 1981) models a mapping function that reconstructs the selected image coordinates onto the real-world plane. Using this mapping function, any subsequent points from the image's (u, v)-coordinate system can be mapped to the pitch's (x, y)-coordinate system.

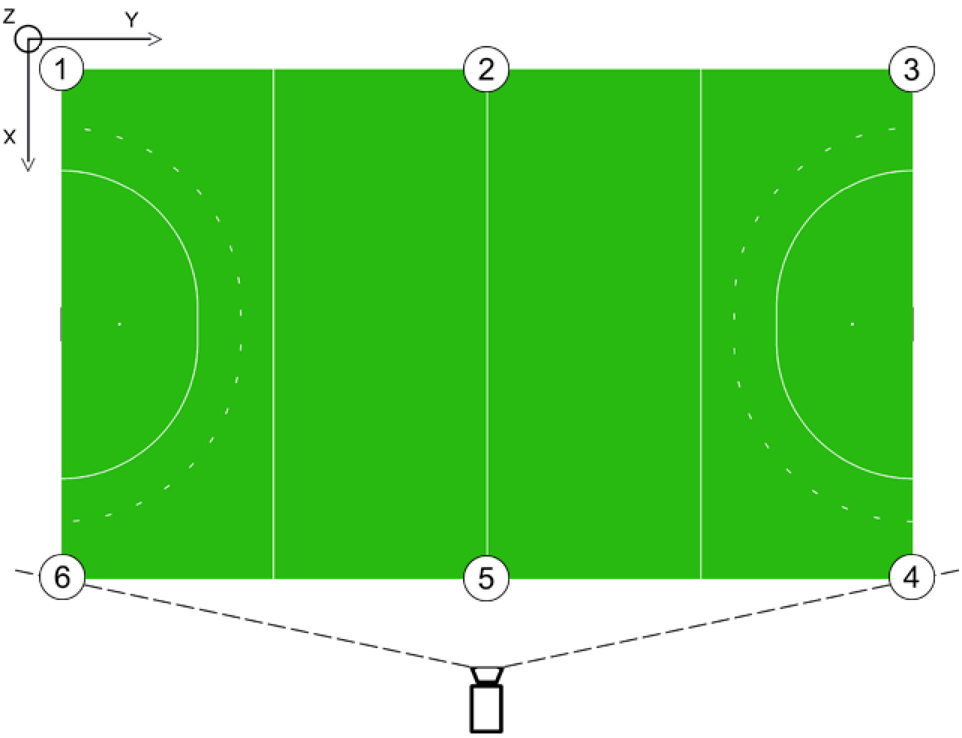


Image removed due to copyright

Figure 2.5 *Top*: Schematic of known locations in the pitch's (x, y)-coordinate system. *Bottom*: Selection of locations in the image's (u, v)-coordinate system

The points of interest are the locations of the players and ball. Many studies do not define the digitised point that is used to represent the player (Konarski 2010; Clemente et al. 2012; Folgado et al. 2012) but the gold standard for representing a person as a single point is the player's centre of mass (Sewell et al. 2013). Calculating a person's centre of mass is non-trivial and requires special equipment to measure the inertial parameters of body segments (Eames et al. 1999). Therefore, the centre of mass is typically estimated from the video image.

The issue that arises is that a player's centre of mass is above the pitch surface but player locations are considered as if they are on the pitch surface. A projection of the centre of mass to the pitch surface is therefore required. Unless the centre of mass is estimated from a bird's-eye view, the projection is likely to cause error in the pitch surface location of the player. This is known as out-of-plane error (Mauthner et al. 2008).

2.4.3 Out-of-plane error

Out-of-plane error is an error in a location measurement caused by the projection of a point to a calibrated plane. An analogy is to consider your shadow as the sun varies its position in the sky. Your shadow represents your location best when the sun is directly above because your shadow is directly beneath you. At any other angle, only your feet will still represent your location well because they are on the ground. Any portion of your body above the ground will cast a shadow that projects away from your location.

Figure 2.6 illustrates how a point digitised out-of-plane leads to an error in the reconstructed location, where out-of-plane means some non-zero value for the height coordinate. This error is exacerbated with non-perpendicular camera angles, just like the analogy with the sun and the shadow (Hinrichs et al. 2005).

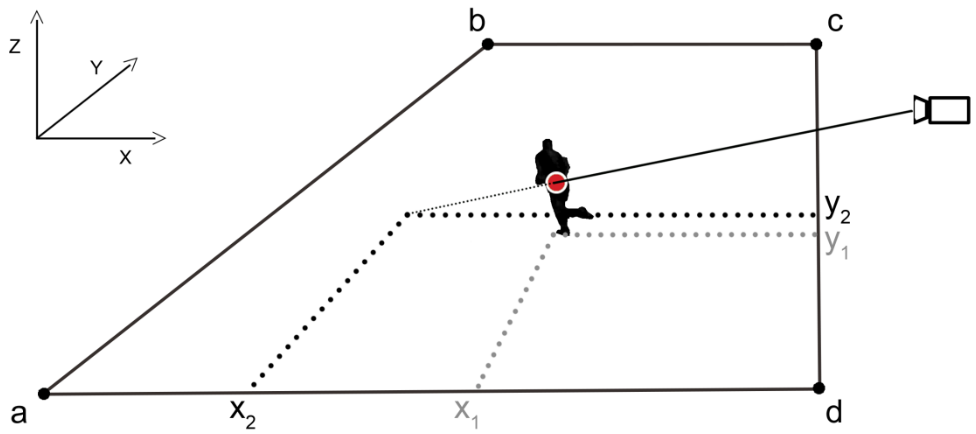


Figure 2.6 Reconstruction error. The digitised point (red dot) does not lie on the calibrated plane, $abcd$. Point (x_1, y_1) is the actual location of the player. Point (x_2, y_2) is the erroneously reconstructed location of the player.

Dunn (2014) used an elevated plane to reduce out-of-plane error in automatic tracking of tennis players (Figure 2.7). Dunn (2014) assumed that players' centre of mass would be at net height for the majority of gameplay. Creating a calibrated plane at this height minimises projection errors, under this assumption. Elevating the plane required access to the tennis court to collect images of a calibration object at the appropriate height. Access to the playing surface is not always possible and the height of players' centre of mass might have large variance. An alternative method is to translate the point of interest perpendicularly to the calibrated plane (Figure 2.8).

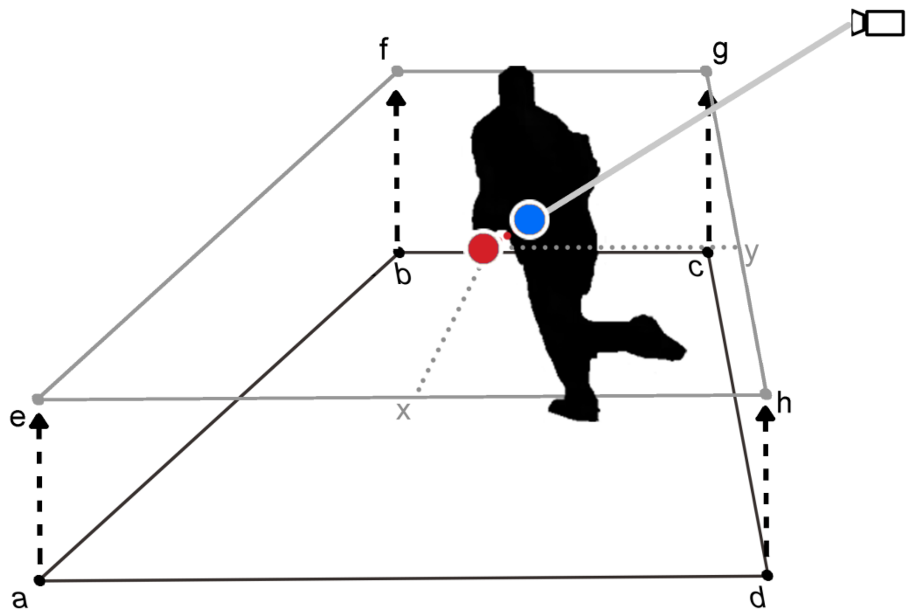


Figure 2.7 Elevating the calibrated plane to mitigate against out-of-plane error. The blue dot is the player's estimated centre of mass. The red dot is the resulting estimate of player location. The calibrated plane is elevated to a height that will minimise the average projection error. The dotted-red line represents the projection of the centre of mass to the elevated pitch surface, $efgh$.

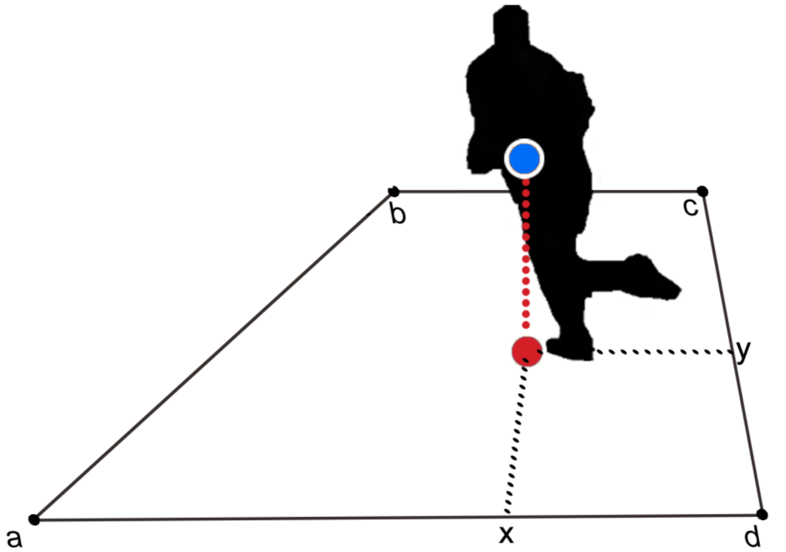


Figure 2.8 Translating the estimated centre of mass to the calibrated plane to mitigate against out-of-plane error. The blue dot is the player's estimated centre of mass. The red dot is the resulting estimate of player location. The dotted-red line represents the subjective translation of the centre of mass to the pitch surface, $abcd$.

Translating the point of interest perpendicularly to the calibrated plane requires subjective estimation, which is subject to random error (Glazier and Irwin, 2001). It is common to use the mid-point between the feet of the player as a representation of the translation of the player's centre of mass onto the ground (Ricardo Duarte et al. 2010; R Duarte et al. 2010; Headrick et al. 2012). This rationale is only valid when the player's mass is evenly distributed between their feet (which is only likely during quiet standing). During locomotion the mid-point between the feet only represents the translation of the centre of mass onto the ground for a fraction of the stance phase. Serrano, Shahidian and Fernandes (2014) quoted a mean absolute difference of 0.70±0.29 m when compared with a GPS worn on a player's back during a 273 m course. The 1 m accuracy of this definition is practically useful given the radius of influence around a player. Manually digitising the point between a player's feet might therefore be a sufficiently accurate and practical method.

2.4.4 Conclusion

The field hockey environment limits the methods of obtaining player location data. Vision-based methods are the only option to non-invasively obtain location data of players from both teams (Kelley et al. 2016). A single-camera system with a wide-angle lens provides an adequate solution that minimises equipment demands and disruption to the running of competitive events.

Single-camera systems can only estimate planar location of players and suffer from out-of-plane error. Where an elevated plane cannot be used, the errors can be reduced by subjectively estimating the point at which the downward projection of the player's centre of mass intersects the calibrated plane. This manual digitisation is the industry standard for obtaining player location from video of sporting performances. Automatic and semi-automatic methods have been developed but are not always available or sufficiently accurate (Bialkowski et al. 2013). Manual methods are practical and demonstrate sufficient accuracy.

2.5 Spatio-temporal metrics

Spatio-temporal metrics use information about player locations through time (Gudmundsson and Horton, 2016) to investigate tactical performance (Memmert et al. 2017). Spatio-temporal metrics use this data to define locations, distances, angles, areas, speeds, player distributions, and timings of and durations of events. These metrics are thought to capture the individual and collective behaviour of teams, but are seldom used to investigate tactical performance in field hockey (Section 1.1). Spatio-temporal metrics are widely used in sports like field hockey. Insights into potentially appropriate metrics can be gained by reviewing the metrics used in the literature associated with these similar sports. Therefore, a definition of what constitutes a similar sport is required.

2.5.1 Invasion game sports

Many taxonomies can be used to classify sports and the choice of taxonomy will depend on the interest of the analyst. If the intensity of gameplay is the focus of investigations then Mitchell, Haskell, Snell and Van Camp (2005) provide a scheme for classifying sports based on the physiological demands of sports' static and dynamic components. In comparison, Read and Edwards' (1997) taxonomy group sports according to tactical concepts and gameplay skills.

In Read and Edwards' (1997) taxonomy, a group called ‘invasion games’ includes sports characterised by two teams attempting to score against their opponent by invading their opponent's territory. This level of grouping deals with tactical similarities. Invasion games can be further divided based on gameplay skills, specifically, their method of scoring (Figure 2.9). The spatio-temporal metrics used by an invasion game sport are likely to be relevant for investigating tactical performance in other invasion game sports because they are tactically similar.

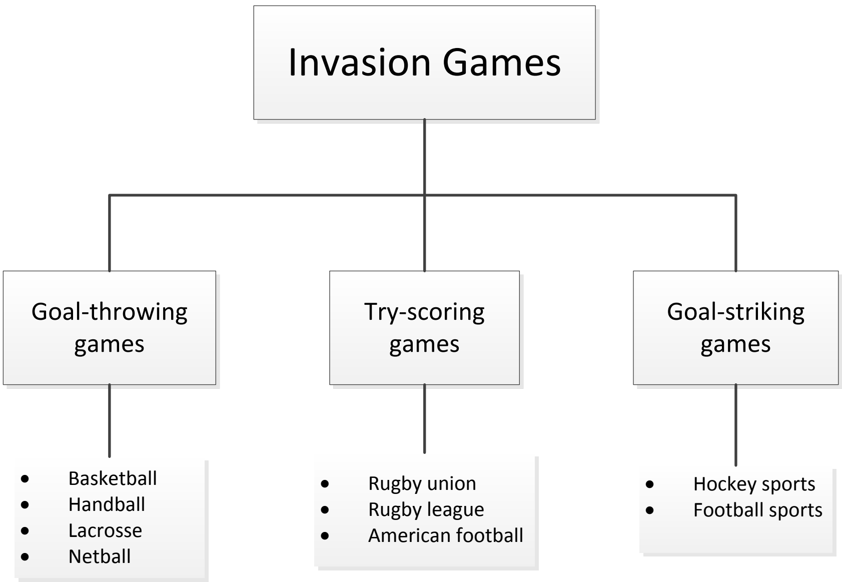


Figure 2.9 Invasion game sports with examples, as per Read and Edwards' (1997) taxonomy

2.5.2 Metric types

There are thousands of spatio-temporal metrics that could be defined depending on the spatio-temporal data that is available. The following paragraphs group them into eight types and provide a brief description. Table 2.2 provides a generalised summary of the spatio-temporal metrics typically used to analyse tactical performance in invasion game sports.

Table 2.2 Spatio-temporal metrics found in the literature relating to invasion game sports.

Type	Definition	Examples metrics	Example study
Location	The pitch/court location of players, player groups, the playing object or points of interest.	Location of turnover (i.e. where possession was gained/lost)	(Stöckl and Morgan, 2013)
		Location of gameplay action, e.g. shot, pass.	(Albinsson and Andersson, 2008)
		Location of group / Centroid	(Sampaio et al. 2014)
		Location of player	(Duarte et al. 2012)
		Location of the playing object	(Bruno Travassos et al. 2012)
		Location of predicted interception point	(Vilar et al. 2013)
Time	A moment during gameplay when an event occurred, or a period between events.	Timestamp of gameplay event	(Nevo and Ritov, 2013)
		Duration of gameplay action	(Platanou 2004)
		Duration between gameplay events	(Meyer et al. 2006)
		Duration of gameplay spent in-play	(James et al. 2004)
		Duration of gameplay spent out-of-play	(Ito et al. 2004)
		Duration of possession	(Kan et al. 2004)
		Duration of attacking plays	(Platanou 2004)
		Time to ball contact	(B. Travassos et al. 2012)
		Time to ball interception	(B. Travassos et al. 2012)
		Time to player collision	(Davids et al. 2013)
Distance	The difference between location metrics.	Inter-player distance	(B. Travassos et al. 2012)
		Inter-group distance / Centroid distance	(Folgado et al. 2012)
		Pass distance	(Travassos et al. 2013)
Speed	The rate of change of location metrics.	Speed of players	(Kan et al. 2004)
		Speed of the playing object	(Kan et al. 2004)
		Predicted speed for successful intercept	(Vilar et al. 2013)
Angle	The angle between three locations or two vectors.	Angle between player-goal centre vector and the baseline	(Dawson et al. 2004)
		Angle between attacker-defender vector and backline	(Passos et al. 2013)
		Angle between group centroid, goal centre and mid-line of the pitch/court	(Bruno Travassos et al. 2012)

Table 2.2 continued Spatio-temporal metrics found in the literature relating to invasion game sports.

Type	Definition	Examples metrics	Example study
Angle	The angle between three locations or two vectors.	Angle between the ball, goal centre and mild-line of the pitch/court	(Bruno Travassos et al. 2012)
		Angle between a player, goal centre and mid-line of the pitch/court	(Vilar, Araújo, Davids and Travassos, 2012)
		Angle between attack, defender and goal centre	(Vilar, Araújo, Davids and Button, 2012)
		Angle between the player vectors before and after an evasive manoeuvre	(Wheeler et al. 2010)
		Angle between the player vector after an evasive manoeuvre and the mid-line of the pitch/court	(Wheeler et al. 2010)
Spread	The dispersion of groups of players.	Stretch index	(Bartlett et al. 2012)
		Frobenius norm	(Moura et al. 2012)
		Group length	(Castellano et al. 2013)
		Group width	(Castellano et al. 2013)
		Major range	(Yue et al. 2008)
		Compactness / Playing area	(Okihara et al. 2004)
Area	Portions of the pitch/court associated with a player or group of players.	Surface area / Convex hull	(Frencken et al. 2011)
		Effective Area of Play	(Clemente et al. 2012)
		Individual playing area	(Zubillaga et al. 2013)
		Dominant region	(Taki and Hasegawa, 2000)
		Voronoi area	(Fonseca et al. 2012)
Context	Descriptions of gameplay scenarios based on player distributions.	Procrustes fit	(Jäger and Schöllhorn, 2012)
		Player density	(Clemente et al. 2012)
		Player distribution	(Couceiro et al. 2014)
		Field side changes	(Pratas et al. 2012)
		Penetration principle metric	(Clemente et al., 2014)
		Offensive unit metric	(Clemente et al., 2014)

Location

Location-type metrics refer to the location of players on the pitch. They are the simplest metric type and form the basis for more complex metrics. Individual player locations inform group centroids, which are the average location of players (Frencken et al. 2011). Figure 2.10 shows a team centroid but a group can be any collection of two or more players (Gonçalves et al. 2014). The locations of gameplay actions like shots (Abdel-Hakim 2014; Abreu et al. 2012; Alcaraz et al. 2012; Connelly 2013) and passes (Albinsson and Andersson, 2008; Barros et al., 2006; Cotta et al., 2013) are often measured and teams have been distinguished by the locations that their plays begin and end (Gómez, Prieto, et al. 2013).

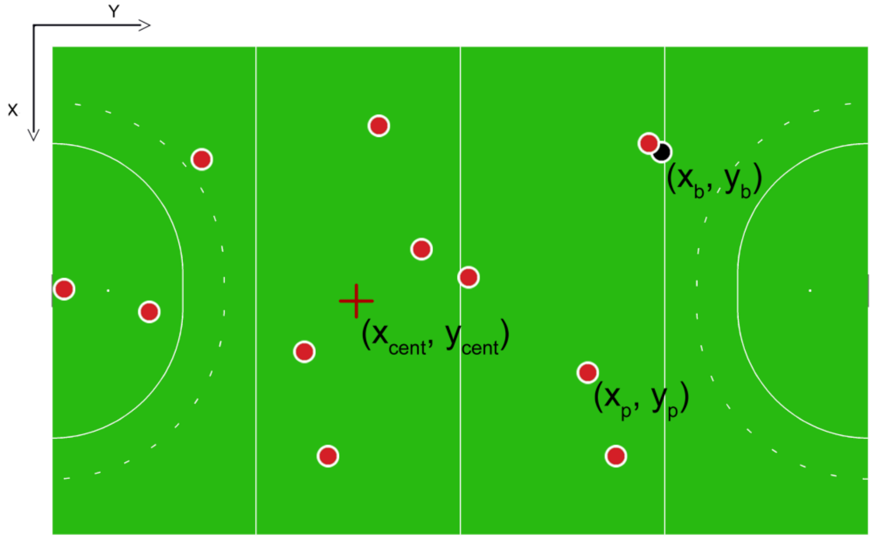


Figure 2.10 Example location-type metrics: Location (x_b, y_b) indicates the location of the ball (black filled circle); Location (x_p, y_p) indicates the location of a player; Location (x_{cent}, y_{cent}) indicates the location of the red team's centroid (red cross).

Time

Time type metrics refer to a moment during gameplay when an event occurred or refer to a period between events. Like location type metrics, they are fundamental to metrics that are more complex. Time is often considered as a categorical variable to partition data, for example, metrics are presented as summaries over the duration of a game period (Kan et al. 2004). Nevo and

Ritov (2013) are an example of a rare case when the time of the event is used as a continuous, interval variable. Nevo and Ritov (2013) used survival analysis to inform a time-dependent model to estimate the likelihood of a second goal when the time of the first goal is known. Stevenson & Brewer (2017) also show an example of using survival analysis to model the progress of batting ability through a cricket Test Match. Survival analysis has similarly been used to investigate the relationship between subsequent substitutions (Del Corral et al. 2008). Such methods lend themselves to modelling the time-dependent relationship between the gameplay events of a play.

Durations are commonly used when investigating playing intensities (Lago et al. 2010). The duration of gameplay actions are rarely investigated (Platanou 2004), presumably because many gameplay actions are momentary, e.g. kick, block or score. Where gameplay actions do take appreciable time, like a pass, other metric types are preferred. In the case of passes, distances are measured most often (Michael David Hughes and Bartlett, 2002; Bruno Travassos et al., 2013).

Durations of possessions and periods of gameplay are of tactical interest (Escalante et al., 2011; Mike D Hughes and Churchill, 2004; Jones et al., 2004). Investigations have characterised playing styles and modelled player behaviours using durations of in-play gameplay (James et al. 2004; Nevill et al. 2008), out-of-play gameplay (Eaves and Evers, 2007; Ito et al., 2004), attacking plays (Janković et al. 2011; Lupo et al. 2009; Platanou 2004; Pratas et al. 2012), and periods in between specific gameplay events (Meyer et al. 2006).

Time has also been an outcome metric for studies into player coordination, e.g. duration spent in and out of phase (Folgado et al. 2014). These studies inform tactical actions by relating player locations to predictions, e.g. the time until defender-attacker collision (Davids et al. 2013), time to ball contact and defender's interception (B. Travassos et al. 2012; Vilar et al. 2013).

Distance

Distance type metrics refer to the difference in locations (Figure 2.11). They are often specified as pitch-width, pitch-length and Euclidean distances. Metrics like the distance from the goal have been associated with the chances of scoring (Ensum et al. 2004). Vilar et al. (2014) analysed relative phases of inter-player distances to support previously held defensive ideas of marking and closing space.

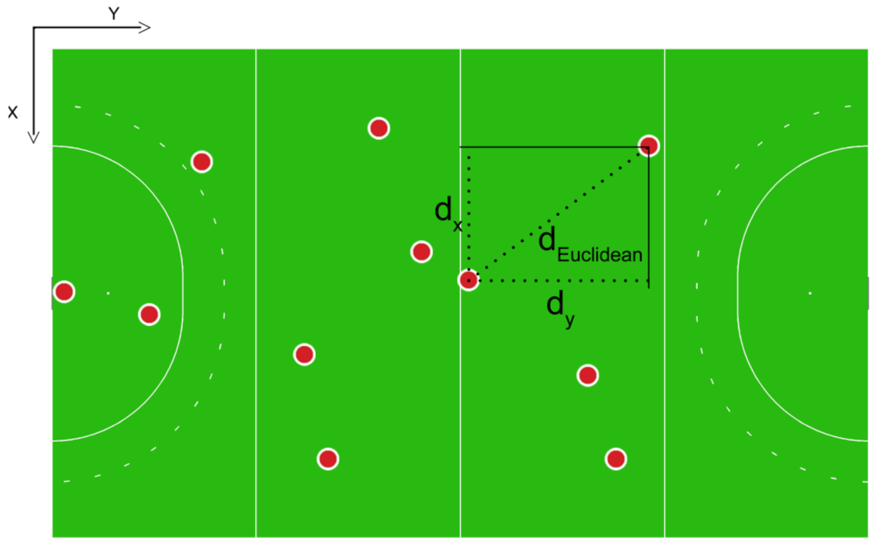


Figure 2.11 Example distance metrics: Euclidean inter-player distance, $d_{Euclidean}$, and axial inter-player distance, $\{d_x, d_y\}$.

Speed

Speed type metrics refer to the rate of change of locations. They are more frequently used as measures of playing intensity to inform training (Lidor and Ziv, 2015). Velocities required of a player to successfully intercept a have been associated with the outcomes of plays. (Vilar et al. 2013).

Angle

Angle type metrics refer to the angle between two vectors that share a point (Figure 2.12). The advantage of angles is that they incorporate information about player location in both axes. While investigating synchrony of attacker-defender interactions in basketball, Esteves et al. (2012) linked the angle of an attacker to baseline with their handedness. Different player coordination modes were attributed to the attacker's proximity to the baseline and each side of the court. It was suggested that the angle metric indicated the constraints of defender pressure and dribbling with preferred and non-preferred hands.

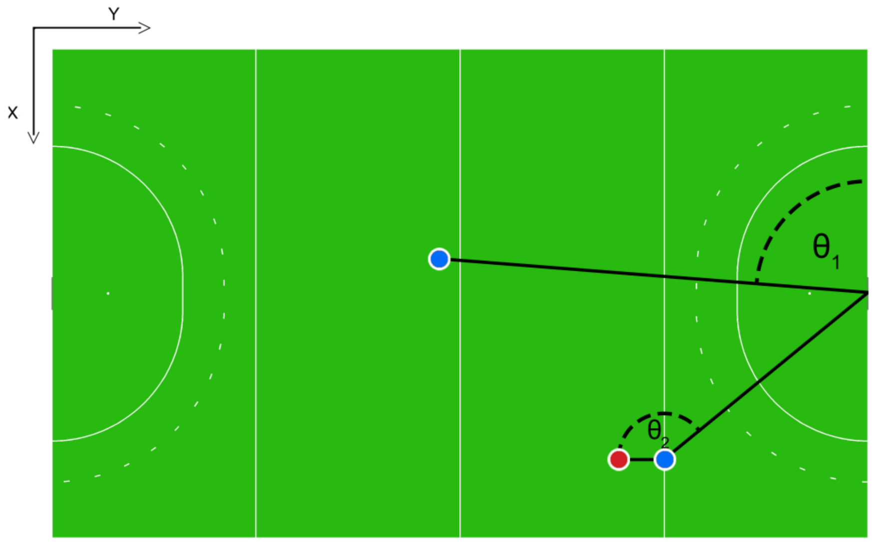


Figure 2.12 Example angle metrics: θ_1 is a Goal line-Goal centre-Player angle; θ_2 is a Goal-Nearest Opponent-Player angle for the red player.

Spread

Spread type metrics refer to how dispersed or spread-out groups of players are (Figure 2.13). Like distance type metrics, they are often specified as pitch-width, pitch-length and Euclidean distances. The court-length dispersion of basketball teams has shown an attraction to in-phase modes with no such attraction for the court-width dispersion (Bourbousson et al. 2010a), suggesting that teams seek to match longitudinal spread but in a variety of latitudinal formations.

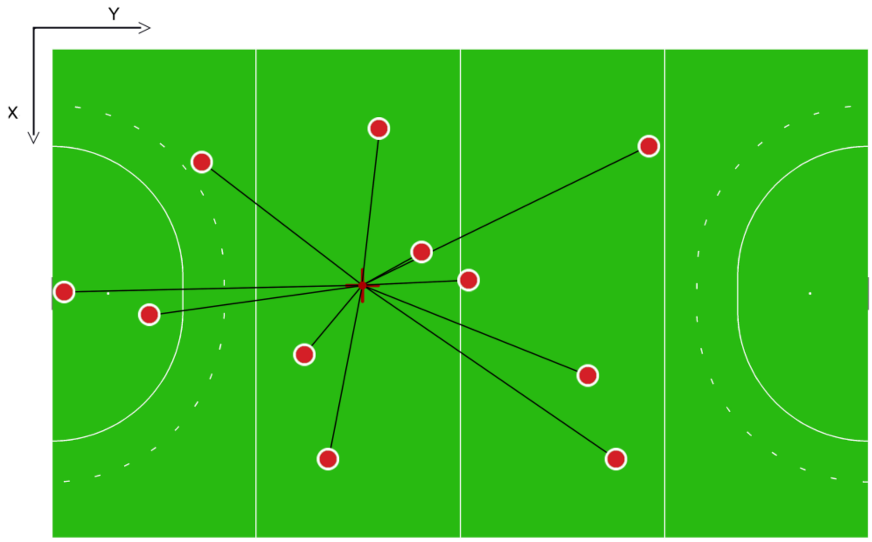


Figure 2.13 Example spread metric, the Radial Stretch Index is the arithmetic mean of the distances (black lines) between players and their centroid (red cross).

Area

Area type metrics refer to portions of the pitch associated with a player or group of players (Figure 2.14). They include the surface area of a group of players (Bourbousson et al. 2010a) and Voronoi areas (Fonseca et al. 2012). Voronoi areas segment the pitch into cells that contain the space closest to a given player. Dominant regions are a version of Voronoi areas that replace the distance function with a time function based on the instantaneous velocity of a player (Taki and Hasegawa, 2000).

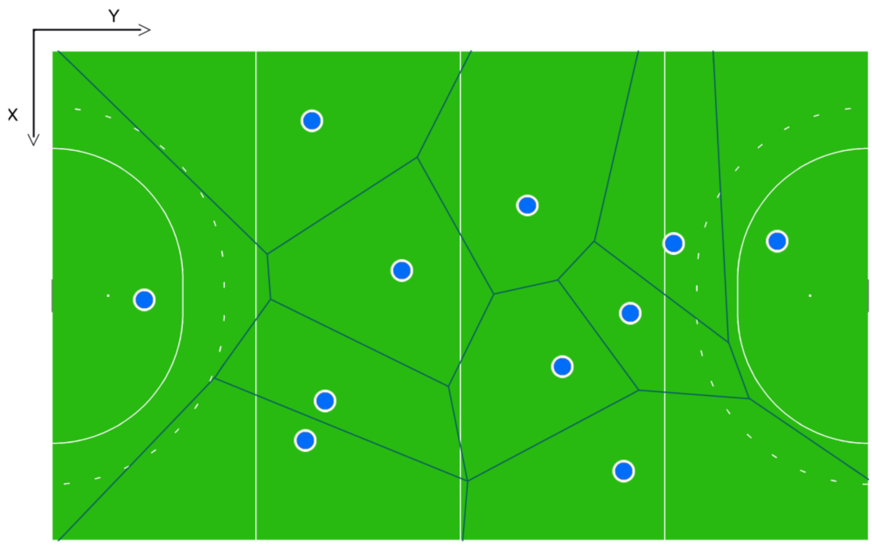


Figure 2.14 Example area metric: the blue team's Voronoi areas.

Context

Context type metrics refer to variables that describe a gameplay scenario. Some of these metrics describe the entire play while others describe the context of specific gameplay events. For example, the chances of scoring are inversely proportional to the number of players between the shooter and the goal (Ensum et al. 2004) and this can be measured by counting the number of players in the region. A whole-play example is the offensive unit metric (Clemente, Martins, et al. 2014b). This metric compares the locations of players at the start and end of a play to indicate whether at least half of the team moved in synchrony with the ball. The metric was one of three developed to measure tactical principles suggested by da Costa et al. (2009).

2.5.3 Metric specifications

Teams

The behaviour of both teams is of interest when investigating tactics because of the inherent cooperative and adversarial interactions (Section 2.2). Almost all metrics can be specified to a team, e.g. the number of attacking and defending players within the Circle.

Specifying the team has important implications for how some metrics are computed. The most noteworthy effect relates to a team centroid, which is the arithmetic mean location of a team's players. Before Clemente et al. (2013), team centroids tended not to include the goalkeeper in the calculations. Including the goalkeeper would draw the centroid back towards the goalkeeper because they tend to stay at their end of the pitch. This effect would be particularly noticeable in field hockey because the rules dictate that the goalkeeper cannot travel farther than 23 m from their goal line. When considering the computation of defensive team centroids, however, Clemente et al. (2013) argues that the goalkeeper should be included because of their implicit role in the defensive phases of gameplay. Including the offensive team's goalkeeper in centroid calculations would unrepresentatively pull the offensive team centroid back because team's often move forward together. Not including the defensive team's goalkeeper in centroid calculations would ignore their important role during intrusions. It might therefore be advisable to consider the goalkeeper for the defensive centroid only.

Player groups

Some metrics require the players be grouped into roles, specifically, defenders, midfielders and forwards. Exemplar metrics are the Defensive Play Area metrics (Clemente et al., 2015). The Defensive Play Area is composed of four regions defined by the roles of the players. For example, the backward region is defined by the space between defenders and the goalkeeper, while the first-half-middle region is defined by the space between the defenders and the midfielders. Several methods have been suggested for assigning these required player roles.

Roles can be assigned based on prior information about players' expected tactical function. Abdelkrim et al. (2007) took advantage of the traditional roles that players assume in basketball. From an offensive perspective, a characteristic of basketball is the use of set plays that traditionally involve guards, forwards and a centre based on their relative distance from the net (Krause and Pim, 2002). Basketball players will often categorise themselves with a specific role and will train for that role. An extreme example of such role specificity is in American football where players are assigned a specific role and assigned to either the offensive or the defensive team (American Football Coaches Association 2000b; American Football Coaches Association 2000a). In contrast, association football has a more open-play style so teams tend to use formations to assign roles to players. Players can be assigned pitch-length labels such as defenders, midfielders or forwards/strikers and pitch-width labels such as centre or left and right wing (Orejan 2011).

Abdelkrim et al. (2007) assumes that the behaviour of a player will stay true to their assigned role throughout the game. In a study investigating physiological demands of specific player positions in associated football, Gonçalves et al. (2014) state that they asked players to assume the standard behaviour of their positional role. This request is perhaps indicative of an understanding that player behaviour is indeed dynamic and not faithfully reflected by pre-defined tactical roles. Clemente et al. (2015) suggest that the demands placed upon a player will change throughout the game and proposed the idea of a tactical location or tactical mission (Clemente et al., 2014). A player's tactical location should reflect the momentary tactical role of a player based on their absolute location on the pitch and relative location to other players.

Lucey et al. (2013) supplemented prior information of team formation with Clemente et al.'s (2015) idea of the tactical location. Based on players' relative location to teammates, players were assigned a role from a set list that was based on a pre-determined team formation. This assignment method allowed for players who might momentarily assume a tactical role that they typically would not, e.g. a defensive wing player assuming a central role because the defensive centre player has been pulled wide.

The drawback of all methods listed so far is that they assume a team formation and that the team endeavours to be true to the formation. Using team formations is a style of play along with many others, which teams may or may not use. An alternative is to base the assignment of player roles purely on the location of the players. Two methods stand out in the literature based on players' absolute or relative location on the pitch. The first, based on absolute location, assigns player roles according to the region of the pitch where that player conducted most of their actions during a game (Di Salvo et al. 2007). This method can be interpreted as a typical tactical location under Clemente et al. (2015). This might be appropriate if the scale of investigation is games or tournaments but it does not allow for play-by-play differences in players' performance. It also assumes that the entire pitch is in use by the team, which might not be true if, for example, the team strategy is to press high up the pitch. In this scenario, many players might be assigned roles that are more forward than expected because the majority of gameplay is biased to one end of the pitch.

Clemente et al. (2015) propose a method based on the relative location of players on the pitch. Arising from the notion of a tactical location, players are assigned roles based on their relative distance from the rearmost player to the foremost player, also known as team length (Figure 2.15). Defenders are any outfield players that lie within 50% of the team length from the rearmost player, midfielders lie within 50% and 75% of the team length from the rearmost player, and forwards lie within 75% to 100% of the team length from the rearmost player. Like Lucey et al. (2013), this method allows player roles to be dynamic throughout the game. To its advantage, it does not require prior information about a team formation but it does lack the pitch-width information about the player that Lucey et al.'s (2013) method provides, i.e. centre or wing player. Other advantages of Clemente et al.'s (2015) method are that it does not assume players have only one role, as in Abdelkrim et al. (2007) and Gonçalves et al. (2014) methods, and although bounded by the rearmost player, it does not assume that the entire pitch is in use at all times, as in Di Salvo et al.'s (2007) method.

Image removed due to copyright

Figure 2.15 Clemente et al.'s (2015, Figure 2) schematic indicating the assignment of roles based on relative player location.

It is important to note that using the Clemente et al. (2015) method can result in there being no members of a specific player group. For example, if all attacking outfield players are near the defensive team's end of the pitch, then it is possible that all are classified as forwards because the goalkeeper is considered when deciding the rearmost player. Although the rearmost of the outfield players will have the responsibility to defend should the ball change possession, Clemente et al. (2015) argue that all players are behaving as forwards in this case so the assignment is valid.

Pitch dimensions

Many metrics can be specified as pitch-width, pitch-length, Euclidean or radial. The stretch index metric provides a good example. The radial stretch index is the arithmetic mean of Euclidean distances from each player to centroid of those players. These Euclidean distances radiate out from the centroid location providing a radial measure of spread (Figure 2.13). The Euclidean distances can also be decomposed into their axial components to provide pitch-width and pitch-length measures of spread (Figure 2.16).

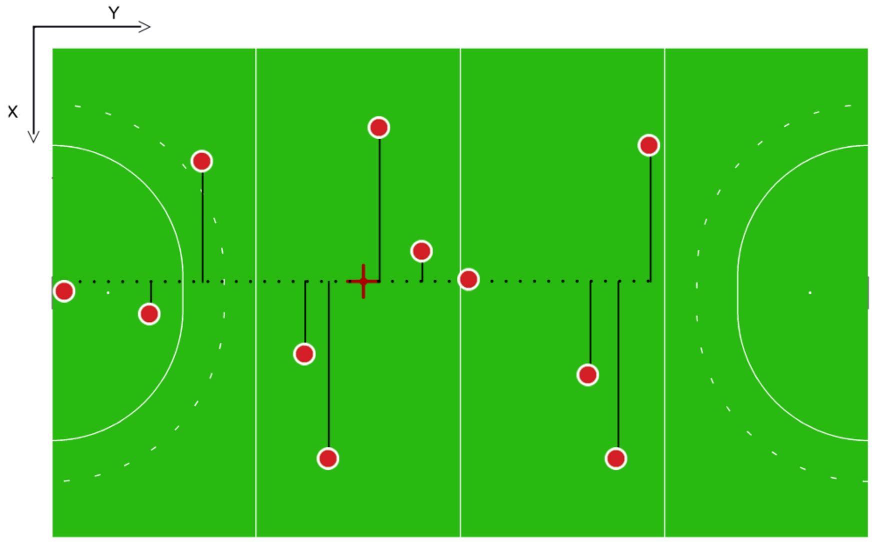


Figure 2.16 Illustration of the pitch-width Axial Stretch Index. The red cross is the team's centroid. Solid vertical lines indicate the pitch-width components of the team's Axial Stretch Index.

Weighting to ball proximity

It was previously noted that goalkeepers can adversely affect team centroids because they tend to stay around their end of the pitch. One solution posed by Clemente et al. (2013) was to only include goalkeepers in the computation of the defensive team's centroid. Another solution by Clemente et al. (2013) was to weight the centroid location according to players' proximity to the ball. The rationale behind this idea is that the ball represents the locus of play and players further from the ball have less of an effect on the game than those that are closer. Clemente et al.'s (2013) player weights, w_i, are calculated as

$$w_i = 1 - \frac{\sqrt{(x_i - x_b)^2 + (y_i - y_b)^2}}{d_{max}} \qquad [2.1]$$

where (x_i, y_i) describes the location of a player, (x_b, y_b) describes the location of the ball and d_{max} is the maximum Euclidean distance from a player to the ball. These weights are only applied to centroids and stretch indices (Clemente et al. 2013) but centroids and stretch indices are the foundations for many metrics, e.g. differences between group centroids (Frencken et al. 2012).

Summary statistics

Summary statistics are used to describe groups of data succinctly. For every metric that refers to a player, an average, maximum and minimum can be computed. Sometimes the summary statistic version of a metric would be of the same metric type, e.g. median and arithmetic mean distances of players to the ball are both distance type metrics. Other times, the summary statistic versions of a metric could be a different metric type, e.g. distances between pairs of players on a team are distance type metrics but the arithmetic mean inter-player distance for a team is a spread type metric.

2.5.4 Constraints on the use of some spatio-temporal metrics

Not all of the metrics found in the review can be applied in all cases. Collecting player locations continuously or discretely constrains options. For example, relative phase metrics can indicate coordination by measuring the degree to which two signals are in-phase (Stergiou 2004). The analysis requires a continuous stream of data. Authors have applied this and other coordination analyses to investigate how teams coordinate their relative expansion and contraction (Sampaio and Maçãs, 2012) and how players' locations coordinate with the location of the ball (Travassos et al. 2011). Similarly, entropy based metrics measure the variability within a time-series and require portions of time-series data to be compared.

Another constraint of some metrics is the unit of performance. Section 2.3 suggested a *play* as the unit of performance. Metrics like the coefficient of determination and the coefficient of variation (Duarte et al. 2013) are the result of analyses that would summarise the dataset and represent all plays rather than individual *plays*.

2.5.5 Conclusion

There are potentially thousands of spatio-temporal metrics than can be computed. For example, consider a tactical investigation using the stretch index metric. The number of metrics doubles if both teams are considered, triples again if player groups are considered, triples again if radial and axial variations are used, and is multiplied proportionally to the number of events of interest if a discrete analysis is conducted. It is easy to see how the number of spatio-temporal metrics increases as more players and perspectives are considered.

Without prior knowledge of the useful variants of metrics, all must be considered as inputs in an analysis because all metrics have the potential to distinguish outcomes. Several analysis methods can help to indicate the importance of the inputs, and some are specifically designed for the task. The following section reviews some of these methods with respect to the current project.

2.6 Analysing spatio-temporal metrics

The previous section presented the multitude of spatio-temporal metrics that can describe tactical behaviour in sports similar to field hockey. Some measure the collective behaviour of player groups. These metrics of collective behaviour include, for example, the total area covered by a team and the pitch-width spread of midfielders. Most, however, measure momentary behaviour of individuals, like instantaneous location, distance to nearest opponent and angle to the goal line. Performance analysts' resources are limited so they want to determine which metrics relate to the outcomes of plays most strongly. The choice of metrics will focus the evaluation and monitoring of performance.

There are a wide variety of methods available to performance analysts to investigate the relationship between the metrics and performance outcome (Gudmundsson and Horton, 2016; Gudmundsson and Wolle, 2014). This section reviews some of the methods already used in sports performance analysis as well as others that might be applicable.

The methods are designed to evaluate the importance of inputs based on their relationship with an outcome variable. Most methods are designed to accommodate only a small number of inputs. The methods discussed toward the end of the section are particularly well suited to the situation where there are many inputs.

It is assumed in the following discussions that the outcome variable of interest is categorical with at least two categories. This is to focus discussions on methods that can be applied the operational definition of a play provided in Section 2.3. An outcome variable with at least two categories can accommodate pair-wise comparisons, if not a comparison between all categories.

2.6.1 Univariate methods

The simplest relationship to investigate is between a single input and the outcome variable. Hereafter, the term ‘univariate’ is used when referring to methods that consider a single input, also known as an independent variable. Users of these methods are interested in finding individual ‘key performance indicators’ to guide training and competitive performance. Univariate methods focus on individual variables at the expense of not controlling for covariates in the model or during interpretation. They are therefore better suited to inputs that consider information about many factors of interest. The spatio-temporal metrics of collective behaviour, like average distance to goal or area of defenders, are good candidates because they summarise information about multiple players (Clemente et al., 2014; Gréhaigne and Godbout, 2013; Yue et al., 2008).

Tests of statistical significance ask whether zero is contained within the uncertainty in the true value of a difference between groups (Hopkins et al. 2009). They provide a probabilistic indication that groups differ with respect to some input. Important inputs can be selected based on this indication, usually when the level of significance is $\alpha = 0.05$. For example, Student’s paired and unpaired *t*-test have been used to indicate important inputs in futsal (Abdel-Hakim 2014), basketball (Gómez et al. 2006) and association football (Rampinini et al. 2009).

A statistically significant variable is not necessarily a good predictive variable (Lo et al. 2015). The assignment of significance based on a harsh threshold is often criticised because it is a rarely critiqued, persistent, arbitrary value (Kirk 1996) that can be achieved simply by careful research design and lacks a robust theoretical capacity to indicate a meaningful difference (Johnson 1999; Hopkins et al. 2009). There is also concern that the type 1 error will increase when multiple univariate comparisons are made. This is only relevant if the user’s interpretation is that the outcome groups are distinguished by a combination of statistically significant inputs (Perneger 1998). The user must remember that each statistical comparison is independent so the results from each must be interpreted independently. Univariate inferential statistical tests only concern an individual metric so their inferences cannot be combined with other univariate tests.

Parametric inferential statistical methods require knowledge of the distribution of variables' values. What's more, they often require them to be a specific distribution. For example, Student's *t*-test and analysis-of-variance all require the data to be normally distributed (Field 2009). Unfortunately, sports performance variables can have many different distributions like Poisson and Negative Binomial distributions (Brillinger, 2007; Casals and Martinez, 2013; Chan, 2011; Hirotsu et al., 2006; Pollard, 1985). This makes parametric inferential statistical methods inappropriate in many cases. Distribution-free or non-parametric tests provide an alternative (Ortega et al. 2009) but they lack power and cannot adjust for covariates (Hopkins et al. 2009).

Magnitude-based methods are suggested (Batterham and Hopkins, 2006; Liu et al., 2015) as an alternative to the theoretical shortcomings of significance testing (Kirk 1996; Johnson 1999). These magnitude-based methods have their own shortcomings (Welsh and Knight, 2015) prompting the inclusion of both tools to help interpret findings. In magnitude-based testing, inferences are based on whether the difference in groups’ measurements are greater than some appropriate measure of variance (Hopkins et al. 2009). They have been used to infer the importance of notational analysis variables in football (Liu et al. 2015) and rugby (Higham et al. 2014).

Regardless of whether statistical significance or magnitude-based methods are used, investigating metrics individually does not account for the possibility of interaction effects. Univariate methods have performed well for identifying important inputs (Haws et al. 2015) but considering metrics' interactions might be more informative (Escalante et al., 2011; Graham and Mayberry, 2014). Tactical performance is multifactorial with respect to players and their behaviour. Some spatio-temporal metrics try to capture this collective behaviour in individual metrics (Clemente et al. 2014; Yue et al. 2008). An alternative is to use analysis methods that attempt to account for the interactions between inputs. In doing so, they evaluate tactical performance from multiple perspectives. These are known as multivariate methods.

2.6.2 Multivariate methods - few variables

Multivariate methods of analysis use multiple inputs and/or outcome variables (Tabachnick and Fidell, 2001). Only methods that use multiple inputs are discussed in this section, particularly those that can be used to reduce an initial set of inputs. Some of these methods use magnitude-based and significance testing to score combinations of inputs or to score individual inputs conditional on others. Other methods use regression or association approaches to score inputs.

Logistic regression

Logistic regression models the relationship between inputs and a categorical outcome variable that is usually binary (Hosmer et al. 2013). An optimisation process determines a multivariate linear model that predicts the outcome category:

$$Y = b_1X_1 + b_2X_2 + \cdots + b_iX_i + c \qquad [2.2]$$

where Y is the outcome category, b_* are beta coefficients, X_* are the inputs, and c is a constant. This model is called a logit model and informs the probability of an outcome category via

$$P(Y) = \frac{e^{Y}}{1 + e^{Y}} \qquad [2.3]$$

where $P(Y)$ is the probability of outcome category Y, given the values of the inputs.

A unique model is created for every combination of inputs. Models can be compared using goodness-of-fit statistics and classification accuracy (Hosmer et al. 2013). Exhaustively building models, using best-subset or stepwise methods, provides a ranked list of models, whose combination of inputs might represent an elite selection or "*parsimonious solution*" (Atkinson and Nevill, 2001). This process is computationally expensive and impractical, especially when there are many possible combinations.

An alternative approach is to consider statistics that indicate the contribution of individual inputs within a given model. The beta coefficients, b, represent the rate of change of the inputs per unit change in the outcome variable, controlling for other inputs (Hosmer et al. 2013). On their own, b coefficients give some indication of the contribution of the input to the outcome classification. However, it is recommended that odds ratios are used instead. The odds ratio of an event is the ratio of the probability of that event occurring and the probability that the event does not occur. The odds ratio can sometimes be confused with probability alone, which represents a difficulty with interpreting the output of logistic regression.

Logistic regression is very common in sports performance analysis. It has been used to evaluate variables in association football (Ensum et al., 2004; Pollard and Reep, 1997a; Tenga et al., 2010), ice-hockey (Gramacy et al. 2013), handball (Massuça et al. 2014), basketball (Gómez, Lorenzo, et al., 2013; Teramoto and Cross, 2010), field hockey (Vinson et al. 2013) and rugby union (Bremner et al. 2013).

The assumptions of logistic regression are (Field 2009):

- No multicollinearity (no inputs can be correlated);
- Linearity (there is a linear relationship between the inputs and the logit of the outcome variable);
- Independence of error.

Other expectations are that the inputs are not extraneous, that the data from the inputs is complete, and that outcome categories cannot be perfectly predicted by the inputs (Field 2009). These assumptions and expectations are easy to uphold in comparison to other methods, which makes it an attractive option for analysis. Logistic regression is also advantageous because it can incorporate both continuous and categorical inputs. The expectation that the inputs are not extraneous, however, poses a problem when an exploratory approach is being used to determine important variables. The expectation is a precaution based on logistic regressions tendency to over-fit models when there are many inputs.

Discrete inputs are also a problem for logistic regression. They are accounted for using dummy variables, which are binary variables representing each of the possible categories of the discrete input. For example, if the variable 'colour' had three categories, {red, green, blue}, then three binary variables could be included in the model to indicate the category. To indicate 'red', the values of the variables would be {1, 0, 0}. To indicate 'blue', the values would be {0, 0, 1}. This method can substantially increase the number of variables in the model when there are many discrete inputs or when they have many categories.

As mentioned earlier, logistic regression is susceptible to over-fitting and inaccuracy when there are many inputs. Shrinkage techniques attempt to mitigate this effect (Steyerberg et al. 2001) but do not overcome the expectation of few inputs (Dreiseitl and Ohno-Machado, 2002). To make matters worse, new inputs must be introduced if interaction effects are included in a model. That is to say, the interaction of X_1 and X_2 is accounted for by creating a variable X_1X_2, which is added to the model and requires another beta coefficient to be estimated. The extent to which logistic regression can otherwise account for interactions between variables is limited to controlling for other inputs when evaluating a specific variable's contribution to the classification.

Discriminant analysis

Discriminant analysis is a regression method that can accommodate more than two categories in the outcome variable (Burns and Burns, 2008). Rather than being an extension of logistic regression, discriminant analysis relates to multivariate analysis of variance and is often used as a follow-up procedure. Linear discriminant analysis is used most often but quadratic discriminant analysis is possible. Like logistic regression, linear discriminant analysis determines a multivariate model, called the discriminant function, that predicts the outcome category:

$$D = v_1X_1 + v_2X_2 + \cdots + v_iX_i + c \qquad [2.4]$$

where D is a linear discriminant function, v_* are the unstandardized discriminant coefficients, X_* are the inputs, and c is a constant. Important inputs will tend to have large unstandardized discriminant coefficients, analogous to logistic regression's b coefficients.

The analysis models discriminant functions that maximise the distinction between outcome categories. The number of discriminant functions is one less than the number of outcome categories or the number of inputs, whichever is smaller. Subsequent discriminant functions attempt to maximise the distinction between outcome categories with models that are uncorrelated with previous discriminant functions. It is advisable to standardise inputs beforehand to eliminate scale differences between them that might skew the estimation of coefficients.

There are three ways to estimate the importance of inputs using discriminant analysis. Firstly, inputs can be dropped during the modelling of discriminant functions, leaving only those inputs that maximise the distinction between outcome categories. The combination of the remaining inputs are considered important discriminators of the outcome variable (Peinado et al. 2011).

The second way to estimate the importance of inputs is to score the variables based on the magnitude of the standardised coefficients of the discriminant function. Similar to multivariate regression, the magnitude of the coefficients indicates its contribution to the classification, controlling for the other variables. This method is not appropriate for assessing the contribution of categorical inputs (Burns and Burns, 2008).

The third way to estimate the importance of inputs is to score variables based on their Pearson correlation with the discriminant functions. These are called structure coefficients and are considered more accurate than the standardised coefficients (Burns and Burns, 2008). Any inputs exceeding a given threshold are considered important (Escalante et al. 2011; Sampaio et al. 2006; Lorenzo et al. 2010; Ortega et al. 2009). The typically quoted threshold is ≥0.3, based on Tabachnick and Fidell (2001). Inferential statistics are sometimes used to assist in making conclusions based on structure coefficients (Lorenzo et al. 2010; Ortega et al. 2009; Saavedra et al. 2010).

The benefit of discriminant analysis is that inputs can be ranked after poorly contributing variables have been dropped during the modelling process. The discriminant functions also allow the combination of remaining inputs to be evaluated for classification accuracy using an unseen dataset. This is like logistic regression, which is sometimes preferred because it is less restrictive and has fewer strict assumptions. The assumptions for discriminant analysis are:

- No multicollinearity (no inputs can be correlated);
- Univariate normality (each input is normally distributed);
- Multivariate normality (any linear combination of inputs must have a normal distribution);
- Homoscedasticity (variances among groups of inputs must be the same across the range of values);
- Inputs are continuous or binary.

Other requirements are that sample sizes for each outcome category are equal and that the sample size is at least five times the number of inputs (Burns and Burns, 2008). This latter point makes discriminant analysis unsuitable for scenarios in sports performance analysis where many variables are being considered and data is difficult to collect. Recent advances have made it possible to use discriminant analysis with higher-dimensional datasets (Kolar and Liu, 2013).

Principal Component Analysis

Principal component analysis is a non-parametric method that transforms inputs into a set of uncorrelated linear combinations (Pearson 1901; Hotelling 1933). It can be used to determine a combination of inputs that explains the greatest portion of variance in the outcome (Choi et al. 2008; O'Donoghue 2008). These linear combinations are called principal components. It can also be used as a dimension-reduction method for problems with many inputs, by selecting the variable in a combination that correlates strongest with the outcome variable (Hawkins 2011). For example, Moura et al. (2014) used principal component analysis to reduce an initial set of inputs for use in a clustering process that distinguished winners and losers.

Principal components are linear regression models of input combinations, ranked according to the variance in the outcome that they explain. Each subsequent principal component has no correlation to the previous principal component. This ensures that subsequent principal components account for as much of the remaining variance as possible, given the higher ranked principal component(s) (Semmlow 2004).

The correlations between the original variables and the principal components indicate how the variables vary together. The principal component would be said to represent the combined association of the variables with the outcome. Principal component analysis therefore has the potential to indicate metric combinations that are associated with specific outcomes.

Principal component analysis assumes that all inputs are equally important, which is not likely for very large numbers of variables and is analogous to logistic regression's expectation of non-extraneous variables. It also expects no outliers and prefers normally distributed inputs (this can be accounted for in the less-often used independent component analysis (Semmlow 2004)).

Principal component analysis is the first of the methods discussed whose primary purpose is to provide variable combinations. Potentially important variable combinations can be extracted using logistic regression and discriminant analysis but this is a welcome consequence rather than an intention. Also, principal component analysis might better account for variable interactions than discriminant analysis's structure coefficients by allowing for inputs to be correlated (Burns and Burns, 2008). These correlations can be viewed positively or negatively: positively, the correlations might indicate interactions between the variables and can also be used for dimension-reduction; negatively, the correlations might indicate redundancy in the variable combination.

Often cited disadvantages of principal component analysis are that it assumes a linear problem (Saxen and Pettersen 2006), that the principal components might be difficult for coaches and athletes to interpret (O'Donoghue 2008) and that requiring orthogonal principal components is too restrictive. Principal component analysis might be too conservative by providing only the best-of-the-best uncorrelated input combinations. This might be at the expense of similar scoring combinations that might be easier to interpret and implement. This is particularly relevant in the applied domain of sports performance analysis.

2.6.3 Multivariate methods - many variables

The current work is concerned with spatio-temporal metrics of which there are thousands to consider (Section 2.5). All methods discussed so far prefer the number of inputs to be minimised before analysis. This might not be possible or valid if there is no prior information to guide the minimisation. These previously discussed methods are insufficiently equipped to determine important spatio-temporal metrics from the large set suggested in section 2.5. Additional shortcomings of previous methods are their inability to model non-linear relationships or to consider interaction effects in a convenient manner. There are, however, methods that have been specifically designed to consider large numbers of inputs and their interaction effects. Some of these methods can also indicate the importance of variables and variable combinations.

Neural networks

Neural networks can be conceptualised as a network of regression models, in series and parallel, where the output of some models become the input for others (non-linear regression models are typically used). Although the parameters of the individual regression models can be adjusted, neural networks are considered non-parametric because the data informs the overall model estimates during the learning process (Bishop 1995). Therefore, no prior knowledge of the data distribution is required.

Neural networks are often depicted as undirected graphs, where nodes represent the regression models and edges represent the transfer of model outputs from one node to another (Figure 2.17). The information passed between nodes is adjusted by connection weights.

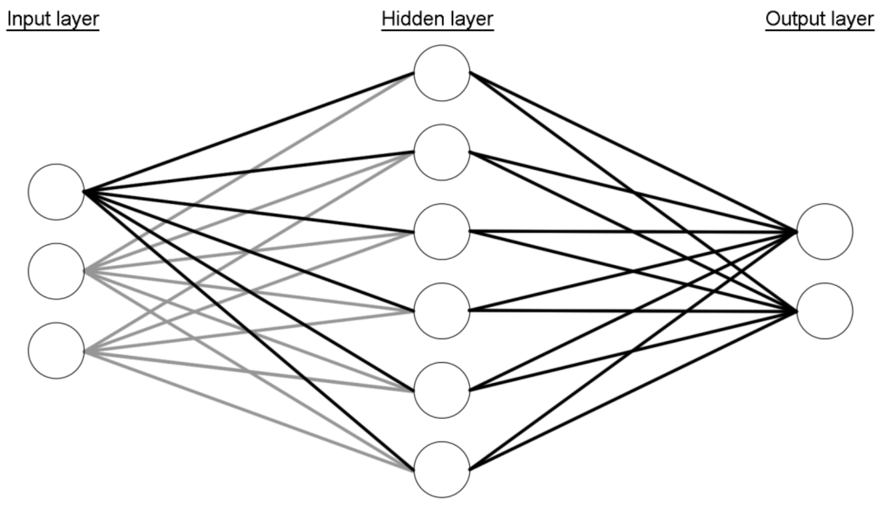


Figure 2.17 Classic depiction of a neural network where nodes (circles) represent regression models and edges (lines) represent the transfer of model outputs from one node to another to be used as input.

Neural networks have been used in sports performance analysis since the 1990s (Nevill et al. 2008; Dutt-Mazumder et al. 2011) for prediction and classification (Condon et al. 1999; Silva et al. 2007; Maszczyk et al. 2010). For example, self-organising maps are unsupervised neural networks used for pattern recognition (Kohonen 1997). The method consolidates similar input trajectories on a neural network architecture to provide a representative output trajectory. Self-organising maps recognise similarities rather than exact matches. This makes them useful for sports performance analysis because of the inherent variability of player behaviour that yields similar outcomes. Self-organising maps have been shown to recognise tactical patterns of play with 84 % accuracy (Perl et al. 2013) and successfully distinguish teams' tactical preferences (Pfeiffer and Perl, 2006).

Neural networks are criticised for being 'black boxes', i.e. a computational mechanism whose parameters are not well understood. This is not so much of an issue for classification tasks because model accuracy is more important than understanding how the model is structured. Nevertheless, several methods have recently been devised to understand the parameters of neural networks to help determine the importance of inputs (Olden and Jackson, 2002).

Satizábal and Pérez-Uribe (2007) discuss three approaches to determine the relevance of inputs to neural networks. There approaches are 'filter', 'embedded' and 'wrapper', and represent pre-, during- and post-training foci. Filter methods are applied before the network training stage, embedded methods make modifications to the learning algorithm, and wrapper methods evaluate the performance of input combinations. Satizábal and Pérez-Uribe (2007) note that all of these approaches discard inputs with low relevance rather than indicate inputs of high relevance. This means that inputs and combinations of inputs cannot be ranked.

Olden, Joy and Death (2004) conducted a review and concluded that the Connection Weight statistic was best. This statistic was used by Kingston, Maier and Lambert (2006) with a Markov chain Monte Carlo method to create distributions of possible weight vectors to provide probabilistic estimates of input relevance. The rationale is that nodes whose outputs have large weights are contributing most to the model and are therefore important. The Connection Weight statistic is the sum of the products of weights in the portion of the network from the input node of interest to the output layer (black connections in Figure 2.18).

None of these methods can indicate whether the remaining 'relevant' inputs are of predictive or causal importance (Sarle 2000). Furthermore, because of the potential complexity of the neural network model, there are many different relationships that might result in similar fits of inputs to outputs (Kingston et al. 2006). Therefore, selected inputs might only be relevant to a model’s training and input-reduction history (Saxén and Pettersson, 2006).

Neural networks are also very susceptible to over-fitting due to their often complex architecture (Sarle 2007). Some methods to mitigate against over-fitting are stopping the learning process early (Sarle 1995), Bayesian learning (Buntine and Weigend, 1991; Neal, 1995), regularisation (Zou 2005), and weight decay (Moody 1992). These mitigations have unresolved shortcomings that maintain concern about over-fitting. For example, after how many training cases does early-stopping optimise learning (Prechelt 2012)? What should the weight reduction factor be (Dreiseitl and Ohno-Machado, 2002)?

For any model, too little data increases bias and does not allow for sufficient 'learning'. There is no accepted rule about how many observations are required to build an optimal neural network and any answer will depend on the network architecture and any methods used to mitigate against over-fitting (Sarle 2007). Suggestions relate to the number of inputs and the number of edges in the network. Examples range from 10-times the number of inputs (Sarle 2007) to 30-times the number of edges in the network (Amari et al. 1995). This is not a problem for typical applications of neural networks like image processing because billions of images can easily be obtained directly or second-hand. These data demands, however, make neural networks impractical for sports performance analysis of many variables where observations can be difficult to collect. The size of the network could be reduced to reduce data demands but only at the risk of under-fitting and subsequently high bias and generalisation error (Geman et al. 1992; Sarle 2007).

Genomic selection

Genomic selection refers to a group of methods that take a genome-wide approach to selecting indicators of an expressed trait (Nakaya and Isobe, 2012). As the term suggests, these methods relate to the study of genetic inheritance, where the identification of predictive genes have many direct and indirect influences on the development of an organism. Because genomic selection methods take a genome-wide approach, they deal with many variables, i.e. genes.

Genomic selection problems differ from the traditional statistics and machine learning domains because the number of observations are drastically smaller than the number of variables (Haws et al. 2015). This is known as the 'large *p*-small *n*' problem (Loh 2012), caused by the large quantity of information gathered from infrequent collection of genetic data. Genomic selection methods assume that variables have small effects and that there are few 'influential' variables in the large set provided. These are more realistic assumptions than principal component analysis's assumption that variables have equal importance and logistic regression's expectation that there are no extraneous variables.

Genomic selection methods include linear methods, like varieties of ridge regression (Whittaker et al. 2000), and non-linear methods like some support vector machines (Smola and Schölkopf, 2004). There are also univariate selection methods, like minimum Redundancy-Maximum Relevance (Peng et al. 2005). The minimal redundancy criterion is similar to neural network relevance metrics in that it penalises variables that show a small contribution to classification. The improvement is to concurrently reward variables based on their mutual information with the classification, hence maximum relevance.

A general disadvantage of genomic selection methods is the computational demand of algorithms. This is mostly attributed to the large set of inputs rather than poor performance. Actions can nevertheless be taken to mitigate against impractical runtimes. For example, almost all multivariate genomic selection methods have either a forward or backward heuristic to their algorithm. These are similar to stepwise regression, where variables are added to less complicated models (forward) or removed from an initial all-encompassing model (backward). Backward elimination is computationally costly because the process must start with all the largest possible models. Conversely, the shortcoming of forward selection is that variables are only added based on large marginal contributions or interaction effects with the limited set of variables already in the model. This biases marginal contributions of individual variables and makes it difficult for variables that might strongly contribute via interactions.

The Backward Dropping Algorithm is a genomic selection method that addresses the concern for variable interactions well. It is an efficient but computationally expensive algorithm that searches for the 'best' subset of variables from a given set (Wang et al. 2012). It was originally introduced by Lo and Zheng (2002) as the backward haplotype-transmission association algorithm, where 'haplotype' refers to a set of genes that tend to be inherited together. It is assumed that these genes might express an interaction effect by virtue of their apparent connection. Lo and Zheng (2002) wanted to develop a genomic selection method that considered these interactions but did not need the genes to be closely located. The parallel with spatio-temporal metrics is to consider combinations of metrics that together distinguish performance but which are not explicitly inter-dependent.

The algorithm is “backward” because it uses stepwise elimination to determine the 'best' subset. The algorithm works as follows:

1. Consider a collection of discrete inputs and a binary response variable.
2. Randomly select a set of inputs of length k.
3. Score this variable set according to its association with the response variable.
4. Tentatively drop one variable from the variable set and score the remaining subset of variables.
5. Reinstate the dropped variable and tentatively drop the next. Repeat the tentative dropping until all k-1 subsets have been scored.

6. Consider all k-1 scores and carry forward the best scoring subset.

7. Repeat the tentative dropping procedure until only individual variables remain.

8. Consider all variable sets and choose the one with the highest score, regardless of the size of the set. This set is called the return set and represents the ‘best’ combination of variables from the initially selected set.

If no input is influential then the variable set scores will not change much in the dropping process (Wang et al. 2012). If some of the inputs are influential then their absence will be noticed in the scoring.

Each iteration of the Backward Dropping Algorithm randomly selects a small set of variables from the large collection of inputs. It is therefore applied many times to gain coverage over all variables. In doing so, it provides a ranked list of metric combinations from which the contribution of individual metrics can be computed, if desired (Chernoff et al. 2009).

The Backward Dropping Algorithm can consider interactions by measuring association with combinations of variables. This is better than regression-based methods that require new variables to represent interactions, adding to the size of the problem. When there are many variables, regression-based approaches become impractical (Wang et al. 2012). The Backward Dropping Algorithm can also handle a modular effect of variables, i.e. an effect associated with a combination of variables, which methods like Sure Independence Screening cannot (Haws et al. 2015; Wang et al. 2012). The Backward Dropping Algorithm is used as part of larger algorithms that, for example, assess association at different orders of interaction (Wang et al. 2012). The order of the interaction refers to the number of possible interactions given a set of variables. A small set of variables only demonstrates a low order of interaction while a larger set demonstrates a high order of interaction. Other algorithms associated with the Backward Dropping Algorithm try to mitigate against conservative filtering by ‘resuscitating’ variables that are underrepresented (Chernoff et al. 2009). These methods are relatively new as far as the author is aware, have not been applied outside of genomic selection problems.

2.6.4 Conclusion

Choosing an appropriate method to identify important metrics must consider the characteristics of the data. It is expected that the spatio-temporal metrics will be a mixture of continuous and discrete data, so linear discriminant analysis is inappropriate. Previous studies suggest that some metrics might be non-normal distributed (Gonzaga et al., 2014; Harrop and Nevill, 2014; James et al., 2005; Sullivan et al., 2014) and thus require non-parametric analysis or transformation. This makes logistic regression a manageable but awkward option. Principal component analysis is not a good choice because there is no prior information to suggest that metrics and metric combinations will be linearly related to the outcome.

The overwhelming concern is the capacity of the method to analyse a large set of metrics because section 2.5 identified so many. The only candidates therefore are neural networks and a genomic selection method like the Backward Dropping Algorithm. The two main disadvantages of neural networks are:

- they require many observations to train the model;
- non-binary categorical variables would require a new variable for every category, thereby increasing the number of inputs further.

The advantages of the Backward Dropping Algorithm over neural networks are:

- few observations are needed;
- variable combinations are provided as output;
- variable and variable combinations are ranked;

It has also been shown to out-perform linear discriminant analysis, support vector machines, random forests and variations of logistic regression in a benchmark test for classification error (Wang et al. 2012). The disadvantage of the Backward Dropping Algorithm is that the input data must be discrete with as few categories as possible but greater than one. Using this method would require the discretisation of continuous metrics, leading to a loss in information. Wang et al. (2012) suggest that the discriminatory capacity of the method outweighs the potential loss of information. Furthermore, although not strictly a univariate method, the algorithm is using marginal effects to inform its choice of variable combinations (Chernoff et al. 2009). The algorithm does not compare all possible combinations of variables in a k-variable subset so it might not present the truly-best combination.

Given these considerations, the Backward Dropping Algorithm is an appropriate method to identify spatio-temporal metrics and metric combinations that distinguish play outcomes in field hockey. Appendix A discusses considerations relating to the algorithm's parameters.

To the author’s knowledge, genomic selection methods have never been applied to problems in sports performance analysis. The univariate and multivariate methods discussed in this section have often been insufficient or misused (Hopkins et al. 2009). The context of the current work is also atypical, asking what variables we should use rather than what these variables tell us. The latter question is often the result of investigators delimiting the number of variables based on familiarity or expert opinion (Hraste et al. 2008; Bremner et al. 2013). Without guidance from domain experts, the current work takes a data-driven, exploratory approach to determining ‘important’ variables. There are clear parallels between the ‘large *p*-small *n*’ structure of genomic selection problems and the current work’s ratio of spatio-temporal metrics and likely number of observations.

2.7 Aim and objectives

The aim of the current work was to determine the spatio-temporal metrics that distinguish play outcomes in field hockey.

The objectives to achieve this aim were:

1. To determine the repeatability of player location measurements.
2. To prepare data for analysis.
3. To reduce the number of metrics based on performance in low-order interactions.
4. To select metrics based on performance in higher-order interactions.
5. To select metrics based on performance in higher-order interactions with previously unselected metrics.

Objectives 3 and 4 follow the Wang et al.'s (2012) algorithm and objective 5 follows an adjunct algorithm by Chernoff et al. (2009).

2.8 Thesis structure

Chapter 3 explains the data collection and data processing methods, and presents a study that investigated the repeatability of player location measurements.

Chapter 4 demonstrates how the data were prepared for subsequent analysis. The chapter explains how the process raised concerns about data integrity and describes how those concerns were handled.

Chapter 5 explains the first stage of the analysis, which reduces the number of metrics to be considered. This first stage involves a two-phase process that evaluates a metric's performance in two-way interactions.

Chapter 6 explains the second stage of the analysis, which measures the influence of metrics based on their performance in higher-order interactions.

This three-phase process also reduces redundancy and removes false positives. The chapter includes details of an adjustment to the originally published method that was required for the current work.

Chapter 7 explains the third stage of the analysis, which assesses the robustness of the previous analysis and gives a second chance to metrics that had not performed well thus far.

Chapter 8 takes a closer look at the best metrics and metric-combinations and describes them in more detail. The thesis concludes with practical implications of the study’s findings.

Chapter 9 concludes the thesis with a summary of the work, insights gained, and discussions on limitations and future work.

Chapter 3 Data collection and assessment of repeatability

3.1 Introduction

This chapter describes how data for the current work was collected. Inclusion-exclusion criteria and the working definition of player location are described. This chapter also addresses the first of the project's objectives: to determine the repeatability of player location measurements. Repeatability of measurements refers to the variation in repeat measurements made on the same subject under identical conditions (J. W. Bartlett and Frost, 2008; Taylor and Kuyatt, 2001). It is an important consideration in sport performance analysis (O'Donoghue 2010) and general sport sciences (Hopkins 2000). For the current work, intra-operator repeatability of digitising is defined as the estimate of variation in spatial data created from the digitising process. An acceptable range of repeatability is stated for spatial and temporal measurements and used in a test-retest design. The results are discussed with respect to maximum and expected errors and possible sources of the errors are provided.

3.2 Method

3.2.1 Data collection

Gameplay footage

With institutional ethical approval, England Hockey provided footage of games from the Men's tournament of the EuroHockey Championships 2015 (Appendix B). Only games played between the four teams that reached the medal stages were used. This provided six games: both medal games, two semi-final games and two games from the pool stage of the tournament. One semi-final game was excluded because the footage was corrupted.

Two sources of footage were provided; one from which player locations were extracted and another that provided an alternative perspective to help locate players on the pitch (Figure 3.1). The first source of footage was collected using a 4K camera (resolution: 3840 x 2160 pixels; Sony FDR-AX1, Japan) positioned approximately at the halfway line (camera 1 in Figure 3.2). This camera was stationary and equipped with a 0.3x fisheye lens (Digital Nc - Optics Nc, USA) capturing at 25 Hz progressive scan. Camera settings like focal length and aperture were optimised using the camera's auto-adjust functionality at the start of each recording. The initial settings were fixed for the duration of the recording because of their effects on the calibration used to convert player locations from image to pitch coordinates (Abdel-Aziz and Karara, 1971).

Image removed due to copyright

Image removed due to copyright

Figure 3.1 Sample views from each camera. *Top*: 4K camera from halfway line, used for manual digitisation. *Bottom*: High-definition camera from behind a goal, used as alternative view to see players.

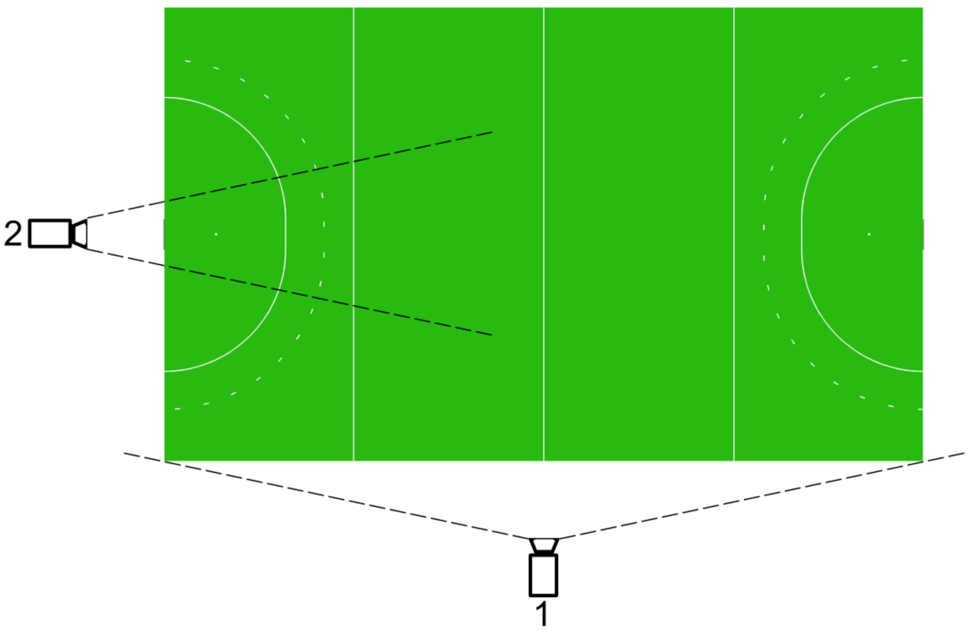


Figure 3.2 Schematic of camera locations. Camera 1 is a 4K camera with 0.3x fisheye lens. Camera 2 is a high-definition camera with a standard lens.

The second source of footage was collected using a high-definition camera (resolution: 1920 x 1080; JVC GY-HM650E, Japan) positioned behind a goal in a viewing tower (camera 2 in Figure 3.2). This camera captured at 25 Hz interlace scan and was free to pan, tilt and zoom to follow the events of the game.

Player location

Player locations were digitised using an in-house software tool developed in Microsoft Visual Studio using the .NET framework. The digitised player location was defined as the point at which the downward translation of the player's estimated centre of mass intersected with the estimated surface of the pitch. This is a common definition used in sports performance analysis (Section 2.4.3).

Figure 3.3 shows the process by which the digitised point was determined when a player was airborne. Briefly, the last frame when the player was in contact with the pitch is considered and the location of the foot that is in contact with the pitch is noted (1 in Figure 3.3). This is called the ‘foot-off’. The first frame when the player re-contacts the pitch is considered and the location of the foot that is in contact with the pitch is noted (2 in Figure 3.3). This is called the ‘foot-on’. The frame of interest is considered and a line is estimated between the locations of the ‘foot-off’ and ‘foot-on’ to provide an estimate of the pitch surface (4 in Figure 3.3). An estimate of the player’s centre of mass is translated downward until it intersects with the estimated pitch surface (5 in Figure 3.3). The point of intersection of these lines is the digitised location of the player (6 in Figure 3.3).

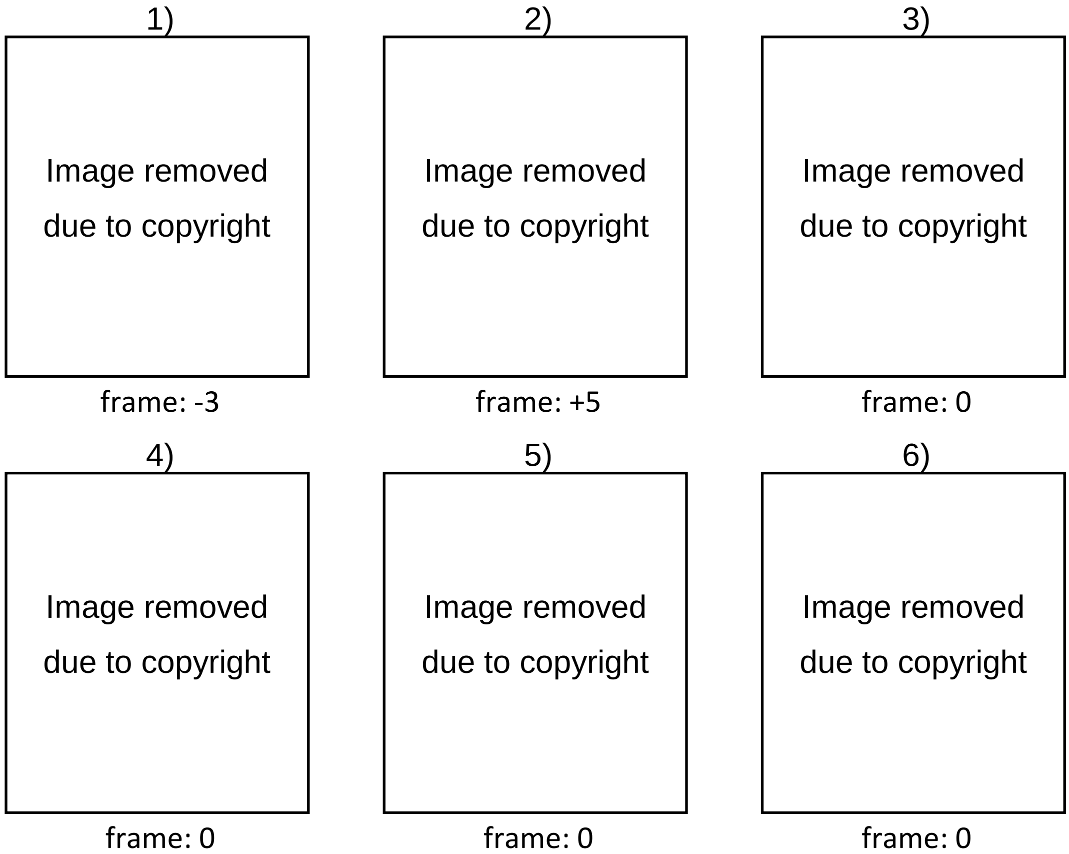


Figure 3.3 The six steps to defining player location (red dot) when the player is airborne. Frame 0 is the frame of interest.

Footage from the 4K camera was used for the digitisation. Footage from the high-definition camera assisted the digitisation process by providing an alternative perspective of gameplay. The high-definition footage was viewed on a separate screen using VLC media player (VideoLAN 2015) because of its advanced zoom and playback functions. All digitisation was completed by the same operator. To avoid operator errors due to fatigue, mandatory breaks were taken every hour and the operator was provided with an adjustable workstation to reduce sedentariness and improve comfort. The operator was also encouraged to keep hydrated and was provided with a large HD screen to reduce eyestrain.

Dataset of interest

The locations of the ball and all players were digitised at the start and end of plays. The definition of a play was informed by general and sport-specific considerations discussed in chapter 2 (Section 2.3). The start of a play was defined by a 23 m intrusion and ended with a 'Circle Entry', a 'Turnover (conceded)' or an 'Other' outcome (Figure 3.3). These outcomes represent positive, negative and partially positive outcomes, from the offensive perspective. Operational definitions are shown in Table 3.1 and 3.2.

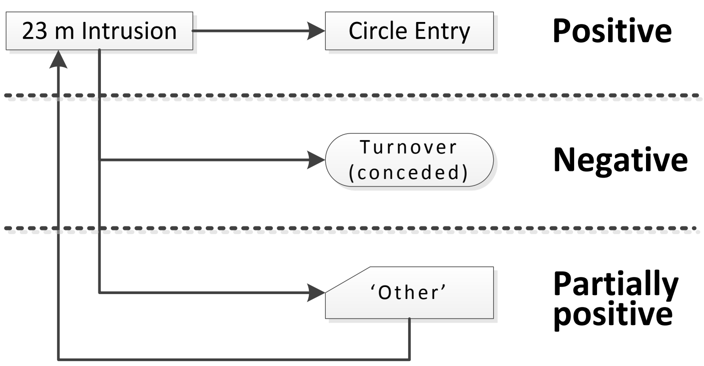


Figure 3.3 Schematic summarising three possible plays (excluding 'Other'-to-23 m intrusion). The defining events are easily identifiable and some are objectively indicated using player location data.

Only plays that occurred during standard 11-v-11 gameplay were included in the dataset. This scenario was predominant in the tournament from which the footage was collected (89.9% of gameplay duration).

Table 3.1 Operational definitions of the positive gameplay events that define the start and end of plays in Figure 3.3.

Positive	
Outcome:	Circle Entry
Definition:	The ball enters the Circle.
Requirements:	*Dribble*: The moment the ball crosses the Circle line.
	Pass: The ball is passed into the Circle and received within the Circle, under control or deflected. The moment the ball crosses the Circle line.
Outcome:	23 m Intrusion
Definition:	The ball enters the 23 m region.
Requirements:	*Dribble*: The moment the ball crosses the 23 m line.
	Pass: The ball is passed into the 23 m region and received within the 23 m region, under control or deflected. The moment the ball crosses the 23 m line.
	Interception or steal: Within the 23 m region, the moment the intercepting or stealing player contacts the ball for the action that relieves the opponent of possession. A 23 m Intrusion is not applied if an intercepting or stealing player only deflects a pass or shot and the ball is subsequently controlled by an opponent.

Table 3.2 Operational definitions of the partially positive and negative gameplay events that define the start and end of plays in Figure 3.3.

Partially positive	
Outcome:	'Other'
Definition:	Any event where the play ends but does not result in a Circle Entry and possession is maintained by the attacking team.
Requirements:	*Penalty corner awarded*: The moment when the umpire blows the whistle to award the penalty corner OR the moment when the awarding team calls the umpire to request a video replay and the umpire agrees (FIH, 2015, Section 12.3). *23 m exit*: The moment the attacking team dribble or pass the ball back over the 23 m. *Free hit* (FIH, 2015, Section 12.2): The moment when the offending event appeared to occur. *Backline by defender* (FIH, 2015, Section 7.4b): The earliest moment the ball is known to be going over the backline off a defender's stick.
Negative	
Outcome:	Turnover
Definition:	Any event where the rules of the game stipulate a change in possession.
Requirements:	*Out-of-bounds*: The earliest moment that the ball is known to be going out of bounds. This is also applied to goal attempts shot from outside the Circle. *Foul committed*: The moment when the offending event appeared to occur. *Interception or steal*: The moment the defending player contacts the ball for the action that relieves the attacker of possession. A Turnover Conceded is not applied if a defender only deflects a pass or shot and the ball is subsequently controlled by an attacker.

3.2.2 Data processing

The digitised locations of players and the ball were defined within images' (u, v) coordinate systems (Figure 3.4, *Top*). These (u, v) coordinates needed converting into real-world (x, y) coordinates to have meaningful discussions about player locations on the pitch (Figure 3.4, *Bottom*). The fisheye lens used to record the 4K footage also distorted the images (Figure 3.1, *Top*), which required correcting.

Image removed due to copyright

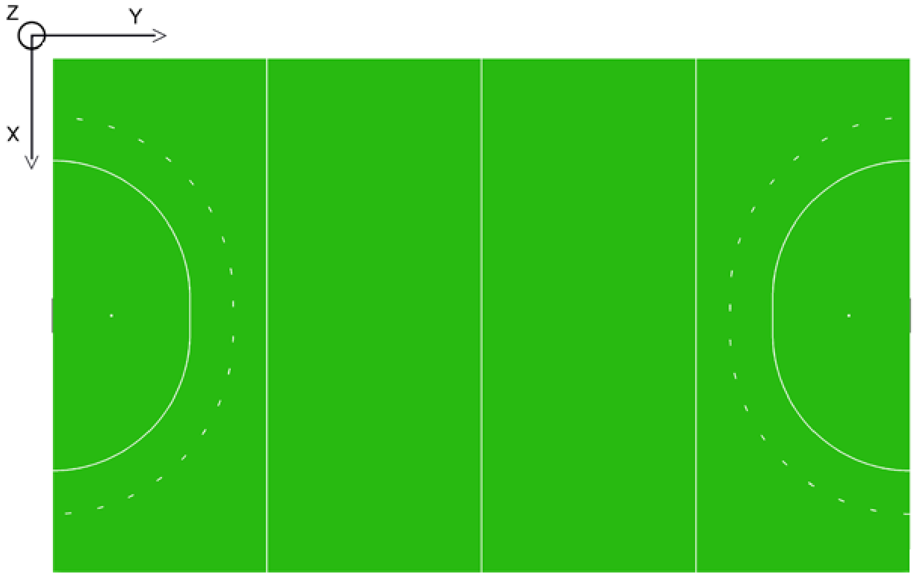


Figure 3.4 *Top*: The (u, v) coordinate system of the 4K camera image. *Bottom*: The (x, y, z) coordinate system of the field hockey pitch.

Digitised locations were converted into undistorted, real-world coordinates using the process shown in Figure 3.5.

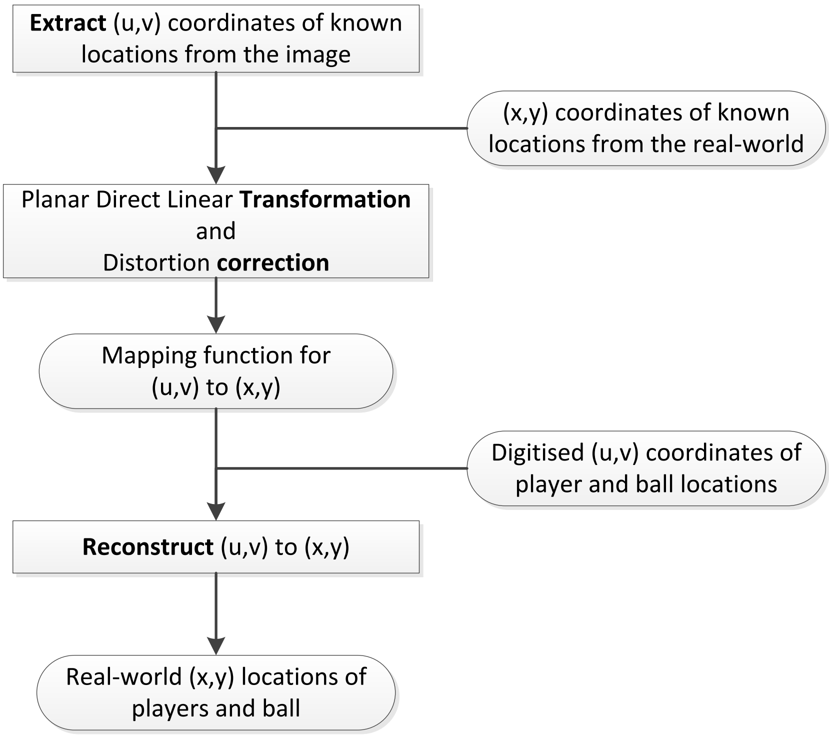


Figure 3.5 Flow diagram showing how digitised locations of players and the ball were converted in undistorted, real-world coordinates.

A real-world plane was defined using 14 known points where pitch lines intersect, in a real-world (x, y) coordinate system (Figure 3.6, *Top*). The locations of these known points within images' (u, v) coordinate systems were obtained using Photoshop CS6 (Adobe 2013) (Figure 3.6, *Bottom*). The (x, y) and (u, v) coordinates of these 14 known locations informed a planar Direct Linear Transformation (Abdel-Aziz and Karara, 1971; Walton, 1981) that could reconstruct images' (u, v) coordinates into real-world coordinates. As part of the reconstruction, the distortion caused by the fisheye lens was corrected using Bouguet's (2010) MATLAB calibration toolbox. The combined process provided a function that reconstructs distorted, digitised coordinates into undistorted, real-world coordinates. This function was applied to the digitised locations of players and the ball to provide coordinates of players and the ball on the pitch's surface.

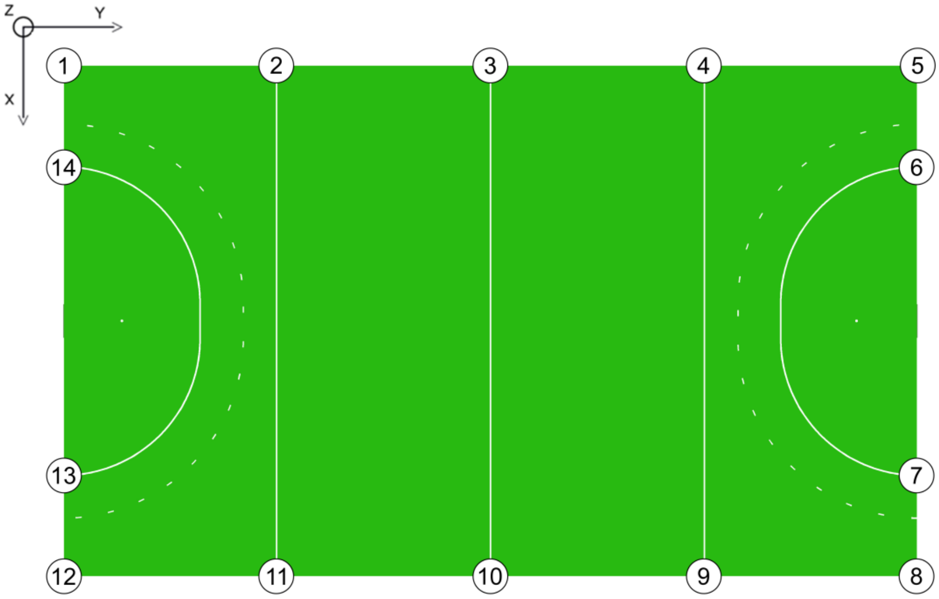


Image removed due to copyright

Figure 3.6 *Top*: The 14 known locations as they appear on an undistorted, real-world plane. *Bottom*: The 14 known locations as they appear in the distorted footage.

Further processing was required to reorient player and ball locations. The dataset contained plays travelling toward both goals depending on which team had possession. Player locations were reoriented such that all plays travelled in the same direction (Figure 3.7). To do this, the attacking team's goal was set at zero pitch-length and the defending goal was at the maximum pitch length. This meant that all plays attack toward larger values of pitch length. All figures presented in this chapter use a left-to-right convention of increasing pitch length.

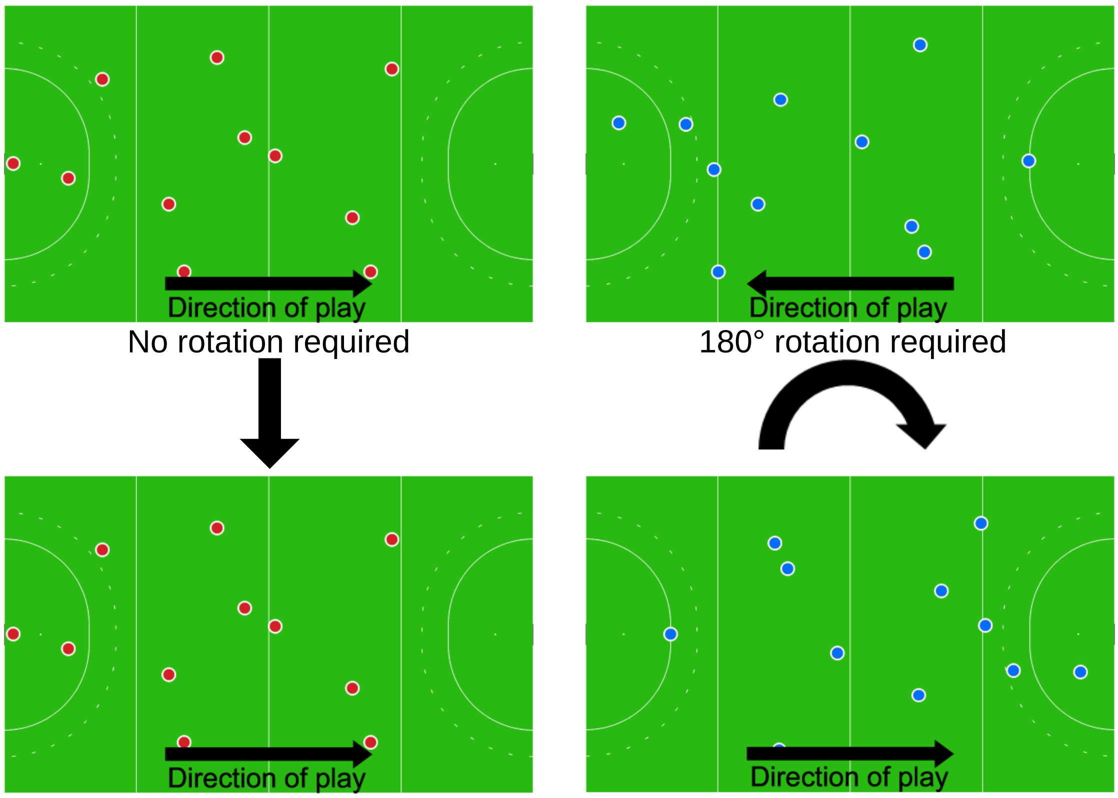


Figure 3.7 Reorienting player locations to align the direction of offensive play.

3.2.3 Data analysis

A test-retest design with one operator was used. This design requires a *test* dataset against which a *retest* dataset is compared. The *test* set contained all digitised plays and the *retest* set was a sample of the plays that were digitised again. The 94 variables constituted the timestamp of the 23 m intrusion and outcome event, and the reconstructed pitch-width and pitch-length locations of all players and the ball. For each observation in the *retest* set, the intrusion timestamp was paired with the closest match in the *test* set. The closest match was assumed to be the appropriate partner in the *test* set.

There was a period of three months between digitisation of the *test* and *retest* sets to minimise familiarisation. All digitising was done by one operator with the same software on the same machine in the same room.

Acceptable range of repeatability

The acceptable ranges for each data type were decided before digitisation (O'Donoghue 2007). The acceptable range for spatial data was ±0.5 m for the pitch-length and pitch-width reconstructed locations. The rationale for this range is that the expected estimated location of the player should lie within the 'footprint' of the player and a ±0.5 m range was considered a reasonable estimate considering the observed movements of standing, jumping, shuffling, jogging, sprinting and lunging. Pitch-length and pitch-width location components were considered separately because the effects of perspective and lens distortion differ between the axes.

The acceptable range for temporal data was 1 frame, which equated to ±0.04 s. The rationale is based on maximum expected distance that a player could travel in 0.04 s. Lidor and Ziv's (2015) review of elite field hockey performance demands suggested a maximum player speed of 9 $m \cdot s^{-1}$ = 32.5 $km \cdot h^{-1}$ (Konarski 2010). At this speed, a player could travel 0.36 m in 0.04 s, which lies within the ±0.5 m spatial window. Any more than a one-frame difference could indicate a maximum potential difference greater than 0.5 m.

Number of retest trials

Three plays from each of the 20 quarters were selected (n_{retest} = 60 = 9.1% of n_{plays}). It was assumed that selecting three plays from each quarter would sufficiently cover the sources of variance in image quality arising different camera settings and lighting conditions between and during games and quarters. The duration of each quarter was normalised between 0 and 1 and timestamps at 0.33 and 0.66 were computed. The soonest occurring play from the start of the quarter, from the 0.33 timestamp and from the 0.66 timestamp were digitised. If the timestamp was in the middle of a play, then that play would be digitised. If the same play was indicated for two selections, then the subsequent play was used for one of the digitisations. Only the locations and times of the intrusion and outcome events were digitised, i.e. passing events surrounding intrusions and outcomes were not considered.

Statistical analyses

Temporal data and spatial data were assessed separately using the median absolute error and the standard error of measurement (Brown 1999; Weir 2005). The median absolute error (MdAE) provides an indication of absolute consistency specific to the dataset used. It is more robust than the arithmetic mean absolute error and especially more representative for skewed distributions. The MdAE is the $\left(\frac{n_{digi}+1}{2}\right)^{th}$ absolute error in an ordered list of absolute errors, where n_{digi} is the number of locations digitised. Absolute error is:

$$\text{Absolute error} = \left|x_{test_i} - x_{retest_i}\right| \qquad [3.1]$$

For a test-retest design, the standard error of measurement (SEMt) can be calculated by:

$$\text{SEMt} = \sqrt{\frac{\sum\left(x_{test_i} - x_{retest_i}\right)^2}{2n}} \qquad [3.2]$$

where x_{test_i} and x_{retest_i} are the paired observations in the *test* and *retest* set, respectively (Bland and Altman, 1996), and n is the number observations being compared. The SEMt is also known as the within-subject standard deviation. It represents the standard deviation of observed scores around the unknown true score (Weir 2005). As such, it provides an indication of the absolute error that could be generalised to similar data digitised by this operator under similar conditions. Other statistics such as the intra-class correlation coefficient (Müller and Büttner, 1994; Shrout and Fleiss, 1979) and the coefficient of repeatability (Vaz et al. 2013; Beckerman et al. 2001) were not used because data were not normally distributed.

3.3 Results

Table 3.3 summarises the number of plays observed in the current work. For the 5 games included in the dataset, the median proportion of gameplay duration involving the standard complement of players was 83.3%. All three outcomes were similarly represented.

Table 3.3 Number of plays, n, observed in the current work.

	n
Number of plays observed	755
Number of 11-v-11 plays observed	660
Number of plays ending in a Circle Entry	238
Number of plays ending in an ‘Other’ outcome	211
Number of plays ending in a Turnover (conceded)	211

Table 3.4 presents the values of the repeatability statistics for all data types. Table 3.5 shows that spatial pitch-length data were consistently more repeatable than pitch-width for the ball or player locations.

Table 3.4 Summary of repeatability statistics. MdAE = median absolute error, SEMt = standard error of the measurement.

Data type	**MdAE**	**SEMt**
Temporal (s [# of frames])	0.04 [1.0]	0.40 [9.9]
$\text{Spatial}_{\text{pitch-width}}$ (m)	0.14	0.28
$\text{Spatial}_{\text{pitch-length}}$ (m)	0.10	0.26

Table 3.5 Repeatability of digitisation of the ball and players. MdAE = median absolute error, SEMt = standard error of the measurement.

Metric type	MdAE (m)		SEMt (m)	
	width	length	width	length
Ball location	0.10	0.09	0.30	0.19
Player location	0.14	0.10	0.28	0.26

3.4 Discussion

As mentioned in the literature review, the similar number of plays in all three outcome groups will provide balanced support for comparisons between the groups (Atkinson and Nevill, 2001). The effect of only considering plays with the standard, 11-v-11 complement of players is unknown. Situations where one team has more players than the other are sometimes called ‘power plays’. Goals have been shown to be more likely during power play situations in ice-hockey (Thomas 2007) but not in water polo (Lupo, Minganti, et al. 2012) so it is unclear how including power plays might have affected the number of Circle Entry outcomes. At over 87%, the number of plays involving the standard, 11-v-11 complement of players well-represents the total number of plays observed in the current work.

Both the median absolute error and the standard error of measurement of spatial data were within the desired threshold of repeatability for both pitch-width and pitch-length data. These results support the repeatability of digitising and reconstructing player locations. The trend for pitch-length values to be more repeatable than pitch-width values is likely because of the video camera's shallow-angle view of the pitch (26° at the near side and 6° at the far side). This meant that there were fewer pixels per unit area for the pitch-width when compared to the pitch-length. Any digitisation errors therefore carry a greater cost for the pitch-width, especially at the far-side of the pitch. Table 3.5 also shows that digitising the ball was more repeatable than digitising players, possibly because the ball presents a smaller range of pixels to choose from.

For temporal data, the median absolute error statistic suggested that observations were repeatable by 0.04 s [1 frame]. This equates to a 0.36 m maximum potential difference in player location assuming Konarski's (2010) maximum player speed of 9 $m \cdot s^{-1}$. This represents an estimated worst-case scenario because players spend most of their time at low- and moderate-intensity at speeds of < 3 $m \cdot s^{-1}$ (Lidor and Ziv 2015). At these lower speeds, the expected potential difference in player location would be closer to 0.12 m. The maximum and typical potential differences are below the predetermined threshold for repeatability of ±0.5 m, and therefore suggest good repeatability of the timing of gameplay events.

The standard error of measurement for temporal data was much larger than the predetermined threshold, equating to maximum potential difference in player location of 3.6 m. The typical potential differences based on Lidor and Ziv's (2015) estimate of < 3 $m \cdot s^{-1}$ is 1.2 m. These values are unacceptably large but this statistics' values are expected to be larger because of its generalisation. The sources of these results were two large differences where the assigned outcome differed between the *test* and *retest* set, i.e. different events were digitised in each case. This meant that the difference in the timing of the digitisations were abnormally large. With these outliers removed, the maximum and typical potential differences in player location are reduced to 0.63 m. The errors in temporal data are likely due to the difficulty in locating the ball. Examples of images that were difficult to digitise are shown in Figure 3.8.

Image removed due to copyright

Image removed due to copyright

Image removed due to copyright

Figure 3.8 Instances where the ball is not easily identifiable in a group of players.

3.5 Summary and Conclusion

A total of 775 plays were observed in 5 games of an elite, international, men's field hockey tournament. Only 660 plays were carried forward because they involved the standard, 11-v-11 complement of players. Player and ball locations were measured by manual digitisation using in-house software. Locations were only digitised at the start and end of plays. Definitions of player locations and plays were provided.

Sixty plays were re-digitised for comparison with the initial dataset in a test-retest design to assess repeatability. An acceptable error range of ±0.5 m was suggested based on players' expected 'footprint' during gameplay actions and the expected speed of players during gameplay. Both spatial and temporal data were deemed repeatable. The conflicting result of the standard error of measurement for the temporal data was attributed to atypical errors. The first objective of the current work was met. The next chapter will discuss the spatio-temporal metrics that make use of the player location measurements.

Chapter 4 Preparing metric data for analysis

4.1 Introduction

The literature review identified the many spatio-temporal metrics that can describe players' behaviour (Section 2.5). The spatio-temporal data collected in the previous chapter constrained the possible number of metrics that could be computed because it only measured locations at the start and end of a play. Table 4.1 shows the number of metrics used in the current work and how many were associated with each of the eight metric types identified in section 2.5.2. For metrics that required information about player roles, player roles were determined using Clemente et al.'s (2015) method that was discussed in the literature review (Section 2.5.3 *Player groups*). Individual metrics will be discussed in the following chapters whenever appropriate.

Table 4.1 Number of metrics used in the current work.

Metric type	Number of metrics
Location	132
Time	7
Distance	2,260
Speed	8
Angle	156
Spread	746
Area	300
Context	43
Total	3,652

The literature review also identified the Backward Dropping Algorithm (Wang et al. 2012) as an appropriate method to determine which of these metrics and metric combinations distinguish play outcomes (Section 2.6.4). Two requirements of the method needed to be addressed before analysis: a binary outcome variable and discrete inputs.

The Backward Dropping Algorithm was designed to handle a binary outcome variable. This is in part due to its use in genetic analytics, where the presence or absence of a trait is of interest (S.-H. Lo and Zheng, 2004). There are three outcomes of interest in the current work: a Circle Entry, a Turnover (conceded) and an 'Other' outcome (Section 2.3.1). For the outcome of the algorithm to be binary, it was necessary to pair the outcomes as three comparisons:

1. Circle Entry and Turnover (conceded)
2. Circle Entry and 'Other'
3. 'Other' and Turnover (conceded)

The first named outcome in the comparisons is the preferred outcome based on the 'positive'-'partially positive'-'negative' paradigm discussed in the literature review (Section 2.3.1). Each of these independent comparisons will be investigated in the current work partly because there is no evidence to suggest that the same metrics will distinguish each pair of outcomes.

The Backward Dropping Algorithm also requires discrete input data, i.e. data values are one of a fixed set. Unfortunately, the variety of spatio-temporal metrics computed were mostly continuous, i.e. values that are measured along a continuum. For example, a player's pitch-length location is not restricted to being 1 m or 2 m from the end of the pitch. Instead, it can be at any infinitely fine location between the ends of the pitch and limited only by the precision of the measurement tool.

Metrics' data therefore had to be transformed into discrete values. Two approaches were considered for this task. The first was to consider each metric individually and logically determine appropriate categories for discretisation. For example, when considering a player's pitch-length location, data could be assigned to one of four values representing each quarter of the pitch. Unfortunately, not all metrics lend themselves to such logical discretisation. For example, how should one discretise the range of values for the surface area of a team? Furthermore, discretising the pitch-length location of a player into the four quarters of the pitch is only useful if players have an equal likelihood of being in any quarter. Consider the goalkeeper: since the goalkeeper is not permitted to leave their quarter of the pitch (FIH, 2015, Section 10.1), it would be useless to consider the other quarters of the pitch to describe their location. All values would relate to their quarter of the pitch and the goalkeeper's pitch-length location would be uninformative. Furthermore, discretising each of the 3,652 metrics individually would also be prohibitively time consuming.

An alternative approach was to apply an automatic data-partitioning algorithm to discretise each metric's data. This empirical approach could take the actual distribution of the metric's data into account rather than assuming any distribution.

This chapter address the second of the project's objectives: to prepare the spatio-temporal metric data for analysis. The following sections explain how continuous metrics were discretised using the *k*-means++ algorithm and how already-discrete metrics with too many categories were handled. Although all data and processing scripts were checked for accuracy, some errors were evident following the discretisation. These errors are discussed alongside the solutions applied to correct them.

4.2 Method

4.2.1 Continuous metrics

Each continuous metric was discretised using a *k*-means++ algorithm that used the squared Euclidean distance statistic. Data from all $n_{plays} = 660$ plays were used. The algorithm used two or three clusters based on the best results from 20 iterations of a silhouette analysis. In this section, the *k*-means++ algorithm and silhouette analysis will be explained briefly.

The *k*-means++ algorithm is an unsupervised clustering algorithm that assigns observations to one of *k* clusters, where each cluster is defined by a centroid (Arthur and Vassilvitskii, 2007). The user inputs a *k* number of 'seeds' as initial cluster centroids, which are then iteratively updated to optimise some cost function. The squared Euclidean distance from observation to centroid was used as the cost function, in the current work. The *k*-means++ algorithm betters the original *k*-means algorithm by improving the seeding of clusters. Whereas *k*-means seeds by uniformly randomly sampling *k* observations, *k*-means++ only uniformly samples for the first seed, at random. All remaining seeds are randomly sampled from the set of observations with a probability proportional to an observation's squared distance from its closest seed. This difference improves convergence time and final error because centroids are expected to be as dispersed as possible.

The number of seeds was constrained to be either two or three, for two reasons. Firstly, the greater the number of discrete values that inputs take, the more computationally expensive the forthcoming analysis method is. Therefore, fewer discrete values are preferred. Secondly, the current work has been conducted in collaboration with England Hockey and communication of findings has always been an important consideration. It was decided that discussing metric values in terms of {Large, Small} or {Large, Moderate, Small} would be useful. The clustering algorithm will provide such discretisation for each metric specifically. Whether two or three clusters were used was determined by a silhouette analysis.

A silhouette analysis indicates how well a data point is suited to its cluster rather than other clusters (Rousseeuw 1987). The arithmetic mean of data points' silhouette scores indicates how distinct the suggested clusters are, with larger values being preferred. The chosen discretisation was the best of 20 iterations of a silhouette analysis for two and three seed initialisations (40 silhouette analyses in total). Both the *k*-means++ and the silhouette analysis were performed using MATLAB R2015b (MathWorks 2015).

For some metrics, it was meaningful to say that a value did not exist rather than assign a value of zero. For example, the length of pass was only measured if a pass was associated with an event. For situations where the outcome was not a pass, a pass length of zero would imply that a very small pass was used, given the metric's margin of error. An additional discrete value was used to represent these situations when it was meaningful for a value to be non-existent. This would mean, for example, that the clusters of pass length could be Large, Moderate, Small or Null.

4.2.2 Discrete metrics

Some metrics were already of a discrete level of measurement and thus suitable for the main analysis. The complexity and runtime of the main analysis increases with the number of categories that the input variables take (Wang et al. 2012). It is preferred that variables have as small a range of categories as possible but at least two. Practically, it is the range of values expressed in the data that will affect the analysis. If a metric had a possible range of six categories but only three were expressed in the dataset, then the effective range of categories would only be three. Most metrics expressed two categories with three and four categories ranking next most frequent. Only the lateral pitch transition type metric expressed more than four categories.

The lateral pitch transition type is a nominal variable indicating how the play transitioned between corridors on the pitch, from intrusion to outcome. The corridors are left-of-pitch, centre and right-of-pitch. The possible options are shown in Figure 4.1.

		End of play		
		Left	*Centre*	*Right*
Start of play	*Left*	Left stay Left	Left to Centre	Left to Right
	Centre	Centre to Left	Centre stay Centre	Centre to Right
	Right	Right to Left	Right to Centre	Right stay Right

Figure 4.1 All nine, possible lateral pitch-transition types.

There are 9 values but only a range of 8 were observed in outcome comparisons. To reduce the number of categories, the values were grouped according to the colour scheme seen in the Figure 4.1. The white values are instances where there was no transition. The light grey values are instances where play transitioned to an adjacent corridor. The dark grey values are instances where play transitioned from one side of the pitch to the other. This reduced the range of categories to three.

4.3 Results

For a given comparison's data, most metrics were discretised into two clusters (Table 4.2). Metrics with four clusters included a null value to represent situations when it was meaningful for a value to be non-existent. A small number of metrics presented with only one cluster.

Table 4.2 The tally of metrics having 1, 2, 3 or 4 clusters per comparison.

Comparison	Number of clusters	Tally of metrics	Proportion of metrics
Circle Entry-Turnover	**1**	21	0.01
	2	2,829	0.77
	3	711	0.20
	4	91	0.03
Circle Entry-Other	**1**	39	0.01
	2	2,863	0.78
	3	665	0.18
	4	85	0.02
Other-Turnover	**1**	36	0.01
	2	2,772	0.76
	3	744	0.20
	4	100	0.03

4.4 Discussion

The trend for fewer clusters was beneficial for the forthcoming analysis because the number of computations increases with the number of values that inputs can take. The presence of metrics that have only one cluster in a comparison's dataset is problematic because they could be outliers or legitimately single-clustered metrics. The discretisation forced either two or three clusters upon each metric's data. If there was an outlier, then the *k*-means++ algorithm might assign it to one cluster and all remaining data to a single other cluster (Figure 4.2). Given that the outlier can only belong to one outcome group, the metric would be minimally informative because both groups would be identical except for the one observation.

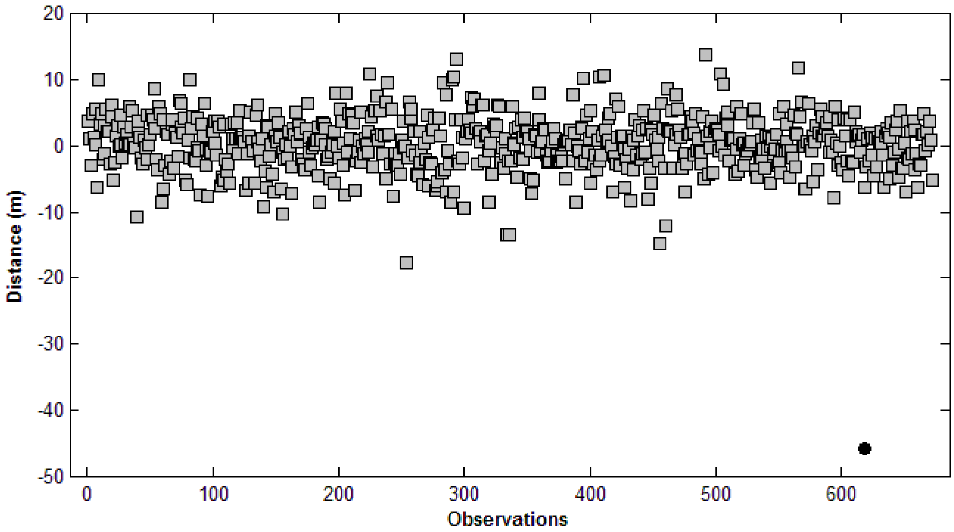


Figure 4.2 Scatter plot of data relating to the Euclidean distance between the centroid location of the offensive forwards and defensive backs at the start of the play. The black circle data point at the bottom of the plot is an outlier that formed its own cluster during the discretisation process.

This data are categorised very differently when the outlier is corrected (Figure 4.3).

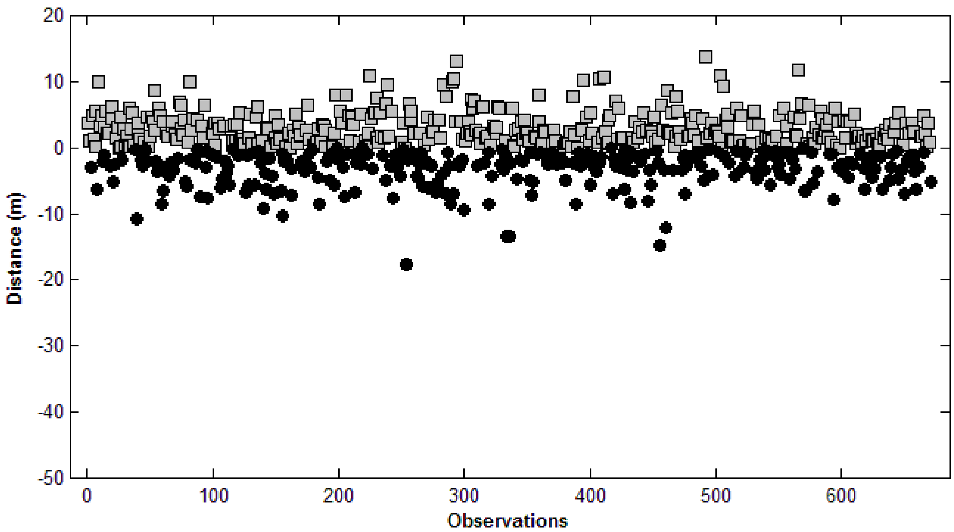


Figure 4.3 Scatter plot of data relating to the Euclidean distance between the centroid location of the offensive forwards and defensive backs at the start of the play. The outlier from Figure 4.2 has been corrected.

The previous example of spurious data was indicated by a metric having only one cluster. This occurred for 58 metrics, which was suspicious because even metrics with very low variance would still be expected to have a more balanced discretisation. Having only one cluster within a comparison's dataset might not be because of errors in the data. A metric could be legitimately single-clustered for the dataset used in the current work. Such metrics can be removed because they are uninformative. The 58 metrics that presented one cluster within a comparison's dataset were investigated and four possible explanations were concluded.

The first explanation for unbalanced discretisation is an extreme value in the dataset that has formed its own cluster during discretisation. This was the case in the previous example and was seen in four of the 58 metrics. A review of player location data revealed digitisation errors that were corrected.

A second explanation is that some observations formed a distinct outlying cluster away from the majority of data points. A metric that was a candidate for this explanation was the arithmetic mean of the defensive team's Euclidean distance to the ball at the start of a play. Figure 4.4 shows that three observations are outliers of the main dataset but they are all within a small range.

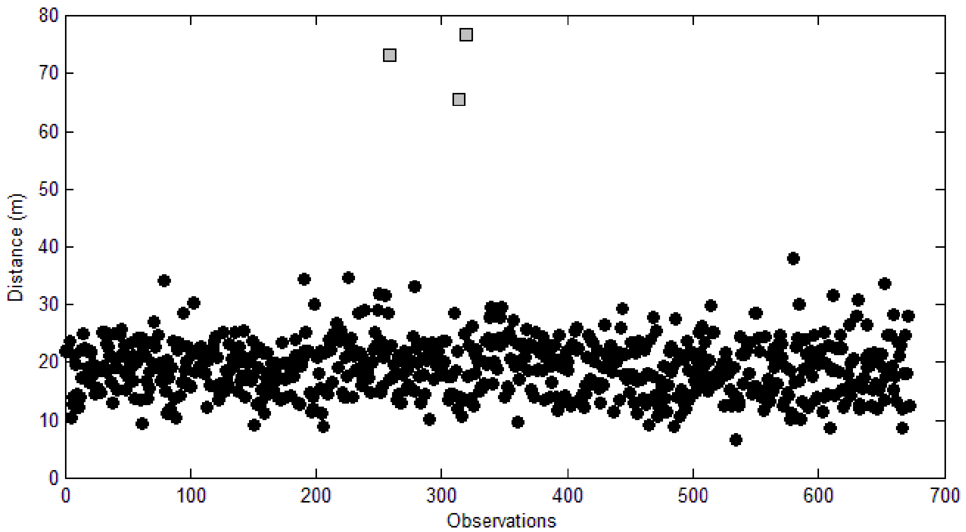


Figure 4.4 Scatter plot of data relating to the arithmetic mean of the defensive team's Euclidean distance to the ball at the start of a play. The grey square data points at the top of the plot possibly belong to a distinct outlying cluster.

It is possible that this outlying cluster has been correctly discretised into its own cluster. It might also be the case that these observations are all errors that fall within a small range. These same three observations were responsible for 18 of the 58 metrics that had only one cluster within a comparison's dataset. A review of player location data revealed digitisation errors that accounted for these outlying observations. These were subsequently corrected.

The third explanation for a metric having only one cluster is that there were errors in the MATLAB scripts that calculated the metric's value. This was the case for Offensive Numerical Superiority, also known as Territorial Dominance (Clemente et al., 2014). The metric can have three values that indicate the ratio of offensive and defensive players in 12 regions of the pitch. Comparisons between the computed values and illustrations of player locations suggested that the calculations might be incorrect. Inspection of the MATLAB script revealed an error. The error was corrected, which accounted for 26 of the 58 spurious metrics.

The final explanation for a metric having only one cluster is that all the metric's data were legitimately single-clustered. This was common for those metrics that did not need discretisation because they were nominal variables to begin with. Their design was such that, for the given set of observations, most cases were only one nominal value.

An exemplar metric would be Tenga's (2010) Defensive Pressure, which was inspired by Olsen et al. (1994). It indicates the ratio of defenders within and outside a 1.5 m radius around the possessing player. For all observations, Defensive Pressure at the start of plays had a value of 2, indicating that more players were outside of a 1.5 m radius of the possessing player than within it. It is for metrics like this that a legitimate single cluster is not considered spurious. The rationale is that an offensive player would be unlikely to have had possession if five or more defenders were within 1.5 m of them, given the reach afforded to the defenders by their sticks. Tenga's (2010) Defensive Pressure was developed for association football where a defensive player's radius of influence would only be as far as their body could reach, which suggests that adjustments might need to be made to this metric to appropriately apply it to other sports. This was not done in the current work.

Legitimate single clusters were also found for continuous metrics, such as the proportion of overlap between the rear regions of both teams (see Defensive Back Region, Clemente et al., 2015). The rear regions of each team are, by definition, on opposite ends of the pitch. Player locations were visualised using illustrations similar to those used in chapter 2. These illustrations confirmed that there were no observations for which the rear regions of the teams overlapped, thus the metric's values were always zero.

Metrics like these, whose data is entirely and legitimately single-clustered, were removed from the dataset because they provide no information for distinguishing outcome groups. Table 4.3 shows which metrics were removed from each comparison dataset.

Table 4.3 Metrics removed because they were legitimately single clustered and therefore uninformative for distinguishing between outcomes.

	Comparisons		
Metrics	**CEvTOc**	**CEvOther**	**OthervTOc**
Defensive Pressure at the intrusion	x	x	x
Defensive Pressure at the outcome	x	x	x
Defensive Back-Up at the intrusion	x	x	x
Defensive Back-Up at the outcome	x	x	x
Defensive Cover at the intrusion	x	x	x
Defensive Cover at the outcome	x	x	x
Proportion of the Offensive Back Region overlapped by the Defensive Back Region, at the intrusion	x	x	x
Proportion of the Offensive Back Region overlapped by the Defensive Back Region, at the outcome	x	x	x
Proportion of the Defensive Back Region overlapped by the Offensive Back Region, at the intrusion	x	x	x
Proportion of the Defensive Back Region overlapped by the Offensive Back Region, at the outcome	x	x	x
Proportion of the Defensive 1st Half Middle Region overlapped by the Offensive Back Region, at the outcome	-	x	-
Numerical superiority of the front-centre region, at the outcome	x	-	-
Numerical superiority of the back-left region, at the intrusion	-	-	x
Numerical superiority of the back-left region, at the outcome	-	-	x
Numerical superiority of the back-centre region, at the intrusion	-	-	x

CEvTOc = Circle Entry-Turnover comparison.
CEvOther = Circle Entry-Other comparison.
OthervTOc = Other -Turnover comparison.

Table 4.4 is an updated version of Table 4.3 that considers the corrections made and the metrics that were removed.

Table 4.4 The tally of metrics having 2, 3 or 4 clusters per comparison.

Comparison	Number of clusters	Tally of metrics	Proportion of metrics
Circle Entry-Turnover	**2**	2,936	0.81
	3	633	0.17
	4	72	0.02
Circle Entry-Other	**2**	2,889	0.79
	3	675	0.19
	4	77	0.02
Other-Turnover	**2**	2,886	0.79
	3	675	0.19
	4	78	0.02

The minimum number of clusters became two but the relative proportion of metrics with each range of clusters remained the same. The removal of legitimately single-clustered metrics also meant that 3,641 metrics were carried forward for the Circle Entry-Turnover comparison and Circle Entry-Other comparison. Only 3,639 metrics were carried forward for the Other-Turnover comparison.

4.4 Summary and conclusion

The purpose of this chapter was to prepare the spatio-temporal metric data for the subsequent analysis. Datasets associated with the three outcome groups were paired to suit the Backward Dropping Algorithm's structure for two-group comparisons. The Backward Dropping Algorithm also requires discrete inputs. The *k*-means++ algorithm was used to discretise continuous metrics. Metrics were discretised mostly with a range of two clusters, which was four times more likely than a range of three clusters and over 37 times more likely than a range of four clusters. The predominance of the two-cluster range was beneficial because smaller cluster ranges reduce the number of processes required in the subsequent analysis.

The discretisation process flagged possibly spurious data for 58 metrics. An investigation revealed four explanations for the possibly spurious data. Errors were corrected and some metrics were removed from comparisons' datasets because they were uninformative. Table 4.5 shows the updated count of metrics associated with each comparison after uninformative metrics were removed. These remaining metrics were carried forward to the first stage of the Backward Dropping Algorithm, which is presented in the next chapter.

Table 4.5 Number of metrics used in the current work after uninformative metrics were removed.

	Number of metrics			
Metric type	**Initial total**	**CEvTOc**	**CEvOther**	**OthervTOc**
Location	132	132	132	132
Time	7	7	7	7
Distance	2,260	2,260	2,260	2,260
Speed	8	8	8	8
Angle	156	156	156	156
Spread	746	746	746	746
Area	300	296	295	296
Context	43	36	37	34
Total	3,652	3,641	3,641	3,639

CEvTOc = Circle Entry-Turnover comparison.
CEvOther = Circle Entry-Other comparison.
OthervTOc = Other-Turnover comparison.

Chapter 5 Selecting metrics and metric combinations. Stage 1: The *I*-score and metric pre-selection

5.1 Introduction

This chapter addresses the third of the current work's objectives: to reduce the number of metrics based on performance in low-order interactions. The order of the interaction refers to the possible number of interactions between a given set of inputs. A small set of inputs only demonstrates a low order of interaction while a larger set demonstrates a higher order of interaction. The objective to reduce the number of metrics is the first of two stages of Wang et al.'s (2012) analysis that uses the Backward Dropping Algorithm. This first stage reduces the initially large number of metrics to a more manageable amount based on performance in pairs.

The first phase of this metric pre-selection procedure uses the *I*-score to rank metric pairs. The *I*-score measures the association of an input or set of inputs with the outcome variable (Chernoff et al. 2009). It is the squared-sum of deviations between the count of actual and expected outcomes (more detail is available in Appendix A). The *I*-score is the cornerstone of the Backward Dropping Algorithm because its expected value does not change with the number of inputs being considered (Wang et al. 2012). In other words, it is not affected by the degrees of freedom of the problem. The *I*-score for combinations of inputs of different sizes can therefore be compared.

Another advantage of the *I*-score is that it measures the association between an outcome variable and one *or more* inputs. This is superior to a Pearson's correlation that can only measure the association between an outcome variable and *one* input. Ideally, every possible combination of inputs would be scored using the *I*-score to determine which is best. Unfortunately, the number of possible combinations is prohibitively large when the number of inputs is already large. For example, a set of 1,000 inputs has 499,500 possible pairs and 166,167,000 possible triplets.

Wang et al. (2012) suggest that an indication of an input's combinatory influence can be estimated by evaluating a sample of the possible interactions - preferably low-order interactions so as not to stress computational resources.

This was based on the assumption that those inputs that score well at low-order interactions, like pairs and triplets, are likely to score well at higher-order interactions. Practically, low-order interactions might also be preferred over complex combinations that might be difficult to interpret and apply. Once all low-order combinations of inputs have been scored, only a portion are carried forward to the next phase of the pre-selection procedure. To do this, a threshold is applied to the pairs' ***I***-scores.

The second phase of the pre-selection procedure ranks inputs based on how often they occur in the combinations selected from the first phase. This statistic is known as the retention frequency, f_r. The rationale is that frequently retained inputs from the top-ranked, low-order combinations are likely to form influential, higher-order combinations because they frequently yield large ***I***-scores when combined with other inputs (Wang et al., 2012). Only a portion of these frequently retained inputs are carried forward for subsequent analyses by applying a threshold to the retention frequencies. The following sections present the application of this metric pre-selection procedure to the spatio-temporal metrics collected in the current work.

5.2 Method

The variable pre-selection procedure of Wang et al. (2012) was applied to each of the three outcome group comparisons using metric pairs. The Other-Turnover comparison had 3,639 metrics and so produced 6,619,341 metric pairs. The Circle Entry-Other and Other-Turnover comparisons had 3,641 metrics, producing 6,626,620 metric pairs. To select which metrics would be carried forward, a threshold ***I***-score was determined using the 2nd difference method (Wang et al. 2012). The protocol for the 2nd difference method is as follows:

1. Sort all N-number of variable combinations in descending order of ***I***-score.
2. Select every p^{th} ***I***-score, where p is some fraction of N so that the trend of p^{th} ordered ***I***-scores is representative of the entire list (No guidance is

provided for choosing an appropriate value for p. A value of $p = 1{,}000$ was used in the current work.)

3. Calculate the 1st differences, $d_{1\,i} : i \in \{1, \dots, N-1\}$, between every p^{th} value by calculating the difference between successive p^{th} values:

$$\left\{d_{1\,1}, d_{1\,2}, \dots, d_{1\,\frac{N}{p}-1}\right\} = \left\{p_1 - p_2, p_2 - p_3, \dots, p_{\frac{N}{p}-1} - p_{\frac{N}{p}}\right\}$$

   where p_j is the p^{th} *I*-score for $j \in \left\{1, \dots, \frac{N}{p}\right\}$.

4. Calculate the 2nd differences, $d_{2\,k} : k \in \left\{1, \dots, \frac{N}{p} - 2\right\}$, by calculating the difference between every successive 1st difference:

$$\left\{d_{2\,1}, d_{2\,2}, \dots, d_{2\,\frac{N}{p}-2}\right\} = \left\{d_{1\,1} - d_{1\,2}, d_{1\,2} - d_{1\,3}, \dots, d_{1\,\frac{N}{p}-2} - d_{1\,\frac{N}{p}-1}\right\}$$

5. Plot the 2nd differences with respect to the p^{th} value, which is expected to follow an 'L-shape' curve (Figure 5.1).

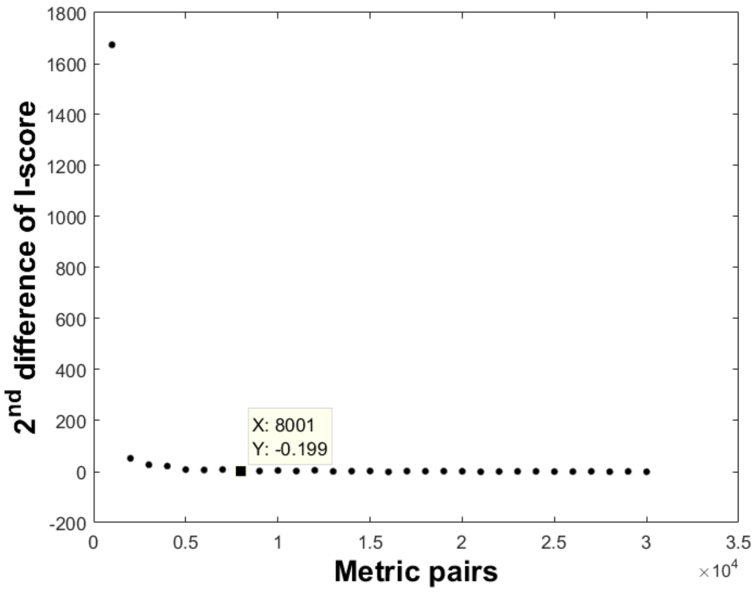


Figure 5.1 The 2nd difference plot for the comparison between plays that ended with a Circle Entry and those that ended with a Turnover Conceded.

6. Select a cut-off threshold, t_0, where the 2nd difference is near zero for the first time as the plot settles. (the first 2nd difference that was ≤1 was used - see Figure 5.1). Record the associated p^{th} value.

7. Select the top $t_0 - 1$ variable combinations from the ordered list. In Figure 5.1, $t_0 = 8{,}001$ so the top 8,000 metric pairs are carried forward.

Metrics in the metric pairs that satisfied the threshold were ranked according to retention frequency. Thresholds for the 1st difference method were applied to select the metrics to be carried forward to the next stage of analysis. A threshold is determined using the 1st difference method (Wang et al. 2012):

1. Sort all, N, variables in descending order of retention frequency, $f_{r\,i}\ , i \in \left\{1, \dots, \frac{N}{p}\right\}$.

2. Calculate the 1st differences, $D_{1\,j} : j \in \left\{1, \dots, \frac{N}{p} - 1\right\}$, by calculating the difference between successive retention frequencies:

$$\left\{D_{1\,1}, D_{1\,2}, \dots, D_{1\,\frac{N}{p}-1}\right\} = \left\{f_{r\,1} - f_{r\,2}, f_{r\,2} - f_{r\,3}, \dots, f_{r\,\frac{N}{p}-1} - f_{r\,\frac{N}{p}}\right\}$$

3. Plot the 1st differences, which is expected to follow an 'L-shape' curve (Figure 5.2).

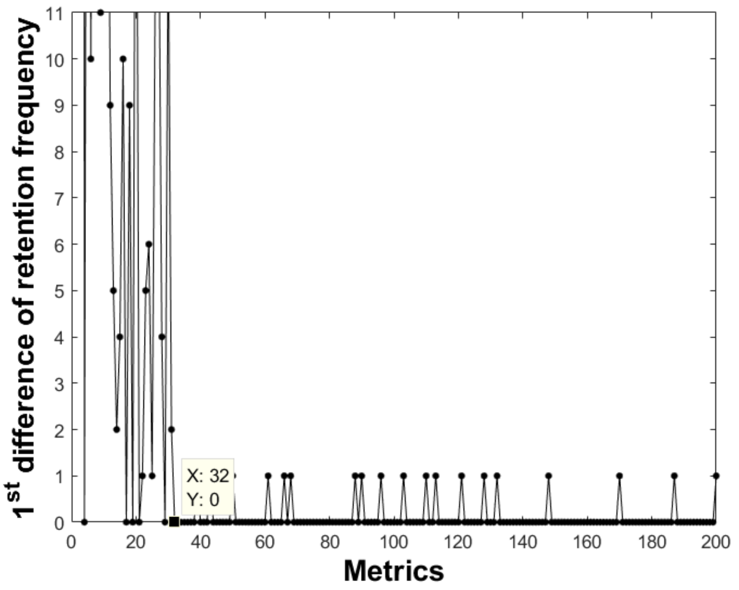


Figure 5.2 The 1st difference plot for the comparison between plays that ended with a Circle Entry and those that ended with a Turnover Conceded.

4. Select a cut-off threshold where the 1st differences "*differ little...* [and]*... retention frequency ties (1st-difference zeros) occur much more frequently*" (Wang et al., 2012). Record the associated threshold number of metrics, t_0'.

5. Select the top t_0' variable tuples from the ordered list. In Figure 5.2, $t_0' = 32$ so the top 32 metrics are carried forward.

Communications with the authors of Wang et al. (2012) revealed that "*rigorous theoretical justification of the* [1st difference and] *2nd difference method is still an open issue*" (Appendix C). The method is only supported by "*empirical and heuristic evidence*" from the authors' experiences, with no supporting data. Similar to Wang et al. (2012), the effect of both threshold choices on the final list of surviving metrics was therefore assessed. Datasets representing ±10% of the metric-pair threshold, t_0, were selected:

- the dataset based on the initial value of t_0 was called Initial;
- the dataset based on +10% of t_0 was called Plus;
- the dataset based on -10% of t_0 was called Minus.

For each of these metric-pair datasets, the ±10% method was applied to the retention-frequency thresholds. This provided the nine datasets shown in Table 5.1

Table 5.1 The nine datasets created during the metric pre-selection procedure.

		Retention-frequency threshold, t'		
		Initial (t'_0)	**Plus ($t'_{+10\%}$)**	**Minus ($t'_{-10\%}$)**
Metric-pair threshold, t	**Initial (t_0)**	Initial Initial (t_0 t'_0)	Initial Plus (t_0 $t'_{+10\%}$)	Initial Minus (t_0 $t'_{-10\%}$)
	Plus ($t_{+10\%}$)	Plus Initial ($t_{+10\%}$ t'_0)	Plus Plus ($t_{+10\%}$ $t'_{+10\%}$)	Plus Minus ($t_{+10\%}$ $t'_{-10\%}$)
	Minus ($t_{-10\%}$)	Minus Initial ($t_{-10\%}$ t'_0)	Minus Plus ($t_{-10\%}$ $t'_{+10\%}$)	Minus Minus ($t_{-10\%}$ $t'_{-10\%}$)

The effect of the threshold will only be understood when the full analysis is complete. Differences in the datasets after this stage will affect the metrics and metric combinations that constitute the final output. Datasets that differ must therefore be carried forward through the full analysis to estimate the effect of threshold choices.

5.3 Results

The highest- and lowest-scoring pair had *I*-scores of 9,817.4 and 0.6, respectively. Table 5.2 summarises the results of applying the nine thresholds.

Table 5.2 The results of the two stages of metric pre-selection. Rows indicate the effect of applying the metric-pair threshold. The rightmost three columns show the effect of applying the retention-frequency threshold.

Comparison	Metric-pair thresholds	Number of metric pairs selected	Number of metrics selected		
			Initial	Plus	Minus
CEvTOc	**Initial**	8,000	32	35	29
	Plus	8,800	32	35	29
	Minus	7,200	31	34	28
CEvOther	**Initial**	11,000	12	13	11
	Plus	12,100	12	13	11
	Minus	9,900	12	13	11
OthervTOc	**Initial**	13,000	27	30	24
	Plus	14,300	37	41	33
	Minus	11,700	35	39	31

CEvTOc = Circle Entry-Turnover comparison.
CEvOther = Circle Entry-Other comparison.
OthervTOc = Other-Turnover comparison.

For all comparisons, the metric selections included less than 1.2% of the original metrics (Table 5.3, 5.4, 5.5). Some metric types are no longer represented in comparisons' selections. For example, there are no longer any Time type or Angle type metrics associated with the Circle Entry-Turnover comparison. Distance type metrics maintain their rank with the largest representation even after suffering the largest loss. Although not evident from the tables provided, metrics relating to the difference between teams demonstrated the largest drop in representation; those selected referred to the offence or defence alone. Metrics relating to the outcome event now dominate the selection with the defensive team being completely unrepresented.

Table 5.3 Number of metrics selected across all Circle Entry-Turnover datasets, per metric type.

Metric type	Initial no. of metrics	No. of metrics selected	Percentage of possible metrics selected (%)
Location	132	6	4.5
Time	7	0	0.0
Distance	2,260	14	0.6
Speed	8	1	12.5
Angle	156	0	0.0
Spread	746	14	1.9
Area	296	3	1.0
Context	36	4	11.1
Total	3,641	42	1.2

Table 5.4 Number of metrics selected across all Circle Entry-Other datasets, per metric type.

Metric type	Initial no. of metrics	No. of metrics selected	Percentage of possible metrics selected (%)
Location	132	4	3.0
Time	7	0	0.0
Distance	2,260	8	0.4
Speed	8	0	0.0
Angle	156	1	0.6
Spread	746	0	0.0
Area	295	0	0.0
Context	37	0	0.0
Total	3,641	13	0.4

Table 5.5 Number of metrics selected across all Other-Turnover datasets, per metric type.

Metric type	Initial no. of metrics	No. of metrics selected	Percentage of possible metrics selected (%)
Location	132	5	3.8
Time	7	0	0.0
Distance	2,260	27	1.2
Speed	8	1	12.5
Angle	156	6	3.8
Spread	746	0	0.0
Area	296	2	0.7
Context	34	0	0.0
Total	3,639	42	1.1

5.4 Discussion

The application of the metric-pair and retention-frequency thresholds substantially reduced the number of metrics, in line with the expectations of Wang et al. (2012). The substantial > 98% reduction could be for three reasons. Firstly, it could be a legitimate reduction because many of the metrics truly are not influential for the given dataset of field hockey performance. This might not be surprising given that the original list of metrics covered a large variety of simple and complex measurements. The current work took a naïve approach by not assuming that any metric was more or less likely to be influential. Instead, the initial list of metrics contained all those that have been used to investigate sports similar to field hockey (Section 2.5.1). But, as the sociologist William Bruce Cameron once said, "*not everything that can be counted counts, and not everything that counts can be counted*" (Cameron, 1963, p 13).

Secondly, the substantial reduction could be because there are, actually, many influential metrics. This would violate an assumption of the method, which expects there to be only a small number of influential variables (Wang et al., 2012). The 1st difference and 2nd difference methods depend on the *I*-scores and retention frequencies following an 'L-shaped' curve similar to an exponential distribution, i.e. a small set of high-scores followed by a precipitous drop to many low-scores. Because the 1st difference and 2nd difference methods look at differences between successive scores, the chosen thresholds are expected to represent the beginning of a flat-lining of the many low-scores. If, on the other hand, there are many high-scores, then the 1st difference and 2nd difference methods might indicate a threshold that represents the beginning of a plateau of high-scores. This scenario would result in a conservative selection of only the highest of high-scores and/or an arbitrary portion of equally high scores. The *I*-scores for the metric pairs and the retention frequencies of the constituent metrics both followed the expected 'L-shape' (Figure 5.3). It is therefore not likely that there were many influential variables and, hence, the substantial reduction is likely to be legitimate.

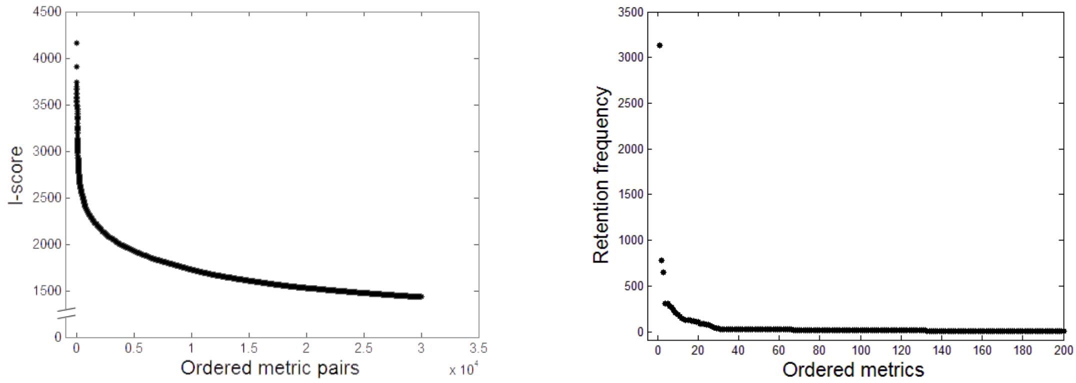


Figure 5.3 *Left*: The *I*-score of the top 30,000 metric pairs. *Right*: The retention frequencies of the top 200 metrics.

A third reason for the substantial reduction in metrics might be that a metric-pair interaction was too low an order of interaction to give metrics a chance of scoring well. This could cause metrics to be illegitimately selected because this stage of the analysis assumes that a metric's performance in low-order interactions is indicative of the metric's performance in higher-order interactions. This assumption cannot be checked without assessing the deselected metrics in higher-order interactions. Some of these metrics might strongly contribute to metric combinations that score well in higher-order interactions. While in the next chapter, the selected metrics will be assessed in higher-order interactions, the deselected metrics will be given a second chance during a later stage of the analysis.

A closer look at the three comparisons shows the effect of applying the thresholds to the metric pairs' *I*-scores and the retention frequency. The reader is reminded that only unique sets of metrics need to be carried forward for further analysis. First, we consider the Circle Entry-Turnover comparison. The effect of applying the metric-pair threshold was examined by comparing the metrics found in the three datasets that it created (Initial-Initial, Plus-Initial and Minus-Initial in Table 5.2). The uniqueness of these meant that they were all carried forward to investigate the effect of applying the retention-frequency threshold. The effect of applying the retention-frequency threshold was examined by comparing all nine datasets. All nine datasets were unique so they all were carried forward for subsequent analysis. For the Circle Entry-Other comparison, all datasets arising from the metric-pair threshold were identical. Applying the retention-frequency threshold produced unique datasets so all three were carried forward for subsequent analyses. For the Other-Turnover comparison, the datasets arising from the metric-pair threshold were all different and further differences were apparent after considering the effect of the retention-frequency threshold. All nine datasets were carried forward for subsequent analyses.

The procedure overwhelmingly selected metrics describing intra-team distances of the offence at the outcome. This is despite attempts to include descriptions of defensive behaviour to counter its underrepresentation in the literature (Wheeler et al. 2013) and because of the defences integral contribution to tactics in gameplay (Gréhaigne and Godbout, 1995). Unlike the metric types, which were unevenly represented, both teams had a similar number of metrics. Nevertheless, the results suggest that the spatio-temporal behaviour of the offence is a better distinguisher of play outcome.

It might be that the initially suggested metrics capture offensive behaviour well but not defensive behaviour. As mentioned previously, the metrics used were taken from or inspired by the sports performance analysis literature and defensive performance is rarely investigated. Some metrics used in the current work were adaptations of offensive metrics to defensive performance, and vice versa. For example, the Effective Area of Play is the combined area of offensive triangles that are not overlapped by defensive triangles with sides greater than 12 m (Clemente et al. 2013; Clemente et al. 2013); In this definition, a triangle is the region of the pitch defined by three teammates. A defensive version of this metric was included in the current work, which measured the combined area of defensive triangles that are not overlapped by offensive triangles. These adapted metrics might not have captured the essence of defensive behaviour as well as they do offensive behaviour. The alternative is to suggest that the offensive team truly have more influence on the outcome of a play. Future work could develop more relevant defensive spatio-temporal metrics.

The preference for metrics describing the outcome also has some practical implications. One could hypothesise that the spatio-temporal behaviour of players at the beginning of an intrusion does not distinguish the result of the play. For both teams, this implies that seemingly difficult scenarios at the intrusion are not indicative of the conclusion. Given that the outcome of the play is certain at the end, future work could try to describe the developing certainty associated with progress from intrusion to outcome. Similar work has been done in soccer using survival analysis (Nevo and Ritov, 2013).

5.5 Summary and conclusion

The purpose of this chapter was to address the third objective of the current work by reducing the number of initially suggested metrics. Wang et al.'s (2012) variable pre-selection procedure was used with the ***I***-score statistic. The first phase of the variable pre-selection procedure scored metric pairs using the ***I***-score statistic. In the second phase, metrics were scored according to how often they appeared in a top portion of the metric pairs. The third objective was considered to have been met because the number of metrics was reduced from thousands to tens. These results substantially reduced the potential number of processes required in later stages of analyses.

Metrics relating to the offensive team at the outcome dominated the selections across all comparisons. Possible explanations included a lack of metrics relevant to defensive behaviour and the legitimately superior influence of the offensive team on the outcome of plays.

The robustness of the procedure's threshold choices was assessed because communication with the method's authors revealed that "*rigorous theoretical justification*" is still outstanding (Appendix C). Table 5.6 indicates the datasets were carried forward to the next stage of the analysis, detailed in the following chapter.

Table 5.6 The datasets carried forward for subsequent analyses.

Comparison	**Metric-pair threshold, t**	**Retention-frequency threshold, t'**		
		Initial (t'_0)	**Plus ($t'_{+10\%}$)**	**Minus ($t'_{-10\%}$)**
CEvTOc	**Initial (t_0)**	Yes	Yes	Yes
	Plus ($t_{+10\%}$)	Yes	Yes	Yes
	Minus ($t_{-10\%}$)	Yes	Yes	Yes
CEvOther	**Initial (t_0)**	Yes	Yes	Yes
	Plus ($t_{+10\%}$)	*No*	*No*	*No*
	Minus ($t_{-10\%}$)	*No*	*No*	*No*
OthervTOc	**Initial (t_0)**	Yes	Yes	Yes
	Plus ($t_{+10\%}$)	Yes	Yes	Yes
	Minus ($t_{-10\%}$)	Yes	Yes	Yes

CEvTOc = Circle Entry-Turnover comparison.
CEvOther = Circle Entry-Other comparison.
OthervTOc = Other-Turnover comparison.

Chapter 6 Selecting metrics and metric combinations. Stage 2: The Backward Dropping Algorithm

6.1 Introduction

The previous chapter detailed how the initial number of metrics was reduced using Wang et al.'s (2012) variable pre-selection procedure. The procedure required the application of two thresholds, which substantially reduced the initial number of metrics by > 98%. Twenty-one unique sets of metrics were created from the pre-selection procedure but their ability to distinguish the outcome groups is still unknown.

This chapter begins to address the fourth of the project's objectives: to measure the association of metrics with the outcome based on performance in higher-order interactions. The order of the interaction refers to the possible number of interactions between a given set of inputs. In the previous chapter, low-order interactions were considered to be interactions between combinations of two and three inputs. In this chapter, high-order interactions are interactions between a combination of greater than three inputs. The association between metrics and the outcome was measured using Wang et al.'s (2012) Backward Dropping Algorithm. Appendix A explains the fundamentals of the Backward Dropping Algorithm and how its parameters are estimated. A brief description is given in the following paragraphs.

The previous chapter introduced the *I*-score, which is the cornerstone of the Backward Dropping Algorithm. It is a measure of association of an input or set of inputs with an outcome variable. It does this by evaluating the interaction between an outcome variable and inputs' partition elements, where a partition element is a unique expression of inputs' values. For example, a set of three binary inputs has $2^3 = 8$ partition elements because there are eight possible combinations of their values (Table 6.1).

Table 6.1 Every partition element for a set of three binary inputs.

	Input values		
Partition element	**Input 1**	**Input 2**	**Input 3**
1	0	0	0
2	0	0	1
3	0	1	0
4	1	0	0
5	0	1	1
6	1	1	0
7	1	0	1
8	1	1	1

Each partition element represents a specific interaction between the variables. The [0, 0, 0] partition element represents the all-zero interaction and [0, 1, 0] the partition element represents one of the two-zero interactions. Each of these partition elements is exclusively associated with a portion of observations, i.e. some observations were [0, 0, 0] and some were [0, 1, 0]. Not all possible partition elements might be expressed in a dataset. For example, there might not be any case were all three inputs recorded a zero value for the same observation. The number of partition elements affects some parameters of the Backward Dropping Algorithm

The steps of the Backward Dropping Algorithm were presented toward the end of section 2.6.3. The algorithm requires a random selection of k inputs from all possible inputs. This random selection is known as the metric sample set. The algorithm determines the best subset from the metric sample set, known as the return set. The process of selecting a metric sample set and determining a return set must be repeated many times to cover the many possible combinations of inputs. The number of required iterations, $2\hat{B}$, depends on the total number of inputs, the size of the metric sample set and the size of the return set that is almost-completely covered.

The size of the return set that is almost-completely covered, r, represents the maximum order of interaction to be assessed thoroughly during the algorithm. For example, if $r = 3$, then the algorithm would need to be repeated enough times so that every possible triplet of inputs would likely have been sampled in a metric sample set. The return set is only ‘almost-completely covered’ because the coverage of r is probabilistic. The number of expected iterations is increased to improve the chances of actually covering the desired order of interaction.

The following sections present the application of the Backward Dropping Algorithm to score metrics based on their performance in high-order interactions. Two additional processes are applied to reduce redundancy and remove false positives. An adjustment is made to the originally published method to account for unexpected behaviour of the current work's data.

6.2 Method

6.2.1 Backward Dropping Algorithm

The Backward Dropping Algorithm was applied to all 21 datasets created in the previous chapter. All processing was done using MATLAB R2016a (MathWorks 2016). For all comparisons, the size of the metric sample set was $k = 6$ and the size of return set that is almost-completely covered was $r = 3$. The number of iterations, $2\hat{B}$, depended on the number of metrics being considered (Table 6.2). Explanations of parameter values are detailed in Appendix A. The metric combinations outputted from the Backward Dropping Algorithm were subjected to two further processes: reduction in between-return set correlation and removal of false positives.

Table 6.2 The number of Backward Dropping Algorithm iterations, $2\hat{B}$, for every dataset, based on the upper bounds suggested by Wang et al. (2012) for r = 3 and k = 6.

Comparison	Metric-pair threshold	Retention-frequency threshold Initial	Plus	Minus
Circle Entry-Turnover	**Initial**	4,221	5,751	2,998
	Plus	4,221	5,751	2,998
	Minus	3,781	5,204	2,652
Circle Entry-Other	**Initial**	119	162	84
Other-Turnover	**Initial**	957	1,325	666
	Plus	1,325	1,780	957
	Minus	666	957	444

6.2.2 Reducing between-return set correlation

Return sets are the best subset of the metric sample set, as determined by the Backward Dropping Algorithm. Two return sets were correlated if they contained a common metric. Between-return set correlation was reduced by removing lower-scoring return sets that had metrics in common with higher-scoring return sets. Wang et al. (2012) used this method because they used their final sets of variables to build classifiers for genetic traits and uncorrelated classifiers were desirable to minimise redundancy.

Each metric combination in the current work has the potential to describe a tactic but not reducing between-return set correlation produced many highly-correlated metric combinations. For example, a series of metric combinations like $\{\{a,b,c\},\{a,b,d\},\{a,b,e\}\}$ often had lower or identical scores to their $\{a,b\}$ subset. The implication is that the lower-scoring metric combinations were at best introducing uninfluential metrics or were at worst producing a less influential metric combination. To reduce between-set correlation, all return sets were ordered from largest to smallest *I*-score. If a return set had metrics in common with higher-scoring return sets, then it was removed. This resulted in a list of return sets containing unique metric combinations.

6.2.3 Removing false positives

The number of iterations used during the Backward Dropping Algorithm will not completely cover all possible metric sample sets. It is possible that some return sets from the Backward Dropping Algorithm are false positives, i.e. they are not the truly-best return set for the given metric sample set. Wang et al. (2012, Suppl. 3.2.3) suggested the following method to remove false positive return sets:

1. Consider all ordered return sets. These sets are referred to as $R_h : h = \{1, \dots, H\}$, where H is the number of return sets.

2. For a given return set, R_h, make a list of all variables that are not in it. There will be $S - |R_h|$ of these variables, where S is the total number of variables and $|R_h|$ is the number of variables in R_h.

3. Append each of the remaining variables to a copy of R_h to make all possible $|R_h| + 1$ sets. There will be $S - |R_h|$ of these $|R_h| + 1$ sets.

4. Perform the Backward Dropping Algorithm on each of the $|R_h| + 1$ sets and retain the metric sample set with the highest *I*-score. Call these metric sample sets the $(|R_h| + 1)_{BDA}$ sets.

5. Keep all $(|R_h| + 1)_{BDA}$ sets that have a higher *I*-score than their respective R_h set. These retained $(|R_h| + 1)_{BDA}$ sets are called forward-one sets.

6. Count the number of forward-one sets for the given return set. This count is A_h.

7. Repeat the steps for all other return sets, R_h, for $h = \{1, \dots, H\}$.

8. Plot a histogram of A_h. It is expected that there will be an obvious threshold for outliers indicated by "*the first big gap after the main body of the histogram*" (Figure 6.1).

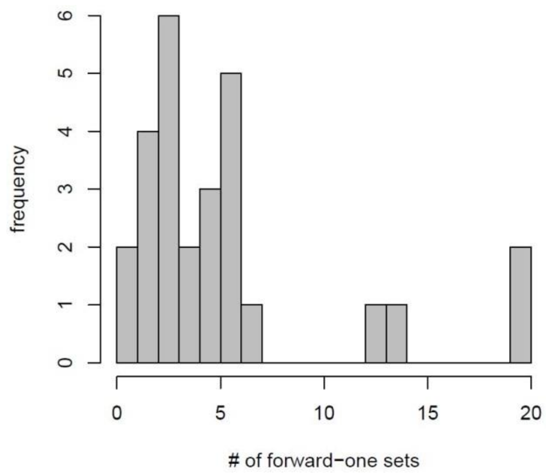


Figure 6.1 An example histogram of A_h given in Wang et al. (2012, Suppl. 3.2.3). In this case, the threshold would be $t^{FPR} = 7$, where the first gap is seen in the histogram.

9. For any R_h whose A_h is above the threshold, replace the R_h with the highest-scoring forward-one set.

10. Keep only unique return sets.

The procedure gives metrics a second chance to improve the return sets. If a return set has few forward-one sets, then it is likely to be a good estimate of the unknown, truly-best return set because there are few better alternatives. If a return set has many forward-one sets, then it is likely to be a bad estimate of the unknown truly-best return set because there are many better alternatives. But, some of the forward-one sets might themselves be false positives. The purpose of the threshold is to balance the risk of the return set being a false positive with the risk of the highest-scoring forward-one set being a false positive.

6.2.4 Adjustment to Wang et al.'s false positive removal procedure

Communications with the authors of Wang et al. (2012) revealed that there is no rigorous theoretical underpinning for choosing the appropriate threshold for the forward-one set histogram. Like the thresholds used during the metric pre-selection procedures, the method is only supported by "*empirical and heuristic evidence*" from the authors' experiences (Appendix C). In the previous chapter, the robustness of the metric pre-selection thresholds was assessed by examining the effect of ±10% of the threshold on the number of metric pairs and number of metrics carried forward (Section 5.2). It was not possible to apply this method to the forward-one threshold because an initial threshold could not be determined based on the guidance provided by the method's authors. For example, where is the main body of the histogram of forward-one sets for the Initial-Plus dataset of the Circle Entry-Turnover comparison (Figure 6.2)?

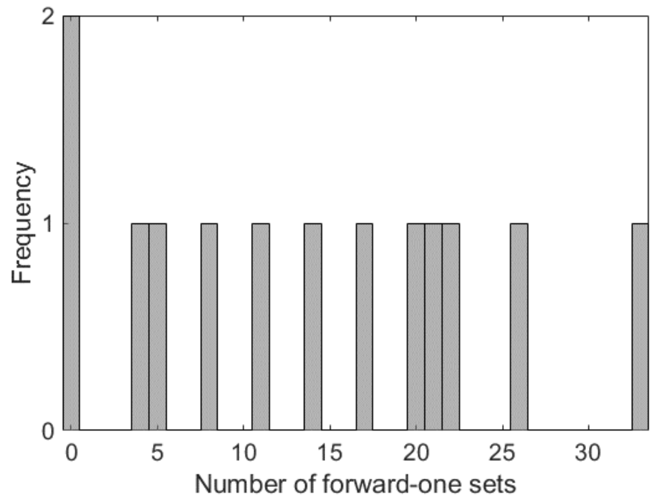


Figure 6.2 The histogram of A_h for the Initial-Plus dataset.

A different method was developed to assess the effect of threshold choice experimentally. Ten equally spaced thresholds were chosen and the differences in the outputted number of unique return sets and the number of unique metrics were examined. The thresholds were rounded-up values of 0%, 10%, 20%, 30%, ..., 90% of the maximum A_h, where A_h is the number of forward-one sets associated with a return set. For example, the maximum A_h for the Initial-Initial dataset was $A_h = 33$ (see Figure 6.2). Thresholds were set 0, 4, 7, 10, 14, 17, 20, 24, 27, 30 (Figure 6.3).

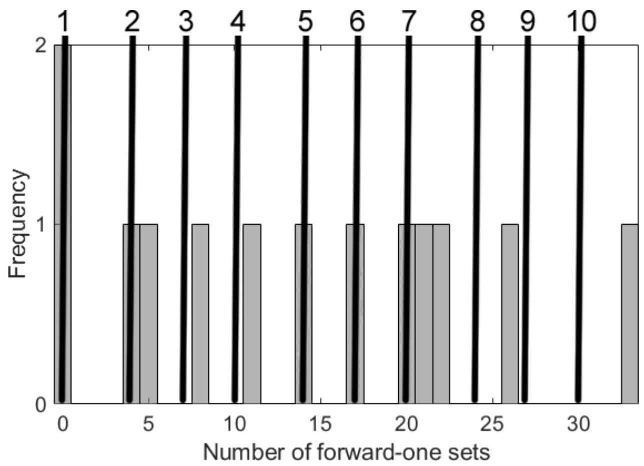


Figure 6.3 The histogram of A_h for the Initial-Plus dataset showing the ten thresholds.

The first threshold represents a case where we are completely confident that the original return set is a false positive and must be replaced by a forward-one set. Increasing the threshold indicates that we are increasingly more confident that the original return set is actually a good estimate of the unknown truly-best return set. At the last threshold, we are indicating that only return sets with the most forward-one sets should be replaced. A hypothetical 11th threshold would be at maximum A_h and would result in no return sets being replaced. This is equivalent to keeping the original return sets and represents the case where we are completely confident that they are not false positives. It can be seen in Figure 6.3 that some thresholds will have the same effect, e.g. both the 9th and 10th threshold mean that only return sets with 33 forward-one sets will adopt their best forward-one set.

6.3 Results

Removing between-return set correlation substantially reduced the number of unique metric combinations (Table 6.3; 6.4; 6.5).

Table 6.3 The number of unique metric combinations produced by each step in the analysis for the Circle Entry-Turnover comparison.

Dataset		Backward Dropping Algorithm	Between-return set correlation reduction
Initial	Initial	497	11
	Minus	324	10
	Plus	1043	13
Minus	Initial	419	11
	Minus	285	10
	Plus	714	13
Plus	Initial	438	12
	Minus	313	11
	Plus	521	12

Table 6.4 The number of unique metric combinations produced by each step in the analysis for the Circle Entry-Other comparison.

Dataset		Backward Dropping Algorithm	Between-return set correlation reduction
Initial	Initial	21	4
	Minus	15	3
	Plus	32	4

Table 6.5 The number of unique metric combinations produced by each step in the analysis for the Other -Turnover comparison

Dataset		Backward Dropping Algorithm	Between-return set correlation reduction
Initial	Initial	184	10
	Minus	123	9
	Plus	301	10
Minus	Initial	753	11
	Minus	344	10
	Plus	911	13
Plus	Initial	836	11
	Minus	514	11
	Plus	1056	14

Applying the adjusted false positive removal procedure created 10 more datasets for each of the 21 brought forward from the previous chapter. Each of the resulting 210 datasets was a unique list of metric combinations. Although each list of combinations was unique, some lists were made of the same set of metrics. For example, a specific set of metrics might make 15 different combinations, ten of which might appear in one list and the remaining five in another. Only unique sets of metrics needed to be carried forward so common sets of metrics created by the application of thresholds were removed. Of the ten possible sets of metrics that could have arisen from applying the ten thresholds, only two to eight unique sets were ever seen. The initially possible 90 unique sets of metrics for the Circle Entry-Turnover comparison were reduced to a more manageable 63. The 30 possible sets of metrics for the Circle Entry-Other comparison were reduced to six and there were 51 unique sets of metrics associated with the Other-Turnover comparison.

6.4 Discussion

The analysis detailed in this chapter selected sets of metrics based on their interactive association with play outcomes. It was noted in the previous chapter that threshold choice during the metric pre-selection procedure affected metric combinations and their constituent metrics. Metric combinations and their constituents were similarly affected by the choice of false positive removal threshold. At this stage of their analysis, the original authors of the method concluded that the variable pre-selection thresholds had no effect on their final set of metrics (Wang et al. 2012). This conclusion was drawn because all of Wang et al.'s (2012) nine datasets from the variable pre-selection produced the same set of final metrics after the application of Backward Dropping Algorithm, between-set correlation removal and the false positive removal procedure. This contrasts with the presence of 63, 6 and 51 unique sets of metrics found for comparisons in the current work.

The contrast is not so surprising because the output of the analysis depends on how the inputs interact. Wang et al. (2012) investigated genes associated with relapse of metastatic breast cancer while the current work investigates spatio-temporal metrics associated with play outcomes in field hockey. There is no expectation that these two sets of inputs should have the same characteristics and interact in similar ways. The results suggest that spatio-temporal metrics associated with play outcomes interact in a more varied manner than genes associated with breast cancer, or that there are more truly-influential metrics in the current work. These differences might have been accounted for by sampling more metrics in each iteration of the Backward Dropping Algorithm or removing fewer metric combinations when reducing between-return set correlation.

The choice of k and reducing between-return set correlations in the manner described in section 6.2.2 might have made the analysis too conservative. Wang et al. (2012) used a metric sample set of size $k = 11$ and the size of the return set that is almost-completely covered was $r = 4$. These parameter values covered a higher order of metric interaction than in the current work. Wang et al. (2012) had this better coverage because their input variables were all binary. The current work had to adjust for the possibility of combinations with metrics that have a range of four values. The range of values that inputs can take affects the number partition elements (Section 6.1) and hence the number of observations per partition element, i.e. the amount of data supporting each possible combination of metrics' values (Appendix A). The increased resolution of metrics' values in the current work was at the expense of covering higher-order interactions. In the next chapter, Chernoff et al.'s (2009) resuscitation analysis is introduced, which can give a second chance to metrics that haven't performed well thus far.

The procedure to reduce between-return set correlation was also very conservative, reducing the number of return sets by 96 - 99%. The one example provided by Wang et al. (2012) expressed a similar reduction in the number of return sets. If the unique datasets that have been outputted are truly representative, then they should be robust to resuscitation analysis detailed in the next chapter.

Finally, the application of multiple thresholds during the false positive removal procedure has not provided any insights into an improved or objective way to choose a threshold. Comparing the unique sets of metrics with the histograms of forward-one sets did not show any obvious relationship. It might have been possible to correlate features of the histograms with watersheds between the unique sets of metrics but no such features were found. Further work is required to inform the choice of a threshold. Work by Chernoff et al. (2009) and Wang et al. (2012) has used simulations of datasets whose response variables are mathematically dependent on a selection of input variables. Applying such methods to the choice of false positive removal threshold might be insightful.

6.5 Summary and conclusions

The fourth objective of the current work was to select metrics and metric combinations that distinguish play outcomes. The purpose of this chapter was to begin this process with the second of three stages of the current work's analysis. The Backward Dropping Algorithm of Wang et al. (2012) was applied to the 21 datasets from the previous chapter.

The metric combinations that were outputted were subject to a false positive removal procedure in accordance with the originally published method. Communications with the authors of the original publication revealed that there is no rigorous theoretical underpinning for choosing a required threshold. The effect of threshold choice was therefore assessed by applying ten equally distributed thresholds. The application of multiple thresholds during the false positive removal procedure did not provided any insights into an improved or objective way to choose a threshold. The assessment produced ten times the number of datasets. The metrics used in each dataset were extracted to create corresponding sets of metrics.

Only unique sets of metrics were carried forward to the final stage of analysis. All comparisons saw a reduction in the number of metric sets because thresholds yielded lists of metric combinations made from common sets of metrics. These 120 unique sets of metrics were carried forward to the resuscitation analysis that is detailed in the next chapter.

Chapter 7 Selecting metrics and metric combinations. Stage 3: Resuscitation analysis

7.1 Introduction

It was discussed in the previous chapter that the pre- and post- filtering methods used by Wang et al. (2012) might be too conservative and prematurely remove metrics and metric combinations. Conservative filtering might therefore reduce the chances of discovering potentially useful tactics. The Backward Dropping Algorithm also does not guarantee complete coverage of all metric combinations. The resuscitation method described in Chernoff et al. (2009) gives a second chance to the metrics that did not pass the metric pre-selection stage or perform well in the Backward Dropping Algorithm.

The resuscitation algorithm requires a small adjustment to the Backward Dropping Algorithm (Appendix A). It differs in the way that the k-number of metrics are selected for each iteration (cf. Appendix D). The standard Backward Dropping Algorithm randomly selects a k-number of metrics from the set of metrics that passed the metric pre-selection. For the resuscitation algorithm, the k-number of metrics are selected by taking a small portion from the highly-ranked metrics determined in the previous chapter and a larger portion from the set of all metrics initially suggested. These latter metrics are known as resuscitation candidates. The method is biased towards the highly-ranked metrics because of their already apparent, strong influence. This is despite more metrics coming from the set of resuscitation candidates.

The resuscitation analysis is designed so that the metrics that were highly ranked are more likely to stay highly ranked while allowing some undiscovered but possibly influential metrics to come to the fore. The resuscitation algorithm differs from the false positive removal procedure (Section 6.2.3) in two ways:

1. The false positive removal procedure selects 'second chance' metrics from the set of metrics already used in the Backward Dropping Algorithm. The resuscitation analysis selects its resuscitation candidates from the set of metrics *not* used in the Backward Dropping Algorithm. For example, 32 metrics were used in the Backward Dropping Algorithm for

the Initial-Initial dataset of the Circle Entry-Turnover comparison. The false positive removal procedure used only these 32 metrics to discover better alternative combinations. The resuscitation analysis will be able to use up to the 3,641 – 32 = 3,609 metrics that never made it past the metric pre-selection stage.

2. The false positive removal procedure appends 'second chance' metrics to existing metrics combinations to test their ability to improve metric combinations. The resuscitation analysis includes the resuscitation candidates in a re-run of the Backward Dropping Algorithm analysis.

There are four special sets of metrics used in the resuscitation analysis:

1. The remaining list
2. The final list
3. The reduced list
4. The resuscitation list

The *remaining list* is a list of resuscitation candidates. It gets its name because it is the portion of the original set of metrics that were left behind after the metric pre-selection procedure. The *final list* is the list of metrics that the user wants to come away with from the resuscitation analysis. The actual number of metrics might end up being less than the desired length of this list (Appendix D). This *final list* is made of the *reduced list* and the *resuscitated list*. The *reduced list* is the portion of the *final list* that the user wants to be made of metrics from the highly-ranked metrics determined in the previous chapter. The *resuscitated list* is the portion of the *final list* that the user wants to be made of resuscitation candidates. Deciding how to portion the *final list* is based on how much confidence there is in the first application of the Backward Dropping Algorithm. The more confident we are that the Backward Dropping Algorithm performed well, the longer the reduced list and the smaller the resuscitated list.

The following sections present the application of the resuscitation analysis to score metrics based on their performance in higher-order interactions. The necessary adjustments to the Backward Dropping Algorithm and new parameters are also explained.

7.2 Method

The resuscitation algorithm was applied to all sets of unique metrics brought forward from the previous analysis, $n_{sets} = 120$. All processing was done using MATLAB R2016a (MathWorks 2016). The Backward Dropping Algorithm parameters were the same for all comparisons: $k = 6$, $r = 3$, $2\hat{B} = 193{,}935$ (Appendix A). The size of metric sample set, k, and the size of the return set that is almost-completely covered, r, remain the same as in the previous application of the Backward Dropping Algorithm. The number of iterations, $2\hat{B}$, was increased to account for the greater number of metrics being considered in the resuscitation analysis. The values of additional parameters are discussed in the following paragraphs

The size of a comparison's final list was determined by the median length of the comparison's metric sets that were brought forward from the previous chapter. As such, it was an estimate of the expected number of unique metrics outputted from applying the Backward Dropping Algorithm.

It was not feasible to include all resuscitation candidates in the remaining list because of resource constraints. For example, with eleven metrics in the reduced list for the Circle Entry-Turnover comparison, there would 3,630 resuscitation candidates (cf. Table 4.5 for number of metrics associated with each comparison). With this many resuscitation candidates to choose from, the number of Backward Dropping Algorithm iterations required would be in the order of billions and beyond the capacity of available resources. A similar problem was addressed with the metric pre-selection procedures in chapter 5. The following method was used to reduce the number of possible resuscitation candidates and create a remaining list specific to each of the datasets brought forward from the previous analysis.

Each of the 120 datasets brought forward from the previous chapter derives from a set of metric pairs that surpassed a threshold applied during metric pre-selection (Section 5.2). The metrics used in these metric pairs were ranked using the retention frequency method, i.e. according to how often they occurred in the pairs. This ranked list of metrics was compared with the associated reduced list and any duplicates were removed. The highest-ranking L_{remain}-number of the non-duplicate metrics were used as the remaining list such that the total number of metrics submitted to the Backward Dropping Algorithm was always 100. For the Circle Entry-Turnover comparison, the remaining list therefore contained 90 metrics for the first application of the Backward Dropping Algorithm because the reduced list contained 10 metrics. A 100 metrics were chosen to accommodate computing resource constraints.

The intention was for the size of reduced and remaining lists to be the same because there was no reason to favour the initial run of the Backward Dropping Algorithm or the resuscitation analysis. However, the resuscitation algorithm constrained $\frac{L_{resus}}{2}$ to be a whole number, where L_{resus} is the size of the resuscitated list. The size of the resuscitated list therefore represented the largest, even, whole-number value less than or equal to half of the size of the final list. The sample set proportion was set at $p = 0.\overline{33}$ so that the number of metrics selected from the reduced list was the largest possible smaller portion of a k-size metric sample set, i.e. $\frac{2}{k} = \frac{2}{6} = 0.\overline{33}$.

The values of these parameters relating to the resuscitation analysis are summarised in Table 7.1.

Table 7.1 The parameter values used in the resuscitation analysis for each comparison.

	Parameter value		
Parameter name	**CEvTOc**	**CEvOther**	**OthervTOc**
Length of final list	21	5	17
Length of reduced list	11	3	9
Length of resuscitated list	10	2	8

CEvTOc = Circle Entry-Turnover comparison.
CEvOther = Circle Entry-Other comparison.
OthervTOc = Other-Turnover comparison.

The resuscitation analysis produced a final list for each of the 120 metric datasets brought forward from the previous chapter. A set of unique metrics were collated and ranked according to their prevalence across final lists.

7.3 Results

The resuscitation analysis added between 7 and 15 new metrics that were not seen in any reduced lists (Table 7.2). Between 14% and 19% of unique final-list metrics appeared in all of a comparison's final lists.

Table 7.2 Count of unique metrics found in the resuscitation analysis. Values for n indicate the number of datasets associated with a comparison.

Unique metrics	**CEvTOc ($n = 63$)**	**CEvOther ($n = 6$)**	**OthervTOc ($n = 51$)**
Input datasets (min-median-max)	15-21-31	4-5.5-7	10-18-32
Common across all reduced lists	21	4	25
Common across all final lists	36	11	32
Appearing in all final lists	5	2	6

CEvTOc = Circle Entry-Turnover comparison.
CEvOther = Circle Entry-Other comparison.
OthervTOc = Other-Turnover comparison.

Tables 7.3, 7.4 and 7.5 show the top three metrics across all final lists. The metrics are ordered, firstly according to the proportion of final lists in which they were found and secondly according the *I*-score of their best combination. Almost all metrics came from reduced lists and were found in all final lists. The first exception was the 3rd-rank metric for the Circle Entry-Other comparison: the pitch-length distance between the ball and the offensive goalkeeper, at the outcome. This metric was found in only 83% of final lists. The other exception was the 3rd-rank metric for the Other-Turnover comparison: pitch-length distance between the offensive midfielders' centroid and the defensive team's centroid, at the intrusion. This metric did not come from a reduced list.

Table 7.3 Top three metrics across all final lists from the resuscitation analysis for the Circle Entry-Turnover comparison.

	Metric	Other metrics in best combination	
1	Pitch-length distance between the goal centre and the possessing player, at the outcome.	Proportion of the Offensive 1st-Half-Middle Region that is overlapped by the Defensive Forward Region, at the intrusion.	Numerical superiority of the offensive team in the region of the pitch where the ball is, at the outcome.
		Proportion of the Defensive Back Region that is overlapped by the Offensive 1st-Half-Middle Region, at the intrusion.	Offensive Unit Principle statistic.
2	Pitch-length location of the pass receiver when the pass is received, if the method of outcome was a pass.	Proportion of the Defensive 1st-Half-Middle Region that is overlapped by the Offensive Back Region, at the outcome.	Numerical superiority of the offensive team in the rear-right region of the pitch, at the intrusion.
		Euclidean distance between the leftmost and the rightmost offensive player, at the outcome.	Pitch-length location of the pass sender when the pass is sent, if the method of outcome was a pass.
3	Pitch-length location of the pass sender when the pass is sent, if the method of outcome was a pass.	Proportion of the Defensive 1st-Half-Middle Region that is overlapped by the Offensive Back Region, at the outcome.	Numerical superiority of the offensive team in the rear-right region of the pitch, at the intrusion.
		Euclidean distance between the leftmost and the rightmost offensive player, at the outcome.	Pitch-length location of the pass receiver when the pass is received, if the method of outcome was a pass.

Table 7.4 Top three metrics across all final lists from the resuscitation analysis for the Circle Entry-Other comparison.

	Metric	Other metrics in best combination	
1	Pitch-length distance between the player who is in possession of the ball and the defender's goal centre, at the outcome.	Proportion of the Defensive 1st-Half-Middle Region that is overlapped by the Offensive Back Region, at the intrusion.	Euclidean distance between the defender's goal centre and the defensive goalkeeper, at the intrusion.
		Numerical superiority of the offensive team in the rear-centre region of the pitch, at the intrusion.	
2	Pitch-length location of the outcome event.	Numerical superiority of the offensive team in the front-centre region of the pitch, at the intrusion.	
3	Pitch-length distance between the ball and the offensive goalkeeper, at the outcome.	Pitch-length distance between the player who is in possession of the ball and the defender's goal centre, at the outcome.	

Table 7.5 Top three metrics across all final lists from the resuscitation analysis for the Other-Turnover comparison. Metrics marked with an asterisk, *, are those whose individual *I*-scores were < 10% of their best combination's.

	Metric	Other metrics in best combination
1	Median pitch-length distance between the ball and players of the defensive team, at the outcome.	Proportion of the Defensive 1st-Half-Middle Region that is overlapped by the Offensive Back Region, at the intrusion.
2	*Pitch-length distance between the offensive midfielders' weighted centroid and the offensive backs' weighted centroid, at the intrusion.	Median pitch-length distance between the ball and players of the defensive team, at the outcome.
3	*Pitch-length distance between the offensive midfielders' centroid and the defensive team's centroid, at the intrusion.	Median pitch-length distance between the ball and players of the defensive team, at the outcome.

7.4 Discussion

The purpose of the resuscitation algorithm was to improve the chances of finding important variables by giving a second chance to those that have not performed well thus far in the current work. Each comparison responded differently to the resuscitation algorithm. The following paragraphs provide a general discussion of the method's performance while the next chapter describes the metrics and combinations in more depth.

7.4.1 False positives

Two metrics shown in Table 7.5 had individual ***I***-scores that were < 10% of their best combination's. If all other metrics in the combination also had low ***I***-scores, then the metrics must demonstrate substantial interaction effects. If the other metrics in the combination had very high individual ***I***-scores, then the low-scoring metric is probably contributing very little to the combination’s performance. As an example, the 7th and 8th ranked metrics from the Circle Entry-Other comparison were, respectively:

- The pitch-length distance between the offensive midfielders’ weighted centroid and the offensive backs' weighted centroid, at the intrusion; and
- The pitch-length distance between the offensive midfielders’ centroid and the defensive team’s centroid, at the intrusion.

Both these metrics formed their best combinations with the metric describing the pitch-length location of the outcome event. These metrics' individual ***I***-scores are almost 10,000 times less than that of pitch-length location of the outcome event metric. This suggests that they do not contribute to the performance of their combinations, which score worse than the individual ***I***-score for pitch-length location of the outcome event metric. If these metrics were demonstrating substantial contribution via interaction effects, then one might expect them to contribute to other combinations, too, but they do not. Although it is possible that they only have a substantial contributory effect for the combination given, it is more likely that these metrics are false positives.

One to three reduced-list metrics had individual *I*-scores < 10% of their best combination's *I*-score. In the Circle Entry-Other and Circle Entry-Turnover comparisons, these metrics appeared infrequently in final lists supporting the idea that they are false positives. In contrast, the one suspect metric in the Other-Turnover comparison appeared in all final lists. This metric was the pitch-length distance between the offensive midfielders' weighted centroid and the offensive backs' weighted centroid, at the intrusion – the same metric mentioned in the previous example. Its surprising appearance in all final lists of the Other-Turnover comparison is likely due to the exceptionally dominant influence of the median pitch-length distance between the ball and players of the defensive team, at the outcome. This metric ranks highest for individual and combination *I*-score, and is involved in the best combination for every final-list metric. This exceptional performance is probably why three of the top five metrics in this comparison have very low individual *I*-scores. That is to say, they are 'piggy-backing' on the performance of the median pitch-length distance between the ball and players of the defensive team, at the outcome.

The purpose of the method used in the current work was to select metrics that strongly distinguish the binary outcome, under the assumption that there are a few, strongly-distinguishing metrics. It applies numerous thresholds and carries forward only the best-of-the-best metrics. For example, the reduced list is only a top portion of well-performing metrics. The method might, therefore, fail to find the few strongly-distinguishing metrics if a single, exceptionally distinguishing metric is also present. In this case, the distinction between the performance of strongly- and weakly- distinguishing metrics might be blurred next to the performance of the exceptional metric. The method, then, might only identify the exceptional metric as important and provide little diversity in the output. This might have been the case for the median pitch-length distance between the ball and players of the defensive team, at the outcome. The final list for the Other-Turnover comparison looks like a random sample of other metrics appended to this exceptional metric. Although it might provide a useful insight into distinguishing spatio-temporal behaviour, its dominance has reduced the breath of possible insights by providing no peers for alternative consideration. Should situations like this arise, it would be advisable to repeat the entire analysis without the exceptional variable to see if others come to the fore and see if the metrics that currently rank high can do so without its interaction. Previously applied procedures have sought to reduce the number of outputs for interpretation but the presence of a suspiciously omnipresent variable warrants the widening of one’s proverbial net.

7.4.2 Reintroduction of intrusion and defensive metrics

The resuscitation analysis has reintroduced defensive metrics, which were completely removed during the metric pre-selection stage (chapter 5). Metrics relating to the start of a play also show improved representation. With few exceptions, these intrusion and defensive metrics are combined with the previously dominant outcome and offensive metrics. These findings counter previous suggestions that play outcomes are not affected by the spatio-temporal behaviour of defensive players or of all players at the start of plays (Section 5.4). The resuscitation analysis presents a more complex relationship where, for example, the pitch-width movement of the play is combined with location of the defending goalkeeper, the possessing player and areas of overlap between the teams.

7.4.3 Performance of reduced-list metrics

It is expected that the reduced-list metrics will generally stay highly ranked, assuming the reduced-list metrics are not false positives. This expectation is based on simulations by Chernoff et al. (2009) using a variety of ranking statistics and is theoretically justifiable given that reduced-list metrics are taken from a set initially selected by a run of the Backward Dropping Algorithm.

Chernoff et al.'s (2009, p17) ranking statistics were not preferred over those used in the current work because they were based on arbitrary thresholds that required further work. Chernoff et al. also did not have to deal with multiple datasets. Nevertheless, it seems appropriate to consider Chernoff et al.'s expectation with respect to the current findings. The top-10 metrics in the Circle Entry-Turnover and Circle Entry-Other final lists were predominantly from reduced lists. This supports Chernoff et al.'s conclusions. However, the reduced-list metrics spread throughout the middle ranks of the Other-Turnover final list and many are in lower ranks of the Circle Entry-Turnover final list. There are three possible explanations.

Firstly, it is possible that the resuscitated metrics that rank above the reduced-list metrics are false positives. It would be surprising to discover false positives so strong that they overwhelm the metrics that consistently performed well in all previous analyses, but, as discussed earlier, it cannot be discounted. Secondly, the statistic used to rank the metrics in the current work does not behave like those of Chernoff et al. (2009). In the current work, final-list metrics were ranked according to the proportion of final lists in which they were found.

A third reason for poor performance by reduced-list metrics is that those data violated an assumption of the metric pre-selection procedure detailed in chapter 5. Part of the procedure selected metrics based on their performance in paired interactions. The reduced-list metrics were a portion of these selected metrics. This early stage of the analysis assumed that metrics' performance in low-order interactions is indicative of their performance in higher-order interactions. The resuscitation analysis was the first opportunity for the unselected metrics to be evaluated in these higher-order interactions. The results might indicate that the metric pre-selection procedure selected metrics that were, actually, poor performers in higher-order interactions even though they performed well in pairs. The particularly outstanding metrics in the current work could be described as those with the following attributes that incorporate metrics' univariate and multivariate performance, and robustness:

1. Presence in greater than 80% of final lists;

2. High-ranking *I*-score for its best combination;

3. High-ranking individual *I*-score.

7.5 Summary and Conclusion

In this chapter, Chernoff et al.'s (2009) resuscitation analysis is presented as a method to improve the chances of finding distinguishing metrics. The resuscitation analysis was designed to give a second chance to metrics that have not fared well thus far in the overall analysis. The outstanding metrics were those that were present in greater than 80% of final lists and ranked highly for individual and best-combination *I*-scores. The metrics with these characteristics are summarised in Table 7.6.

Table 7.6 Outstanding metrics for each comparison.

Circle Entry-Turnover	
	Pitch-length distance between the goal centre and the possessing player, at the outcome.
	Pitch-length location of the pass receiver when the pass is received, if the method of outcome was a pass.
	Pitch-length location of the pass sender when the pass is sent, if the method of outcome was a pass.
	Pitch-length location of the pass receiver when the pass is sent, if the method of outcome was a pass.
	Pitch-length location of the pass sender when the pass is received, if the method of outcome was a pass.
	Pitch-length distance between the passing pair at the moment the pass is received, if the method of outcome was a pass.
	Length of pass, if the method of outcome was a pass.
Circle Entry-Other	
	Pitch-length distance between the player who is in possession of the ball and the defender's goal centre, at the outcome.
	Pitch-length location of the outcome event.
	Pitch-length distance between the ball and the offensive goalkeeper, at the outcome
Other-Turnover	
	Median pitch-length distance between the ball and players of the defensive team, at the outcome.
	Pitch-length distance between the ball and the leftmost offensive player, at the outcome.
	Euclidean distance between the offensive goalkeeper, at the intrusion.

The chapters so far have presented the methods used to determine metrics that distinguish the outcome of plays in field hockey. The findings must now be communicated to coaches and athletes, for whom the current work was conducted. This chapter marks the end to discussions about analysis methods and the beginning of discussions about the application of such efforts. The next chapter details the author's interpretation of the findings, which can be a starting point for discussions with coaches and athletes.

Chapter 8 Interpreting spatio-temporal metrics that distinguish play outcomes

8.1 Introduction

The content of previous chapters explained how the many metrics of players' spatio-temporal behaviour were distilled into a small set that distinguished play outcomes. The findings can be used to inform tactical behaviour by describing the spatial arrangements of players at the start and end of plays. This information can help to signpost coaches and players toward developing effective tactics without being overly prescriptive.

The three outcomes of interest in the current work were a Circle Entry (positive), a Turnover conceded (negative) and an 'Other' outcome (partially positive), which included a variety of gameplay events (see Section 3.2.1). The data from pairs of outcomes were compared to find distinguishing spatio-temporal metrics. A small selection of each comparison's metrics was outstanding because they appeared in many final lists and ranked highly, both individually and in combination with other metrics. This chapter describes these metrics, discusses how to interpret the findings and provides general insights about the pitfalls, assumptions and limitations associated with them.

Each section of this chapter considers one of the three comparisons and its outstanding metrics. Steps have been taken to help communicate findings to coaches and athletes, for whom the current work was conducted. Metric values are mostly presented as Small, Moderate or Large to help swift comparison, with the corresponding numeric values described in the text. Graphical depictions of example scenarios are also presented. The proportion of observations associated with each partition element is presented, where a partition element is a unique expression of inputs' values (Section 6.1). This proportion indicates the likelihood that a given outcome was described by a given expression. For example, consider two metrics that can take the values of Small or Large. If 51% of positive outcomes are associated with a Small value for both metrics, then that implies positive outcomes are characterised by the Small-Small expression more often than chance. Coaches and athletes are therefore encouraged to adopt such spatio-temporal behaviour to achieve positive outcomes.

8.2 Circle Entry-Turnover comparison

Table 8.1 shows the outstanding metrics for this comparison. Some conclusions can immediately be drawn:

1. Observing that all metrics relate to the outcome suggests that plays ending in a Circle Entry and Turnover are indistinguishable at the start of the play.

2. Observing that six of the seven metrics relate to passing suggests that passing could be a focus for intervention.

What follows are examples of how the combinations of these outstanding metrics can be interpreted.

Table 8.1 Outstanding metrics for the Circle Entry-Turnover comparison.

Pitch-length distance between the goal centre and the possessing player, at the outcome.
Pitch-length location of the pass receiver when the pass is received, if the method of outcome was a pass.
Pitch-length location of the pass sender when the pass is sent, if the method of outcome was a pass.
Pitch-length location of the pass receiver when the pass is sent, if the method of outcome was a pass.
Pitch-length location of the pass sender when the pass is received, if the method of outcome was a pass.
Pitch-length distance between the passing pair at the moment the pass is received, if the method of outcome was a pass.
Length of pass, if the method of outcome was a pass.

Pitch-length distance between the goal centre and the possessing player, at the outcome.

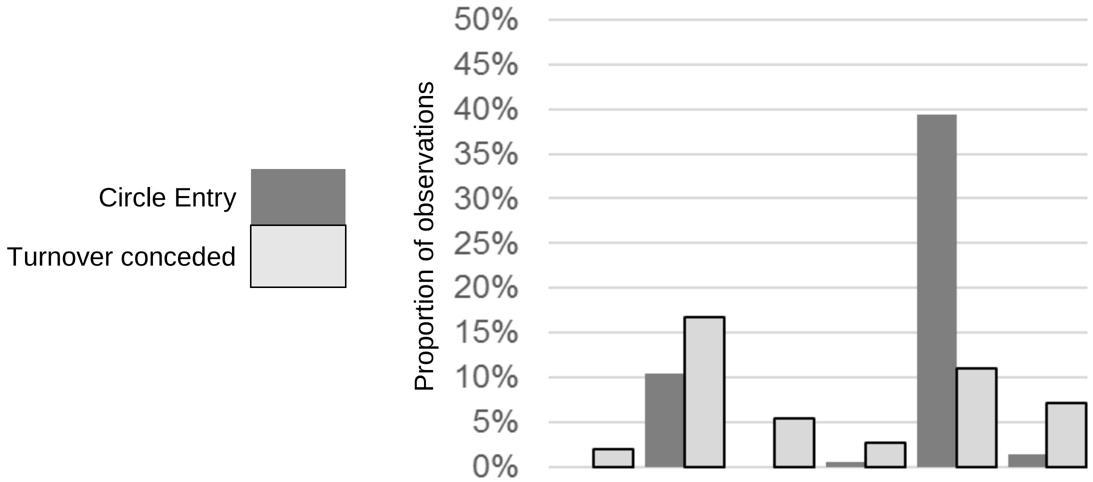
This was a reduced-list metric that appeared in all final lists and boasts the highest individual and combination *I*-score. In its highest-scoring combination, it is combined with:

- The numerical superiority of the offensive team in the region of the pitch where the ball is, at the outcome;
- The proportion of the Offensive 1st Half Middle Region that is overlapped by the Defensive Forward Region, at the intrusion;
- The proportion of the Defensive Back Region that is overlapped by the Offensive 1st Half Middle Region, at the intrusion;
- Offensive Unit principle statistic.

The second and third metrics relate to the Defensive Play Area metric, which segments a team’s surface area in triangles that link players with specific roles (Clemente et al., 2015). The Defensive Back Region is the defensive team's area that is defined by the triangles that link the goalkeeper and the backs (see Section 2.5.2 *Player groups* for description of how player roles were determined). The Defensive Forward Region is the defensive team's area that is defined by the triangles that link the forwards. The Offensive 1st Half Middle Region is the offensive team's area that is defined by the triangles that link the midfielders and backs (the offence's defenders). The original Defensive Play Area metrics were only concerned with the defensive team so equivalent metrics for the offensive team were developed for the current work. The Offensive Unit principle statistic is a binary variable that is true when at least half of the offensive players behind the ball move in the same direction as the ball (Clemente, Martins, et al. 2014b).

Figure 8.1 presents the proportions of outcomes associated with each expression of the metric combination. The figure shows that the latter three metrics had the same values for all plays, i.e. overlaps of regions were always

Small (≈ 0 m^2) and at least half of the offensive players behind the ball moved in the direction of the ball during the play. The differences in the likelihood of either outcome was therefore indicated by the values of the first two metrics.



Pitch-length distance between the goal centre and the possessing player, at the outcome.	Large	Large	Large	Small	Small	Small
Numerical superiority of the offensive team in the region of the pitch where the ball is.	+	-	=	+	-	=
Proportion of the Offensive 1st Half Middle Region that is overlapped by the Defensive Forward Region.	Small	Small	Small	Small	Small	Small
Proportion of the Defensive Back Region that is overlapped by the Offensive 1st Half Middle Region.	Small	Small	Small	Small	Small	Small
Offensive Unit principle statistic.	Pos.	Pos.	Pos.	Pos.	Pos.	Pos.

Figure 8.1 Likelihood of each outcome for the given combinations of metrics in the Circle Entry-Turnover comparison. Only expressions with likelihoods > 1% are shown.'+' indicates more offensive than defensive players. '–' indicates fewer offensive than defensive players. '=' indicates equal numbers of offensive and defensive players. 'Pos.' indicates a positive outcome.

At 17%, the most likely spatio-temporal behaviour that would illicit a Turnover would see the possessing player at greater than 15 m from the goal line with more opponents than teammates in his or her region, at the moment of the outcome event (exemplar observation in Figure 8.2). The practical advice to defenders is therefore to close the space surrounding the attacker within 8 m of them entering the 23 m region. This empirically derived conclusion echoes widely accepted principles that attackers should seek to create space and defenders should seek to constrain space (Vilar, Araújo, Davids and Button, 2012).

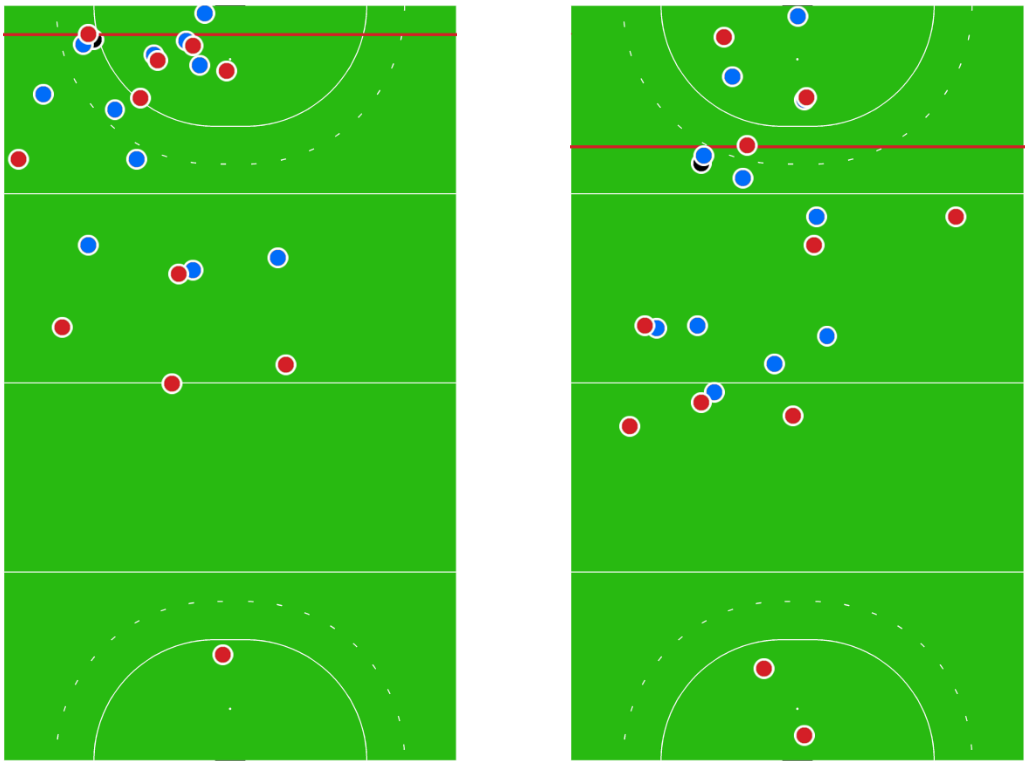
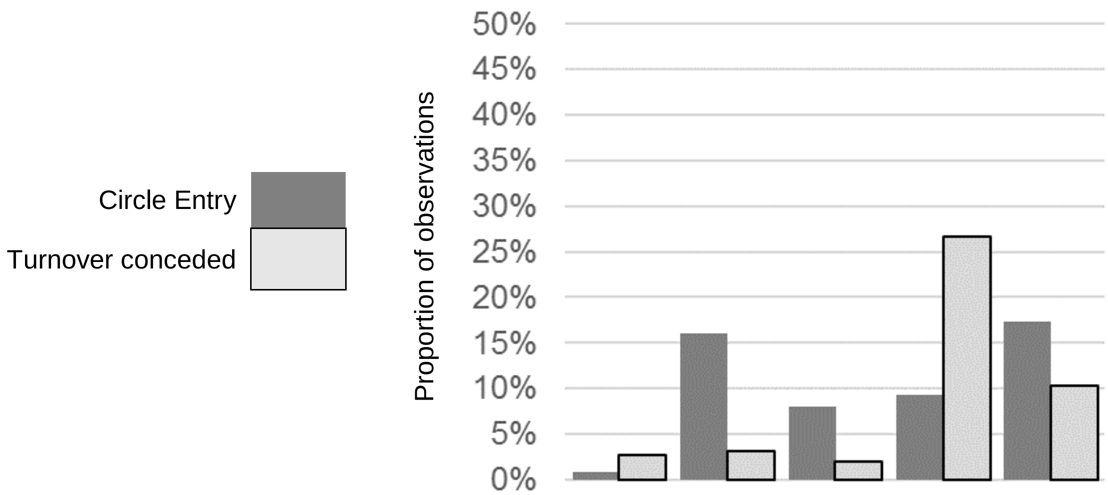


Figure 8.2 Player locations at play outcome. Offensive team in red, attacking upward. Defensive team in blue. Red line indicates possessing player's pitch-length location. *Left*: Circle Entry outcome showing Small value for possessing player's pitch-length distance to goal. *Right*: Turnover outcome showing Large value for possessing player's pitch-length distance to goal.

The most likely positive outcome and most likely negative outcome only differ for the value of the possessing player's pitch-length distance to the goal line. From an offensive perspective, 39% of Circle entries occurred when the possessing player was within 15 m of the goal line with more opponents than teammates in his or her region, at the moment of the outcome. Considering the aforementioned characteristics of a Turnover, this finding suggests that defensive pressure on an attacker is insufficient to tip the odds in the defence's favour. Instead, pitch-length distance to the goal line is the key distinction. When all other metric values are equal, a smaller distance to the goal line changes the odds from 1.6-times in favour of a Turnover to 3.6-times in favour of a Circle Entry.

Passing metrics

The other outstanding metrics all related to passing, they share many of the same metrics in their best combinations and were all reduced-list metrics. Two equally high-scoring combinations contain two of these metrics and distinguish equivalent proportions of outcomes. Figures 8.3 presents the proportions of outcomes associated with the expressions of one of these metric combinations.



Euclidean distance between the leftmost and the rightmost offensive player, at the outcome.	Large	Large	Small	Large	Small
Pitch-length location of the pass sender when the pass is sent, if the method of outcome was a pass.	Small	Small	Large	*null*	*null*
Pitch-length location of the pass receiver when the pass is received, if the method of outcome was a pass.	Large	Large	Large	*null*	*null*
Numerical superiority of the offensive team in the rear-right region of the pitch, at the intrusion.	=	=	=	=	=
Proportion of the Defensive 1st-Half-Middle Region that is overlapped by the Offensive Back Region, at the outcome.	Small	Small	Small	Small	Small

Figure 8.3 Likelihood of each outcome for the given combinations of metrics in the Circle Entry-Turnover comparison. Only expressions with likelihoods > 1% are shown. = indicates equal numbers of offensive and defensive players.

As in the previous example, some metrics present the same values for all expressions, thereby offering no distinction. They are not false positives because their presence in the combination improves the *I*-score beyond what would have been in their absence. In doing so, they offer greater context to the distinguishing spatio-temporal behaviour of players that this metric combination describes. The spatio-temporal behaviours that are most likely to result in a Circle Entry or Turnover are only distinguished by one metric: the Euclidean distance between the leftmost and the rightmost offensive player, at the outcome. An exemplar observation from the current work suggests that this metric gives an indication of the team spread, with shorter distances being associated with Circle entries (Figure 8.4).

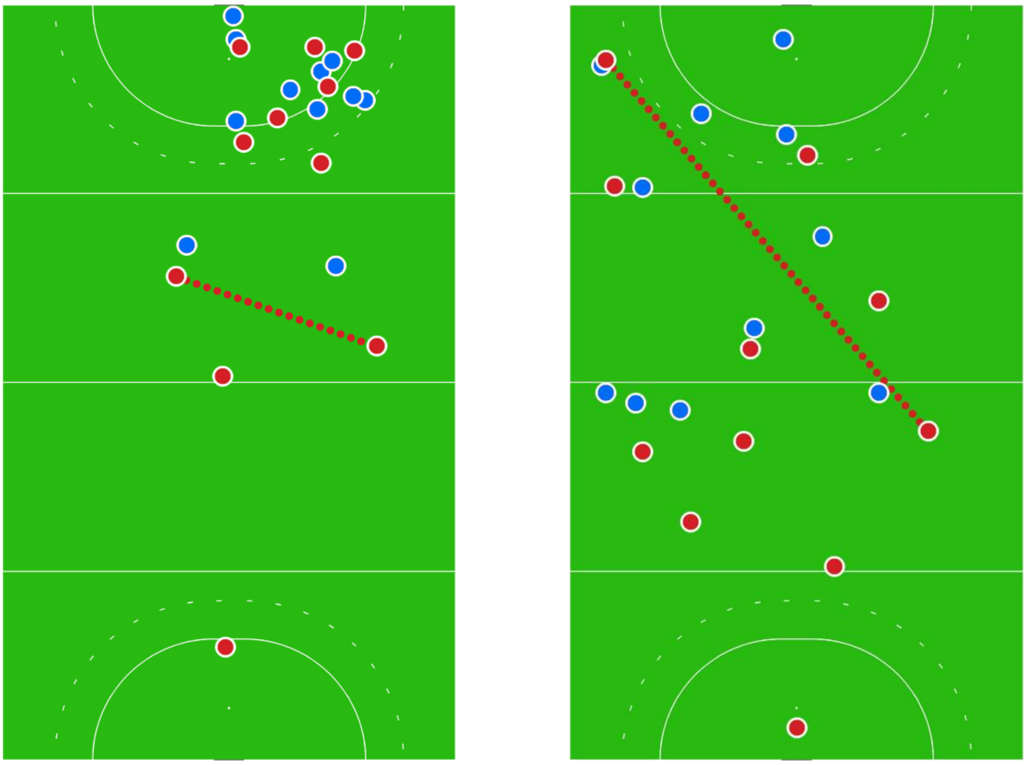


Figure 8.4 Player locations at play outcome. Offensive team in red, attacking upward. Defensive team in blue. Dashed red line indicates the Euclidean distance between the widest offensive players. *Left*: Circle Entry outcome showing a Small value. *Right*: Turnover outcome showing a Large value.

Figure 8.3 shows that 'no pass' events were associated with the spatio-temporal behaviours most likely to lead to a Circle Entry and Turnover outcomes. The proportion of ‘no pass’ events (read as *null* in Figure 8.3) was greater for turnovers than Circle entries at 80% and 52%, respectively. At first glance, these proportions suggest that passing should be a preferred option for the offensive team and the defensive team should limit attackers’ passing options. It is important at this point to note a bias created by the research design: A Circle Entry outcome did not consider which team received the ball within the Circle. Therefore, some Circle Entry outcomes ended as turnovers but only after the ball entered the Circle. Discussions with the collaborating partner suggested that penetrating the Circle is still a positive outcome because it at least offers an opportunity for scoring. The data related to Circle Entry outcomes therefore describe scenarios of scoring opportunities that are both favourable and unfavourable. Future work can interrogate the Circle Entry data to distinguish those that are ‘good’ and ‘bad’. Assuming that some passed Circle entries were intercepted, and thus actually turnovers, turnovers are likely to be further associated with passing. This would reduce the proportion of ‘no pass’ events to a number more similar to that of Circle entries. The implication being that it is the execution of the pass that distinguishes the outcomes and not merely the presence of one.

The histograms of metric values in Figure 8.5 suggest that, in comparison to turnovers, passes for Circle entries are characterised by:

1. The pass sender and receiver being closer to the goal line when the pass is sent and received;
2. A shorter pass length;
3. A smaller distance between the passer and receiver when the pass is sent.

These findings translate into tactical advice for both teams: The offence should penetrate deep into the 23 m region and close to the Circle before attempting a pass; The defence should keep attackers at the edges of the 23 m region.

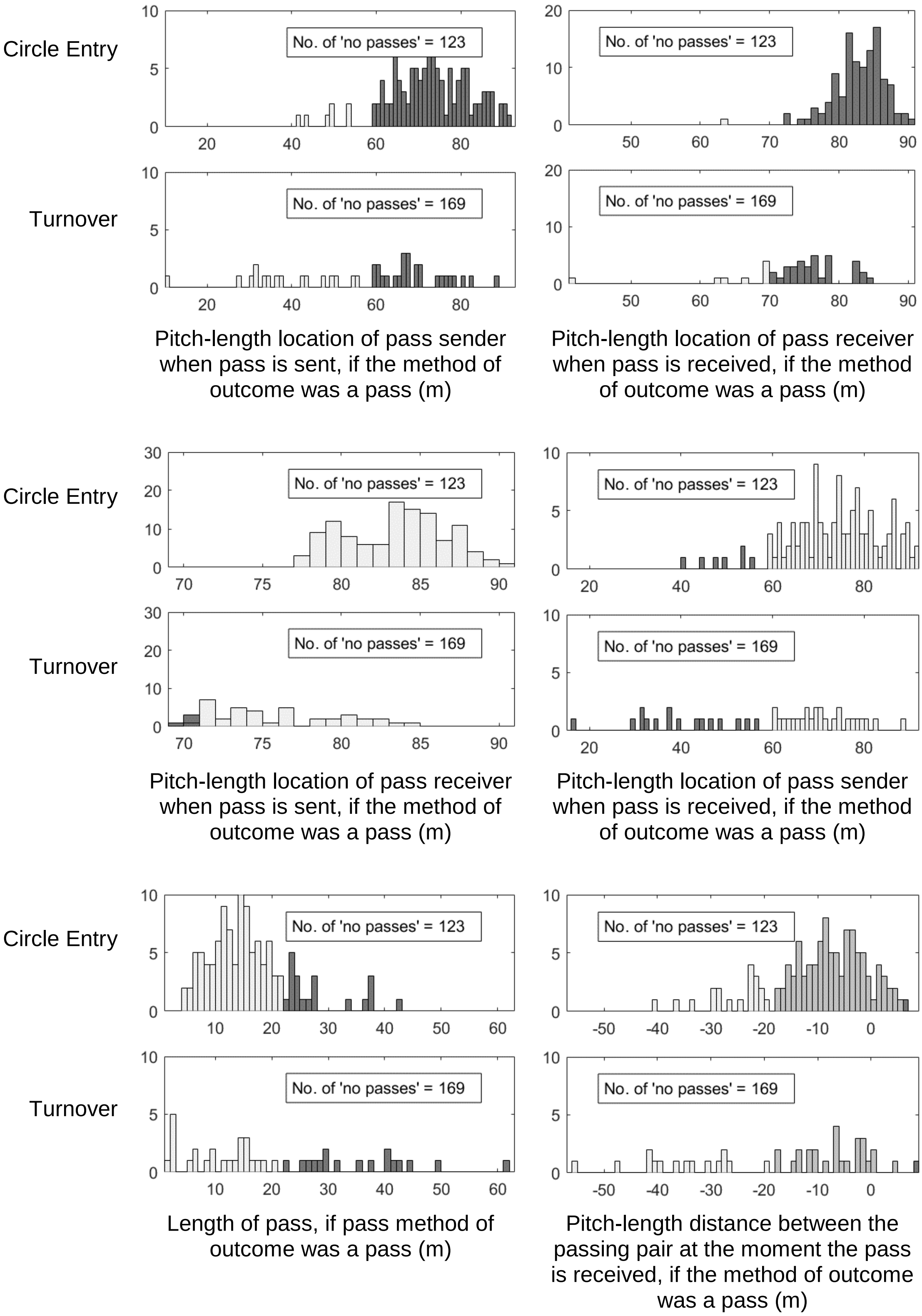


Figure 8.5 Histogram of metric values for metrics passing metrics. Colours of columns indicate membership to the discrete values used in the analysis.

8.3 Circle Entry-Other comparison

Table 8.2 shows the outstanding metrics for this comparison. Some conclusions can immediately be drawn:

1. Observing that all metrics relate to the outcome suggests that plays ending in a Circle Entry and an 'Other' outcome are indistinguishable at the start of the play.

2. Observing that all metrics relate to pitch-length distances suggests that depth of penetration into the 23 m region is important.

What follows are examples of how the combinations of these outstanding metrics can be interpreted.

Table 8.2 Outstanding metrics for the Circle Entry-Other comparison.

Pitch-length distance between the player who is in possession of the ball and the defender's goal centre, at the outcome.
Pitch-length location of the outcome event.
Pitch-length distance between the ball and the offensive goalkeeper, at the outcome.

Pitch-length location of the outcome event

This reduced-list metric appeared in all final lists and had an individual *I*-score larger than its best combination. In its highest-scoring combination, it was combined with the numerical superiority of the offensive team in the front centre region of the pitch. This appending metric indicates the difference in the count of offensive and defensive players in 12 regions of the pitch (Figure 8.6).

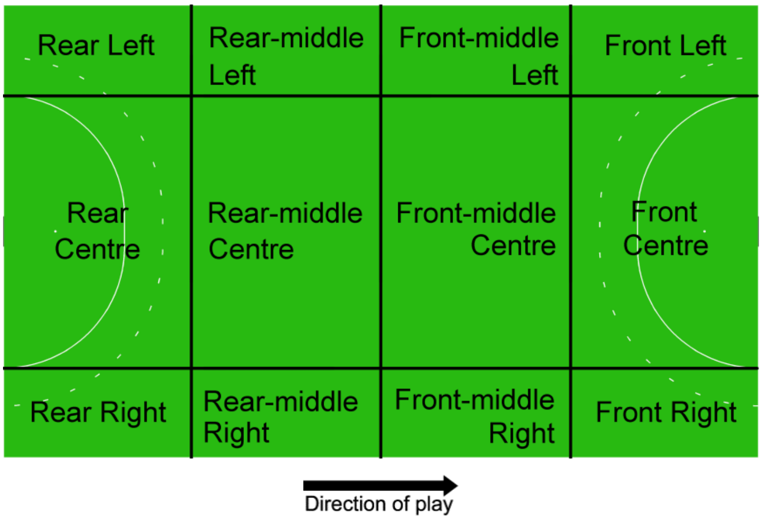
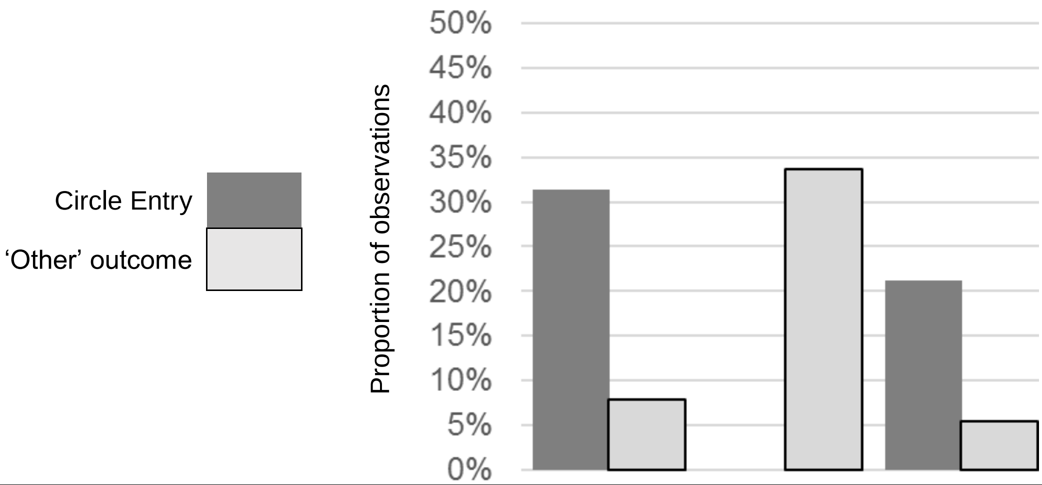


Figure 8.6 Regions of the pitch considered metrics relating to numerical superiority of offensive players.

Figure 8.7 shows the proportions of outcomes associated with each expression of the metric combination. The numerical superiority metric is likely to be a false positive because it offers no distinction between outcomes and lowers the *I*-score when included. The pitch-length location of the outcome therefore appears to distinguish this pair of outcome events single-handedly. Small values (< 74 m) are associated with 'Other' outcomes and Moderate values (> 76 m, < 84 m) are associated with Circle entries. The reader is reminded that the definition of the calibrated plane was such that the offensive team's goal line was at zero pitch-length; Larger values of the pitch-length location of the outcome are closer to the defensive team's goal line.



Pitch-length location of the outcome	Moderate	Small	Large
Numerical superiority of the offensive team in the rear-right region of the pitch, at the intrusion	-	-	-

Figure 8.7 Likelihood of each outcome for the given combinations of metrics in the Circle Entry-'Other' comparison. Only expressions with likelihoods > 1% are shown. – indicates fewer offensive than defensive players.

It is important to consider the events that defined ‘Other’ outcomes when interpreting them (see Table 3.2). All but one of the events could happen anywhere within the 23 m region but outside of the Circle. A 23 m Exit, however, was always recorded at the 23 m line - a pitch-length of 68.4 m. Metrics involving the location of the ball, the outcome event or the possessing player will be biased to this pitch-length location. This is because 62% of ‘Other’ outcomes were 23 m Exit events. Similarly, metrics relating to the Euclidean distance between the goal centre and the ball, Circle entries or the possessing player will be biased to the approximate radius of the Circle, i.e. 14.63 m. The histograms of metric values (Figure 8.8) clearly show the bias of ‘Other’ outcomes to the 23 m line with a peak around 68 m, that is, 23.4 m from the defensive team’s goal line. The Circle bias is also demonstrated by a peak around 78 m – approximately at the length of Circle’s radius – which suggests that the Circle was more often entered at the middle of the arc. Sunderland et al. (2006) also found that Circle entries occurred predominantly down the centre but Sofwan et al.'s (2012) findings suggest that winning teams enter from the sides. These studies and the current work all had different definitions of a central Circle Entry and only Sofwan et al. (2012) distinguished winning and losing teams. It might be that the findings of Sunderland et al. (2006) and the current would support Sofwan et al. (2012) if data were similarly grouped. When collating research on field hockey, one must also be aware of the frequent rule changes that might render some findings obsolete (FIH, 2015, p4 *Rules Review)*. For example, Tromp and Holmes (2011) found that a self-pass rule sped-up the game and changed typical tactics.

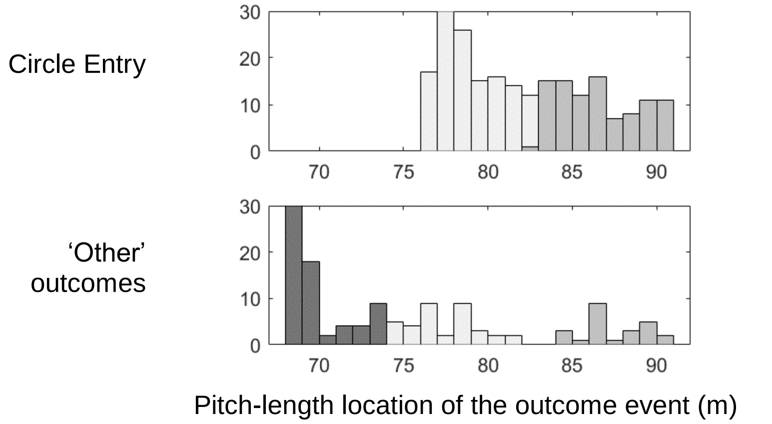


Figure 8.8 Histogram of metric values. Colours of columns indicate membership to the discrete values used in the analysis.

The pitch-length location of the outcome logically and empirically distinguishes the outcomes. The analysis' selection of this metric is testament to its ability to detect distinguishing metrics. To better detect distinguishing metric combinations, it might be useful in future applications to iterate the analysis and remove any metrics like this that individually score better than many combinations.

Pitch-length distance between the ball and the offensive goalkeeper, at the outcome

In its highest-scoring combination, this metric is combined with the pitch-length distance between the goal centre and the possessing player (Figure 8.9). This appending metric has some credentials that suggests the combination's score might be heavily dependent on it: It was one of the outstanding metrics identified in the previous chapter, it appends eight other final-list metrics and was part of the highest-scoring metric pair in the first stage of the analysis.

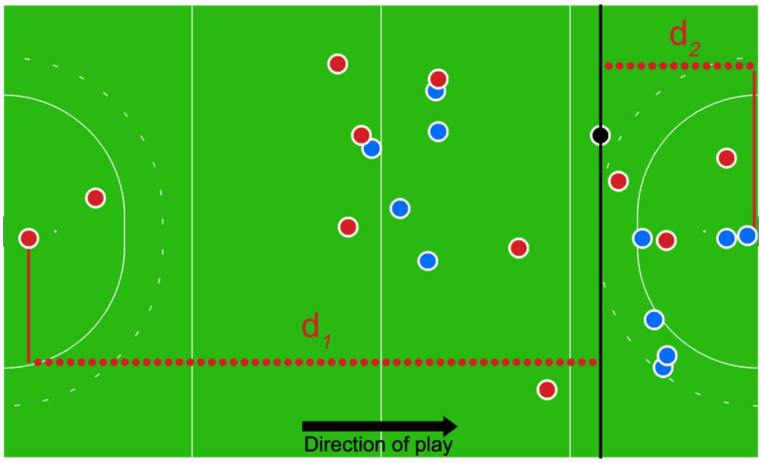
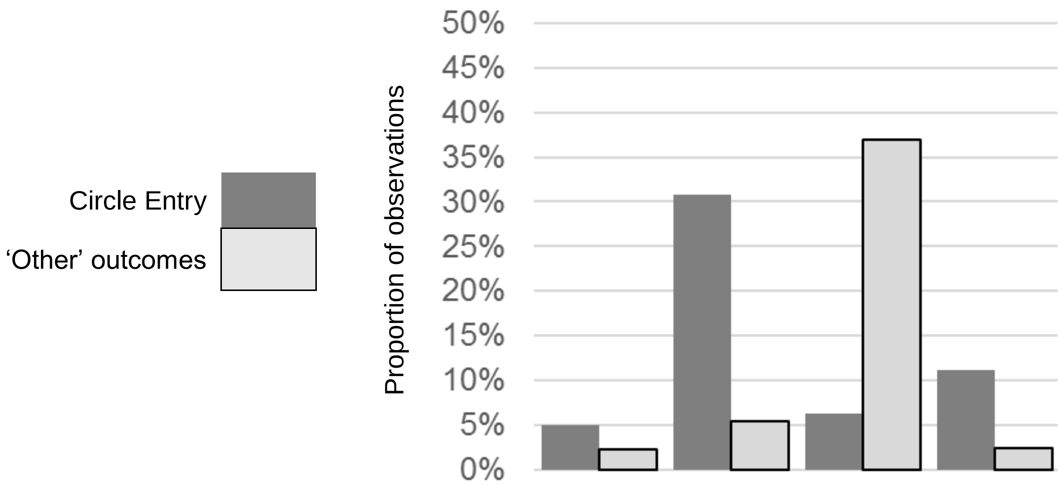


Figure 8.9 Defence in blue, offence in red. *Solid black line*: Pitch-length location of the ball. d_1: The pitch-length distance between the ball and the offensive goalkeeper. d_2: The pitch-length distance between the goal centre and the possessing player.

Circle entries are almost six-times more likely than 'Other' outcomes when Large values for the ball-to-offensive goalkeeper distance are seen (> 74 m) with Small values for goal-to-possessing player distance (< 15 m) (Figure 8.10). Conversely, 'Other' outcomes are even more likely than Circle entries when Small values for the ball-to-offensive goalkeeper distance are seen (< 74 m) with Large values for goal-to-possessing player distance (> 15 m).



Pitch-length distance between the ball and the offensive goalkeeper, at the outcome	Large	Large	Small	Small
Pitch-length distance between the goal centre and the possessing player, at the outcome	Large	Small	Large	Small

Figure 8.10 Likelihood of each outcome for the given combination of metrics in the Circle Entry-'Other' comparison. Only expressions with likelihoods > 1% are shown.

The bias of the 23 m Exit events is clear in the histograms of metric values (Figure 8.11). Circle entries occur farther from the offensive goalkeeper because the Circle is farther than the 23 m line. Similarly, the histogram of the appending metric shows the 23 m bias and the constraints on the possible pitch-length locations of a Circle Entry, i.e. the radius at 14.63 m.

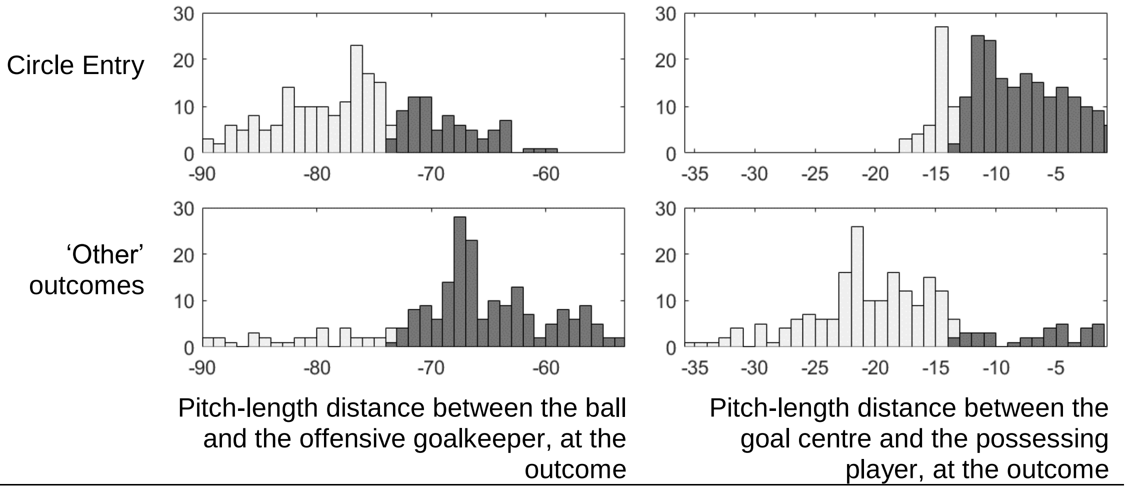


Figure 8.11 Histogram of metric values. Colours of columns indicate membership to the discrete values used in the analysis.

This metric combination is an example of an obvious distinction and further evidences the bias introduced by including metrics that indirectly define the outcome groups. It can be summarised by saying that attackers are more likely to enter the Circle if there are far from their goal and close to the opposition's. Positively, this finding not only provides empirical support for preconceptions, but also suggests that the method used in the current work can identify distinguishing metric combinations.

8.4 Other-Turnover comparison

Table 8.3 shows the outstanding metrics that distinguished plays ending in 'Other' outcomes from those ending in a Turnover. Unlike for the previous comparisons, no broad conclusions are immediately evident from the list of outstanding metrics. In the previous chapter, it was acknowledged that the best combination for all final-list metrics was a pairing with the same, exceptionally performing metric. This exceptionally performing metric was the median pitch length distance between the ball and players of the defensive team, at the outcome. The overwhelming performance of this metric was likely responsible for the lack of diversity in final-list metrics for this comparison. It was also identified that many of the metrics that appended this exceptional metric were very poor performers and were likely 'piggy-backing' with minimal interference. What follows are examples of how the combinations of the outstanding metrics can be interpreted.

Table 8.3 Outstanding metrics for the Other-Turnover comparison.

Median pitch-length distance between the ball and players of the defensive team, at the outcome.
Pitch-length distance between the ball and the leftmost offensive player, at the outcome.
Euclidean distance between the offensive goalkeeper, at the intrusion.

Median of pitch-length distances between the ball and players of the defensive team, at the outcome

The previous chapter discussed how this metric performed exceptionally well (Section 7.4.1). It appends every final-list metric in their best combination and, like the pitch-length location of the outcome in the Circle Entry-Other comparison, has an individual *I*-score larger than its best combination. Only four metrics were involved in all best-combinations of the outstanding final-list metrics. Figure 8.12 shows the histograms of metric values for these metrics. The top-left histogram suggests that, on average, defensive players are ahead of the ball by a little over 10 m when ‘Other’ outcomes occur. Conversely, defensive players are, on average, in line or behind the ball when turnovers occur. Figure 8.13 shows two exemplar cases to illustrate these characteristics, which overwhelmingly distinguish ‘Other’ and Turnover outcomes. Both cases indicate high defensive pressure that, in the case of the ‘Other’ outcome, forced a retreat and a passed 23 m Exit.

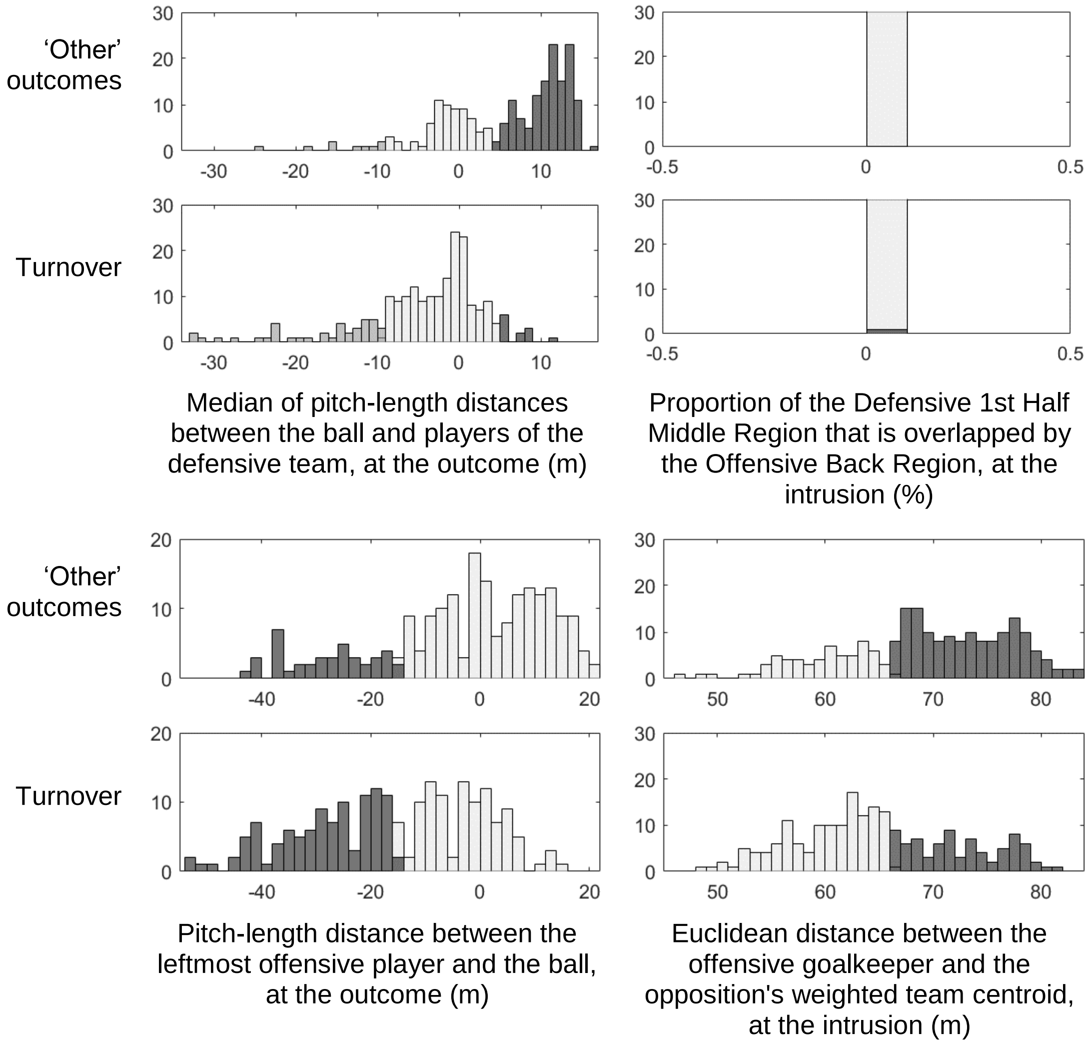


Figure 8.12 Histogram of metric values. Colours of columns indicate membership to the discrete values used in the analysis.

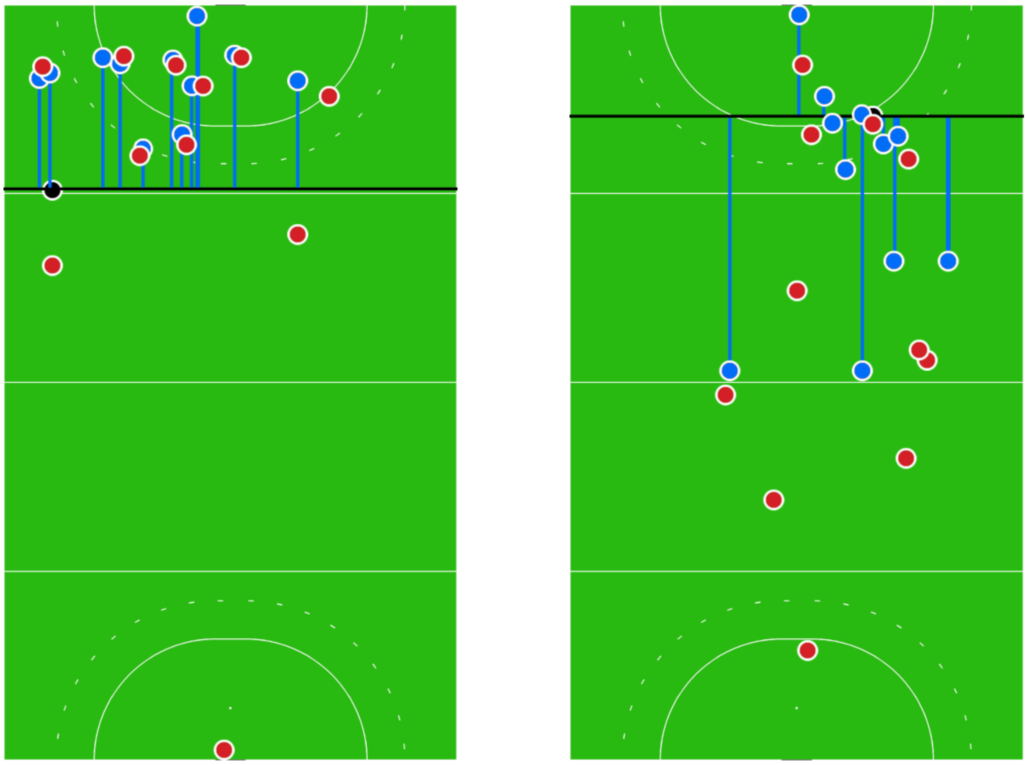


Figure 8.13 Pitch-length distances of defensive players to the ball. Defence in blue, offence in red. *Solid black line*: The pitch-length location of the ball. *Solid blue line*: Defensive players' pitch-length distances to the ball. *Left*: Exemplar case of 'Other' outcomes where players are, on average, ahead of the ball. *Right*: Exemplar case of Turnover outcomes where players are, on average, behind the ball.

The highest-scoring metric combination included the proportion of the Defensive 1st Half Middle Region that is overlapped by the Offensive Back Region, at the intrusion. This metric is likely to be a false positive because it offers no distinction between outcomes (top-right of Figure 8.12) and lowers the *I*-score when appended. The other appending metrics, however, show potential.

The distributions of values for the other appending metrics are skewed for 'Other' outcomes and more symmetrical for turnovers (bottom row of Figure 8.12). The skew suggests that 'Other' outcomes occur when the leftmost attacker is ahead or at least in-line with the ball at the moment of the outcome. Conversely, turnovers are characterised by the leftmost attacker being behind the ball. This latter observation might indicate a lack of offensive support. Looking again at the exemplar cases in Figure 8.13, the leftmost attacker is ahead of the ball for the 'Other' outcome and the difference in offensive support is clear.

The peak in the ‘Other’ outcome’s distribution at 0 m is probably due to the leftmost player being the possessing player because the possessing player was defined as the attacking player closest to the ball. If this is the case, then we can infer that attacking on the far left is more likely to result in an ‘Other’ outcome than a Turnover outcome. In Stöckl and Morgan's (2013) study of spatial characteristics of field hockey attacks, they also noted that more positive outcomes were associated with left-side attacks. This is contrary to conventional field hockey wisdom, which encourages right-side attacks (Sunderland et al. 2006) - indeed, Stöckl and Morgan (2013) found that most attacks were right-sided. Dribbling the ball with the one-sided field hockey stick is safer on the right side of the pitch because the attacker positions themselves between the ball and defenders. Controlling the ball down the left side of the pitch requires more skill or a long, open pass. The greater skill required for a left-sided attack might explain the greater number of preferred outcomes that result, assuming better skilled players create more positive outcomes. Furthermore, left-side attacks might be rare because of insufficient space. When space is available to permit a lift-side attack, it might also facilitate further positive progress.

The final metric to be discussed is the Euclidean distance between the offensive goalkeeper and the opposition’s weighted team centroid, at the intrusion (Figure 8.14). The weighted team centroid is the arithmetic mean location of the team, weighted by players’ proximity to the ball. The exemplar cases presented in Figure 8.14 echo previous observations that more positive offensive outcomes are associated with a left-side attack.

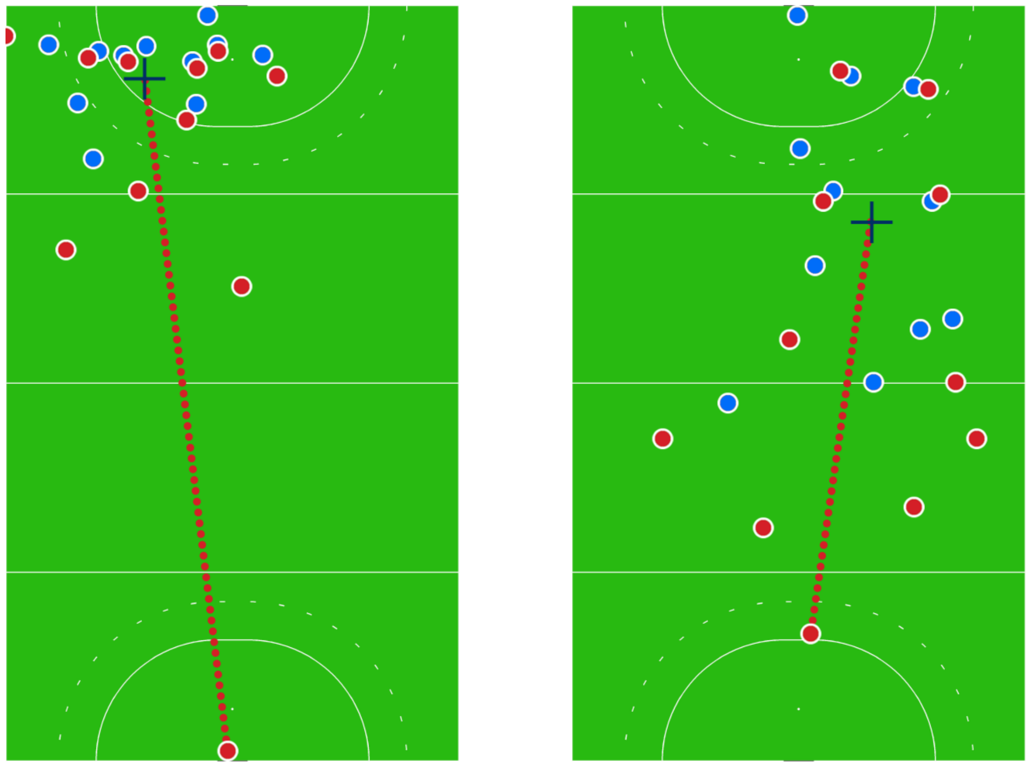
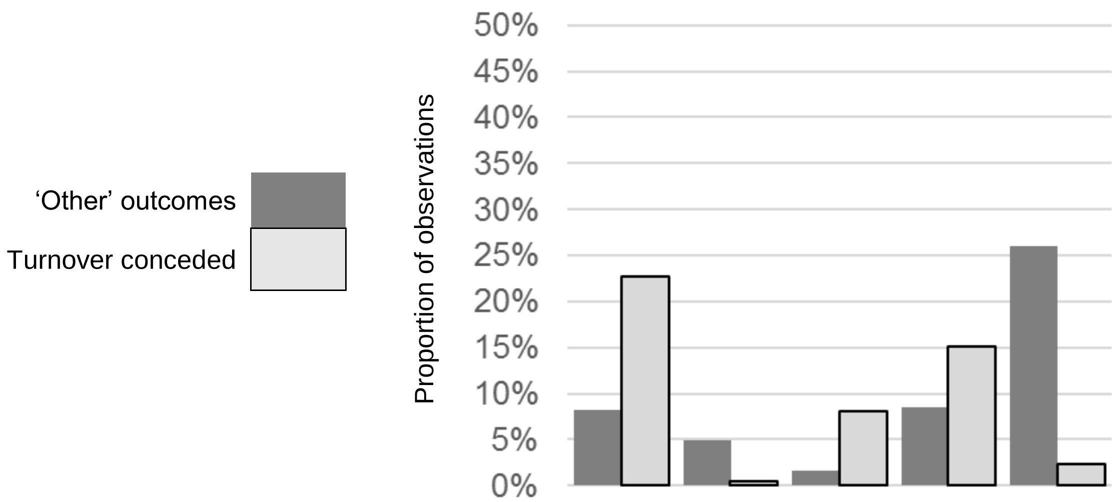


Figure 8.14 Euclidean distance between the offensive goalkeeper and the opposition's weighted team centroid (dashed red line). Defence in blue, offence in red. *Blue cross*: The defensive team's weighted centroid. *Left*: Exemplar case of 'Other' outcomes. *Right*: Exemplar case of Turnover outcomes.

The histogram of metric values suggests that turnovers are associated with a smaller distance than 'Other' outcomes. It could be that plays ending in 'Other' outcomes start within the 23 m region more often, or that plays are more likely to be neutralised when restarted at the 23 m line or by crossing it. Plays are almost three-times more likely to end in a Turnover if defenders are within 5 m of the ball at the start of the play and are < 66 m from the offensive goalkeeper (Figure 8.15). Plays are eleven-times more likely to end in an 'Other' outcome if defenders are over 5 m ahead of the ball and, subsequently, farther from the offensive goalkeeper.



Median of pitch-length distances between the ball and players of the defensive team, at the outcome	Mod.	Large	Small	Mod.	Large
Euclidean distance between the offensive goalkeeper and the opposition's weighted team centroid, at the intrusion	Small	Small	Small	Large	Large

Figure 8.15 Likelihood of each outcome for the given combination of metrics in the 'Other'-Turnover comparison. Only expressions with likelihoods > 1% are shown.

8.5 Summary and Conclusion

In this chapter, outstanding metrics for each comparison were discussed in greater depth. Colour-coded histograms showed the thresholds and distributions of metric values and discussions provided practical advice for team tactics. The variety of metrics and conclusions among the outcome comparisons support the research design that compared all pairs of outcomes separately. No expression of metric values was associated with more than 50% of observations. This means that no expression of spatio-temporal behaviour would likely lead to a specific outcome more often than chance.

The lack of metrics relating to the intrusion event reaffirmed that play outcomes were generally indistinguishable at the start of play. As noted in previous chapters, this implies that scenarios at the intrusion are not indicative of the conclusion. Both teams can take solace in this idea and work toward their preferred outcome regardless of the state of the intrusion. For future research, an alternative definition of a play should be sought whose starting point can provide some certainty about the outcome.

It was noted that passed Circle Entry outcomes did not distinguish between successful or intercepted passes so long as the ball entered the Circle. A portion of Circle entries were therefore actually turnovers but discussions with the collaborating partner suggested that penetrating the Circle is still a positive outcome because it at least offers an opportunity for scoring. Metrics involving the location of the ball, the outcome event or the possessing player were biased to the pitch-length location of the 23 m line, i.e. 68.4 m, because 62% of ‘Other’ outcomes were 23 m Exit events. Similarly, metrics relating to the Euclidean distance between the goal centre and the ball, Circle entries or the possessing player were biased to the approximate radius of the Circle, i.e. 14.63 m.

Circle entries and turnovers were distinguished by passing execution and pitch-length distance to the goal line. Distinguishing Circle entries and ‘Other’ outcomes by the location of the outcome event is likely to be a consequence of bias introduced by including metrics that indirectly define the outcome groups. The Other-Turnover comparison was the only one to include a metric relating to the start of the play. Findings support previous studies with evidence of a tactical preference for left-sided attacks.

Chapter 9 Conclusion

This chapter summarises the current work and discusses practical implications. The chapter concludes with a note on the suitability of genetic analytics-inspired methods for sports performance analysis and provides suggestions for future work.

9.1 Summary of work

The aim of the current work was to determine which spatio-temporal metrics distinguish play outcomes in field hockey. Spatio-temporal metrics can be used to investigate tactical intentions and methods. These metrics measure the cooperative and adversarial interactions between players using information about their locations over time.

With institutional ethical approval, the project's collaborators, England Hockey, provided video of games from the Men's tournament of the EuroHockey Championships 2015. Player locations at the start and end of a play were manually digitised using in-house software and data were divided among three outcome groups: Circle Entry, Turnover conceded, and 'Other' outcomes. This spatio-temporal data informed over 3,639 spatio-temporal metrics.

The Backward Dropping Algorithm is a method used in genetic analytics that addresses variable interactions and was suited to the current work's dataset, which constituted many variables with relatively few observations. The algorithm required the outcome variable to be binary and the inputs to be discrete. Observations were divided into three binary comparisons based on possible pairings of play outcomes and each metrics' data were subjected to a clustering algorithm to discretise values. Some metrics were removed from the dataset because their discretised values were uninformative for distinguishing play outcomes.

The first stage of selecting distinguishing metrics evaluated metric pairs' association with the outcome. The method of Wang et al. (2012) used the ***I***-score as a measure of association and applied two thresholds to reduce the number of metrics by two orders of magnitude. Metrics relating to the offensive team at the outcome dominated the selections across all comparisons. The method's sensitivity to threshold choice was examined and resulted in 21 unique datasets being carried forward for subsequent analysis.

The Backward Dropping Algorithm was applied to all 21 datasets and required the application of another threshold. The algorithm's sensitivity to threshold choice was examined and resulted in 120 datasets being carried forward. The behaviour of the current work's data did not resemble that of genetic data for which the method was originally designed but explanations were provided.

Chernoff et al.'s (2009) resuscitation analysis was applied to all 120 datasets to give a second chance to the metrics that did not perform well thus far in the analysis. Again, the behaviour of the current work's data did not resemble that of simulations on which the method was originally tested. The complete analysis reintroduced metrics relating to defensive team and the start of play, which complicated the relationship between spatio-temporal behaviour and play outcome. Circle entries and turnovers were distinguished by passing execution and pitch-length distance to the goal line. Circle entries and ‘Other’ outcomes were distinguished by the location of the outcome event but this was likely to be a consequence of bias introduced by including metrics that indirectly define the outcome groups. The Other-Turnover comparison was the only one to include a metric relating to the start of the play and findings supported previous studies with evidence of a tactical preference for left-sided attacks. Findings also provided empirical support for tactical preconceptions and suggested the method used in the current work can identify distinguishing metric combinations.

9.2 Limitations

Although the current work might be helpful, there are several philosophical and methodological limitations. Firstly, the reader must acknowledge that it is not necessarily sufficient to copy behaviour without copying motivation. Even if athletes follow the suggestions of the current work, it is not certain that they will achieve their expectations. The advantage of spatio-temporal metrics is that they are non-invasive but this very characteristic means that they cannot inform us of player intentions, despite measuring the result of them. The metrics used in the current work will not have described all relevant variables so coaches and athletes should include the findings in their tactical considerations rather than think of them as 'magic bullets'. This highlights another limitation: the findings are constrained by metrics provided. Efforts were made to incorporate as many metrics as possible from relevant literature, but some did not suit the research design.

Interpretation of the findings are also limited to the dataset used. Only 660 plays involving the full complement of players were observed from five games of one continental tournament. These observations only considered six teams, all of which were Men's. Potential differences in playing style and tactics would make it inappropriate to generalise to other teams (Stöckl & Morgan 2013), the Women's game (Mosquera et al. 2007), and scenarios with unequal player numbers. The culture within field hockey for regular rule changes (Tromp & Holmes 2011) also makes any research findings relatively more transient than other sports.

9.3 Future work

The methods used in the current work are still relatively new and, as far as the author is aware, have not been applied outside of genomic selection problems. In addition to testing the methods' generalisability, several problems still exist that need addressing in the coming years:

- Development of a more objective and theoretically justified method to determine the many required thresholds.
- Restructuring the method to handle more than two outcomes.
- Assessing the effect of including procedures for false positive removal and between-return set correlation in the resuscitation analysis.
- Begin discussions with domain experts to translate findings into useful tactical advice.

- Carry forward the current results to boosting ensemble regression models (Wang et al. 2012), network analyses of metrics within return sets (Lo et al. 2008), and decision tree analysis (Adams et al. 2013; Morgan et al. 2013) with the intention of empirically suggesting action rules (Gréhaigne & Godbout 1995; Gréhaigne 1996; Turner et al. 2001).

9.4 Conclusions

The current work produced lists of spatio-temporal metrics that distinguished pairs of play outcomes and might infer differences in tactics. Some of the metric combinations are almost obvious. If metrics do seem obvious, then users can be contented to learn that their preconceptions have empirical support. If metric combinations are surprising, then they provide new tactical possibilities. These statements exemplify the two ways that the current work's output can be used. The first is to take a deductive approach by starting with previous held beliefs about ‘good’ and ‘bad’ tactics, describing them with spatio-temporal metrics and checking them against the current work's final lists. The second way is to discuss the current work's metric combinations with coaches and athletes to translate the descriptions of player behaviour into potentially useful tactics.

The literature review highlighted some advantages of applying genetic analytics methods to the current work, such as the capacity to handle many variables, few observations, non-linearity, and non-parametric data. The methods used in the current work are also, in the author's opinion, less complicated than other statistical techniques and easier to explain to those who are unfamiliar with them. These are valuable attributes to a performance analyst who tries to provide the best quality insights and convince coaches and athletes to accept them. Choosing parameters is, however, non-trivial and thus a barrier to widespread use, and the method's computing resource requirement is beyond what might be available to most sports teams.

These methods still have their place in sports performance analysis outside of short-term analysis, like feedback during the game. Between games and seasons, strategies, tactics and training can be informed by thorough, possibly time-consuming analysis that has long-term and generalizable goals in mind. The methods used in the current work lend themselves to this kind of analysis.

9.5 Contribution to knowledge

The key finding of this thesis is that play outcomes in field hockey are distinguished by proximity to the goal and passing execution. The following contributions to knowledge have also been made:

1. A suite of metrics combinations that distinguish play outcomes in field hockey.
2. Empirical support for some tactical preconceptions.
3. The first application of genetic analytics-inspired methods to tactical sports performance analysis.
4. An improved awareness of the practical difficulties in applying the Backward Dropping Algorithm.

References

Abdel-Aziz, Y.I. & Karara, H.M., 1971. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *ASP Symposium on Close-Range Photogrammetry*. Falls Church, VA, USA, pp. 1–18.

Abdel-Hakim, H.H., 2014. Quantitative analysis of performance indicators of goals scored in the futsal World Cup Thailand 2012. *Pamukkale Journal of Sport Sciences*, 5(1), pp.113–127. Available at: http://psbd.pau.edu.tr/index.php/pjss/article/view/190.

Abdelkrim, N. Ben, El Fazaa, S. & El Ati, J., 2007. Time-motion analysis and physiological data of elite under-19-year-old basketball players during competition. *British Journal of Sports Medicine*, 41, pp.69–75.

Abidin, H.Z., 2010. Fundamentals of GPS signals and Data. In J. D. Bossler et al., eds. *Manual of Geospatial Science and Technology*. CRC Press.

Abraldes, J.A. et al., 2012. Tactical and shooting variables that determine the difference between win or loss in top-level female water polo : analysis by phases of the game. *International Journal of Performance Analysis in Sport*, 12(2), pp.373–384.

Abreu, P. et al., 2012. Human versus virtual robotics soccer : A technical analysis. *European Journal of Sport Science*, 12(1), pp.26–35.

Adams, D. et al., 2013. Successful short passing frequency of defenders differentiates between top and bottom four English Premier League teams. *International Journal of Performance Analysis in Sport*, 13(3), pp.653–668.

Adobe, 2013. Photoshop CS6.

Albinsson, P.-A. & Andersson, D., 2008. Extending the attribute explorer to support professional team-sport analysis. *Information Visualization*, 7, pp.163–169. Available at: http://ezproxy.lib.ucf.edu/login?url=http://search.proquest.com/docview/336

07955?accountid=10003%5Cnhttp://sfx.fcla.edu/ucf?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ:computerinfo&atitle=Extending+the+attribute+.

Alcaraz, P. et al., 2012. Relationship between characteristics of water polo players and efficacy indices. *Journal of Strength and Conditioning Research*, 26(7), pp.1852–1857.

Alcaraz, P.E. et al., 2011. Throwing velocities, anthropometric characteristics, and efficacy indices of women's European water polo subchampions. *Journal of Strength and Conditioning Research*, 25(11), pp.3051–3058.

Almeida, C.H., Ferreira, A.P. & Volossovitch, A., 2013. Offensive sequences in youth soccer: effects of experience and small-sided games. *Journal of human kinetics*, 36(March), pp.97–106. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3661899&tool=pmcentrez&rendertype=abstract.

Amari, S.-I. et al., 1995. Asymptotic Statistical Theory of Overtraining and Cross Validation. *IEEE Transactions on Neural Networks*, 8(5), pp.985–996.

American Football Coaches Association, 2000a. *Defensive football strategies*, Chicago, IL, USA: Human Kinetics.

American Football Coaches Association, 2000b. *Offensive football strategies*, Chicago, IL, USA: Human Kinetics.

Amjad, I., Hussain, I. & Asadullah, M., 2013. Comparison between long corners and short corners in field hockey. *Rawal Medical Journal*, 38(4), pp.428–431.

Arthur, D. & Vassilvitskii, S., 2007. k-means++: The advantages of careful seeding. In *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithims*. New Orleans, LA, pp. 1027–1035.

Atkinson, G. & Nevill, A.M., 2001. Selected issues in the design and analysis of sport performance research. *Journal of Sports Sciences*, 19(10), pp.811–

827.

Barreira, D., Garganta, J., et al., 2013. Ball recovery patterns as a performance indicator in elite soccer. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 228, pp.61–72. Available at: http://pip.sagepub.com/lookup/doi/10.1177/1754337113493083.

Barreira, D., Garganta, J., et al., 2013. SoccerEye: A software solution to observe and record behaviours in sport settings. *The Open Sports Sciences Journal*, 6, pp.47–55.

Barros, R.M. et al., 2006. Representation and analysis of soccer players' actions using principal components. *Journal of Human Movement Studies*, 51, pp.103–116.

Bartlett, J.W. & Frost, C., 2008. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics and Gynecology*, 31(4), pp.466–475.

Bartlett, R. et al., 2012. Analysing Team Coordination Patterns from Player Movement Trajectories in Soccer: Methodological Considerations. *International Journal of Performance Analysis in Sport*, 12(2), pp.398–424.

Batterham, A.M. & Hopkins, W.G., 2006. Making meaningful inferences about magnitudes. *International Journal of Sports Physiology and Performance*, 1(1), pp.50–57.

Beckerman, H. et al., 2001. Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research*, 10(7), pp.571–578.

Bialkowski, A. et al., 2013. Recognising team activities from noisy data. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp.984–990.

Bishop, C.M., 1995. *Neural networks for pattern recognition*, Oxford: Oxford

University Press.

Bland, J.M. & Altman, D.G., 1996. Statistics notes: Measurement error. *British Medical Journal*, 312(7047), p.1654.

Bouguet, J.Y., 2010. Camera Calibration Toolbox for MATLAB. Available at: http://www.vision.caltech.edu/bouguetj/calib_doc/ [Accessed October 1, 2015].

Bourbousson, J., Sève, C. & McGarry, T., 2010a. Space-time coordination dynamics in basketball: Part 2. The interaction between the two teams. *Journal of Sports Sciences*, 28(3), pp.349–358.

Bourbousson, J., Sève, C. & McGarry, T., 2010b. Space–time coordination dynamics in basketball: Part 1. Intra- and inter-couplings among player dyads. *Journal of Sports Sciences*, 28(3), pp.339–347. Available at: http://www.tandfonline.com/doi/abs/10.1080/02640410903503632%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/20131146.

Bradley, J. & O'Donoghue, P.G., 2011. Counterattacks in elite Gaelic football competition. *International Journal of Performance Analysis in Sport*, 11, pp.159–170. Available at: http://www.ingentaconnect.com/content/uwic/ujpa/2011/00000011/00000001/art00015.

Bradley, P.S. et al., 2013. The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. *Journal of Sports Sciences*, 31(12), pp.1261–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23697463.

Bremner, S., Robinson, G. & Williams, M.D., 2013. A retrospective evaluation of team performance indicators in rugby union. *International Journal of Performance Analysis in Sport*, 2013, pp.461–473.

Brillinger, D.R., 2007. A Potential Function Approach to the Flow of Play in Soccer. *Journal of Quantitative Analysis in Sports*, 3(1).

Brown, J.D., 1999. Standard error vs . Standard error of measurement. *JALT Testing & Evaluation SIG Newsletter*, 3(April), pp.20–25.

Buntine, W.L. & Weigend, A.S., 1991. Bayesian back-propagation. *Complex Systems*, 5, pp.603–643.

Burns, R. & Burns, R., 2008. Discriminant analysis. In *Business research methods and statistics using SPSS*. Sage, pp. 589–608. Available at: http://doi.wiley.com/10.1111/j.1600-0404.1986.tb04441.x.

Van Calster, B., Smits, T. & Van Huffel, S., 2008. The Curse of Scoreless Draws in Soccer: The Relationship with a Team's Offensive, Defensive, and Overall Performance. *Journal of Quantitative Analysis in Sports*, 4(1).

Cameron, W.B., 1963. *Informal Sociology: A casual introduction to sociological thinking.*, New York: Random House.

Carling, C., Williams, M.A. & Reilly, T., 2005. *Handbook of soccer match analysis: A systematics approach to improving performance*, Oxon: Routledge.

Carroll, R., 2013. Team performance indicators in Gaelic football and opposition effects. *International Journal of Performance Analysis in Sport*, 13(3), pp.703–715.

Casal, C.A. et al., 2015. Analysis of Corner Kick Success in Elite Football. *International Journal of Performance Analysis in Sport*, 15(2), pp.430–451.

Casals, M. & Martinez, J.A., 2013. Modelling player performance in basketball through mixed models. *International Journal of Performance Analysis in Sport*, 13(1), pp.64–82. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84876382745&partnerID=tZOtx3y1.

Castellano, J. et al., 2013. Identifying the effects from the quality of opposition in a football team positioning strategy. *Int J Perf Anal Sport*, 13(3), pp.822–832.

Chan, V., 2011. Prediction Accuracy of Linear Models for Paired Comparisons in Sports. *Journal of Quantitative Analysis in Sports*, 7(3).

Chernoff, H., Lo, S.H. & Zheng, T., 2009. Discovering influential variables: A method of partitions. *Annals of Applied Statistics*, 3(4), pp.1335–1369.

Choi, H.G., O'Donoghue, P.G. & Hughes, M.D., 2008. The identification of an optimal set of performance indicators for real-time analysis using principle components analysis. In A. Hökelmann & M. Brummond, eds. *Performance Analysis of Sport VIII2*. pp. 295–301.

Choppin, S. et al., 2007. 3D player testing at the Wimbledon qualifying tournament. In S. Miller & J. Capel-Davies, eds. *Tennis science and technology 3*. International tennis federation, pp. 333–340.

Choppin, S. & Wheat, J., 2013. The potential of the Microsoft Kinect in sports analysis and biomechanics. *Sports Technology*, 6(2), pp.78–85. Available at: http://www.tandfonline.com/doi/abs/10.1080/19346182.2013.819008.

Clemente, F. et al., 2012. Influence of task constraints on attacker trajectories during 1V1 sub-phases in soccer practice. *Sportlogia*, 8(1), pp.13–20. Available at: http://www.sportlogia.com/no5engl/2.html.

Clemente, F.M. et al., 2013. An online tactical metrics applied to football game. *Research Journal of Applied Sciences, Engineering and Technology*, 5(5), pp.1700–1719.

Clemente, F.M., Martins, F.M.L., et al., 2014a. Developing a football tactical metric to estimate the sectorial lines: A case study. In B. Murgante et al., eds. *Computational Science and Its Applications – ICCSA 2014: 14th International Conference, Guimarães, Portugal, June 30 – July 3, 2014, Proceedings, Part I*. Guimarães, Portugal: Springer International Publishing, pp. 743–753.

Clemente, F.M. et al., 2015. Developing a tactical metric to estimate the defensive area of soccer teams: The defensive play area. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports*

Engineering and Technology, pp.1–9. Available at: http://pip.sagepub.com/lookup/doi/10.1177/1754337115583198.

Clemente, F.M., Martins, F.M.L., et al., 2014b. Evaluating the offensive definition zone in football: A case study. *South African Journal for Research in Sport, Physical Education and Recreation*, 36(3), pp.25–37.

Clemente, F.M. et al., 2013. Measuring tactical behaviour using technological metrics: Case study of a football game. *International Journal of Sports Science and Coaching*, 8(4), pp.723–739. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84891365282&partnerID=40&md5=7bfa1fed48591b96f70fb4ae3fdc3055.

Clemente, F.M., Couceiro, M.S., et al., 2014. Practical implementation of computational tactical metrics for the football game: Towards an augmenting perception of coaches and sport analysts. In B. Murgante et al., eds. *Computational Science and Its Applications – ICCSA 2014: 14th International Conference, Guimarães, Portugal, June 30 – July 3, 2014, Proceedings, Part I*. Guimarães, Portugal: Springer International Publishing, pp. 712–727.

Clemente, F.M. et al., 2014. Using collective metrics to inspect spatio-temporal relationships between football players. *South African Journal for Research in Sport, Physical Education and Recreation*, 36(2), pp.47–59.

Clemente, F.M., Couceiro, M.S. & Martins, F.M.L., 2013. Soccer teams' behaviors: Analysis of the team's distribution in function to ball possession. *Research Journal of Applied Sciences, Engineering and Technology*, 6(1), pp.130–136.

Clemente, F.M., Couceiro, M.S. & Martins, F.M.L., 2012. Towards a new method to analyze the soccer teams tactical behaviour: Measuring the effective area of play. *Indian Journal of Science and Technology*, 5(12), pp.3792–3801.

Collet, C., 2013. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *Journal of Sports Sciences*, 31(2), pp.123–136.

Condon, E.M., Golden, B.L. & Wasil, E.A., 1999. Predicting the success of nations at the Summer Olympics using neural networks. *Computers and Operations Research*, 26(13), pp.1243–1265.

Connelly, C.D., 2013. *British handball: How can performance analysis aid the coaching process?* University of Chester. Available at: http://hdl.handle.net/10034/320711.

Cordes, O., Lamb, P.F. & Lames, M., 2012. Concepts and Methods for Strategy Building and Tactical Adherence: A Case Study in Football. *International Journal of Sports Science and Coaching*, 7(2), pp.241–254.

Del Corral, J., Barros, C.P. & Prieto-Rodríguez, J., 2008. The Determinants of Soccer Player Substitutions: A Survival Analysis of the Spanish Soccer League. *Journal of Sports Economics*, 9(2), pp.160–172. Available at: http://jse.sagepub.com/cgi/content/abstract/9/2/160.

Correia, V. et al., 2013. From recording discrete actions to studying continuous goal-directed behaviours in team sports. *Journal of sports sciences*, 31(June 2014), pp.546–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23140581.

da Costa, I.T. et al., 2009. Tactical Principles of Soccer Game: concepts and application. *Revista Motriz*, 15(3), pp.657–668. Available at: https://scholar.google.com.br/scholar?hl=pt-BR&q=comportamento+t�tico+-+Israel+teoldo+da+costa&btnG=&lr=#9.

Cotta, C. et al., 2013. A network analysis of the 2010 FIFA world cup champion team play. *Journal of Systems Science and Complexity*, 26, pp.21–42.

Couceiro, M.S. et al., 2014. Dynamical stability and predictability of football players: The study of one match. *Entropy*, 16, pp.645–674.

Coutts, A.J. & Duffield, R., 2010. Validity and reliability of GPS devices for measuring movement demands of team sports. *Journal of Science and Medicine in Sport*, 13, pp.133–135.

Crego, R., 2003. *Sports and Games of the 18th and 19th Century*, Westport: Greenwood Publishing Group.

Crum, P., 2013. *Space creation dynamics in basketball: A comparison between British and Spanish leagues*. University of Chester. Available at: http://hdl.handle.net/10034/320714.

Davids, K. et al., 2013. How small-sided and conditioned games enhance acquisition of movement and decision-making skills. *Exercise and sport sciences reviews*, 41(3), pp.154–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23558693.

Dawson, B., Appleby, B. & Stewart, G., 2004. Analysis of a 16-game winning streak in Australian rules football. *Journal of Sports Sciences*, 22(6), pp.503–504.

Diedrick, E. & van Rooyen, M.K., 2011. Line break situations in international rugby. *International Journal of Performance Analysis in Sport*, 11(3), pp.522–534. Available at: http://www.ingentaconnect.com/content/uwic/ujpa/2011/00000011/00000003/art00012.

Dreiseitl, S. & Ohno-Machado, L., 2002. Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, 35(5–6), pp.352–359.

Duarte, R. et al., 2013. Capturing complex, non-linear team behaviours during competitive football performance. *Journal of Systems Science and Complexity*, 26(1), pp.62–72.

Duarte, R. et al., 2010. Capturing complex human behaviors in representative sports contexts with a single camera. *Medicina (Kaunas, Lithuania)*, 46(6), pp.408–414.

Duarte, R. et al., 2012. Interpersonal coordination tendencies shape 1-vs-1 sub-phase performance outcomes in youth soccer. *Journal of Sports Sciences*, 30(9), pp.871–877.

Duarte, R. et al., 2010. The ecological dynamics of 1v1 sub-phases in association football. , pp.16–18. Available at: http://shura.shu.ac.uk/3343/.

Dunn, M.D., 2014. *Video-based step measurement in sport and daily living*. Sheffield Hallam University.

Dutt-Mazumder, A. et al., 2011. Neural network modelling and dynamical system theory: Are they relevant to study the governing dynamics of association football players? *Sports Medicine*, 41(12), pp.1003–1017.

Eames, M.H.A., Cosgrove, A. & Baker, R., 1999. Comparing methods of estimating the total body centre of mass in three-dimensions in normal and pathological gaits. *Human Movement Science*, 18(5), pp.637–646.

Eaves, S.J. & Evers, A.L., 2007. The relationship between the “play the ball” time, post-ruck action and the occurrence of perturbations in professional rugby league football. *International Journal of Performance Analysis of Sport*, 7(3), pp.18–25.

Elferink-Gemser, M.T. et al., 2004. Development of the Tactical Skills Inventory for Sports. *Perceptual and motor skills*, 99, pp.883–895.

Elferink-Gemser, M.T. et al., 2007. Multidimensional performance characteristics and standard of performance in talented youth field hockey players: a longitudinal study. *Journal of sports sciences*, 25(April 2013), pp.481–489.

Elferink-Gemser, M.T. et al., 2011. Relation between multidimensional performance characteristics and level of performance in talented youth field hockey players. *Journal of sports sciences*, 22(April 2013), pp.1053–1063.

Elferink-Gemser, M.T., 2005. *Today’s talented youth field hockey players, the stars of tomorrow? A study on talent development in field hockey*.

University of Groningen.

Ensum, J., Pollard, R. & Taylor, S., 2004. Applications of logistic regression to shots at goal in association footbal: calculation of shot probabilitoes, quantification of factors and player/team. *Journal of Sports Sciences*, 22(6), pp.500–520.

Escalante, Y. et al., 2011. Discriminatory power of water polo game-related statistics at the 2008 Olympic Games. *Journal of sports sciences*, 29(3), pp.291–298.

Esteves, P.T. et al., 2012. Interpersonal dynamics and relative positioning to scoring target of performers in 1 vs. 1 sub-phases of team sports. *Journal of Sports Sciences*, 30(12), pp.1285–1293.

Field, A., 2009. *Discovering statistics using SPSS*, Londn: Sage.

FIH, 2016. FIH Men’s and Women’s World Ranking. Available at: http://www.fih.ch/rankings/outdoor/ [Accessed October 27, 2016].

FIH, 2015. Rules of Hockey. *Rules of Hockey including explanations*, pp.1–72.

Folgado, H. et al., 2014. Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS ONE*, 9(5), p.e97145.

Folgado, H. et al., 2012. Length, width and centroid distance as measures of teams tactical performance in youth football. *European Journal of Sport Science*, (April 2015), pp.1–6.

Fonseca, S. et al., 2012. Spatial dynamics of team sports exposed by Voronoi diagrams. *Human Movement Science*, 31(6), pp.1652–1659. Available at: http://dx.doi.org/10.1016/j.humov.2012.04.006.

Foretić, N., Rogulj, N. & Papić, V., 2013. Empirical model for evaluating situational efficiency in top level handball. *International Journal of Performance Analysis in Sport*, 13(2), pp.275–293.

Frencken, W. et al., 2011. Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science*, 11(4), pp.215–223.

Frencken, W. et al., 2012. Variability of inter-team distances associated with match events in elite-standard soccer. *Journal of Sports Sciences*, 30(12), pp.1207–1213.

Geman, S., Bienenstock, E. & Doursat, R., 1992. Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4(1), pp.1–58.

Glazier, P. & Irwin, G., 2001. Validity of stride length estimates obtained from optojump. *Biomechanics Symposia 2001*, pp.98–101.

Gómez, M.Á., Lorenzo, A., et al., 2013. Ball possession effectiveness in men's and women's elite basketball according to situational variables in different game periods. *Journal of Sports Sciences*, 31(14), pp.1578–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23679867.

Gómez, M.Á., Prieto, M., et al., 2013. Ball possession effectiveness in men's elite floorball according to quality of opposition and game period. *Journal of human kinetics*, 38(September), pp.227–37. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3827758&tool=pmcentrez&rendertype=abstract.

Gómez, M.Á., Evangelos, T. & Lorenzo, A., 2006. Defensive systems in basketball ball possessions. *International Journal of Performance Analysis in Sport*, 6(1), pp.98–107.

Gonçalves, B. V et al., 2014. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *Journal of Sports Sciences*, 32(2), pp.191–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24016056.

Gonzaga, A.D.S. et al., 2014. Affective decision-making and tactical behavior of under-15 soccer players. *PLoS ONE*, 9(6), p.e101231.

Graham, J. & Mayberry, J., 2014. Measures of tactical efficiency in water polo. *Journal of Quantitative Analysis in Sports*, 10(1), pp.67–79.

Gramacy, R.B., Jensen, S.T. & Taddy, M., 2013. Estimating player contribution in hockey with regularized logistic regression. *Journal of Quantitative Analysis in Sports*, 9, pp.97–111.

Gréhaigne, J.-F., 1996. Les règles d'action : un support pour les apprentissages. *Revue EPS*, 260, pp.71–73.

Gréhaigne, J.-F. & Godbout, P., 2013. Collective Variables for Analysing Performance in Team Sports. In T. McGarry, P. G. O'Donoghue, & J. E. Sampaio, eds. *Routledge handbook of sports performance analysis2*. Oxon: Routledge, pp. 101–114.

Gréhaigne, J.-F. & Godbout, P., 1995. Tactical Knowledge in Team Sports From a Constructivist and Cognitivist Perspective. *Quest*, 47(4), pp.490–505.

Gréhaigne, J.-F., Godbout, P. & Bouthier, D., 1999. The foundations of tactics and strategy in team sports. *Journal of Teaching in Physical Education*, 18, pp.159–174.

Gudmundsson, J. & Horton, M., 2016. Spatio-Temporal Analysis of Team Sports - A Survey. *arXiv preprint arXiv:1602.06994*. Available at: http://arxiv.org/abs/1602.06994.

Gudmundsson, J. & Wolle, T., 2014. Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, 47, pp.16–27.

Harrop, K. & Nevill, A.M., 2014. Performance indicators that predict success in an English professional League One soccer team. *International Journal of Performance Analysis in Sport*, 14, pp.907–920.

Hawkins, R., 2011. Identifying mechanic measures that best predict air-pistol shooting performance. *International Journal of Performance Analysis in Sport*, 11(1990), pp.499–509.

Haws, D.C. et al., 2015. Variable-selection emerges on top in empirical comparison of whole-genome complex-trait prediction methods. *PLoS ONE*, 10(10), pp.1–22. Available at: http://dx.doi.org/10.1371/journal.pone.0138903.

Headrick, J. et al., 2012. Proximity-to-goal as a constraint on patterns of behaviour in attacker–defender dyads in team games. *Journal of Sports Sciences*, 30(3), pp.247–253.

Hewitt, A., Greenham, G. & Norton, K., 2016. Game style in soccer: what is it and can we quantify it? *International Journal of Performance Analysis in Sport*, pp.355–372.

Higham, D.G. et al., 2014. Performance indicators related to points scoring and winning in international rugby sevens. *Journal of Sports Science and Medicine*, 13, pp.358–364.

Hinrichs, R.N. et al., 2005. Predicting out-of-plane point locations using the 2D-DLT. In *29th Annual meeting of the American Society of Biomechanics*. Portland, OR, USA, pp. 249–251.

Hirotsu, N. et al., 2006. Modeling Tactical Changes of Formation in Association Football as a Zero-Sum Game. *Journal of Quantitative Analysis in Sports*, 2(2).

Hopkins, W.G., 2000. Measures of reliability in sports medicine and science. *Sports medicine (Auckland, N.Z.)*, 30(1), pp.1–15.

Hopkins, W.G. et al., 2009. Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, 41(1), pp.3–12.

Hosmer, D.W., Lemeshow, S. & Sturdivant, R., 2013. *Applied logistic regression* 3rd ed., Hoboken: Wiley.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology1*, 24, pp.417–441.

Hraste, M., Dizdar, D. & Trninić, V., 2008. Experts opinion about system of the performance evaluation criteria weighted per positons in the water polo game. *Collegium antropologicum*, 32(3), pp.851–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18982761.

Hughes, M. & Franks, I., 2005. Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences*, 23(5), pp.509–514.

Hughes, M.D. & Bartlett, R.M., 2002. The use of performance indicators in performance analysis. *Journal of Sports Sciences*, 20(10), pp.37–41.

Hughes, M.D. & Churchill, S., 2004. Attacking profiles of successful and unsuccessful teams in Copa America 2001. *Journal of Sports Sciences*, 22(6), p.505.

International Olympic Committee, 2011. HOCKEY: participation during the history of the Olympic games – Reference document. *Olympic Studies Centre*. Available at: http://www.olympic.org/Assets/OSC Section/pdf/QR_sports_summer/Sports_Olympiques_hockey_eng.pdf [Accessed November 5, 2016].

Ito, K. et al., 2004. Effectiveness of changes to the laws of association football. In T. Reilly, J. Cabri, & D. Araújo, eds. *Journal of Sports Sciences*. Lisbon, Portugal: Routledge, pp. 506–507.

Jäger, J.M. & Schöllhorn, W.I., 2012. Identifying individuality and variability in team tactics by means of statistical shape analysis and multilayer perceptrons. *Human Movement Science*, 31(2), pp.303–317.

James, N., Jones, P. & Mellalieu, S.D., 2004. Possession as a performance indicator in soccer as a function of successful and unsuccessful teams. *Journal of Sports Sciences*, 22(6), pp.507–508.

James, N., Mellalieu, S.D. & Jones, N.M., 2005. The development of position-specific performance indicators in professional rugby union. *Journal of Sports Sciences*, 23(1), pp.63–72.

Janković, A. et al., 2011. Influence of tactics efficiency on results in serbian soccer super league in season 2009-2010. *Journal of Physical Education and Sport*, 11(1), pp.32–41.

Johnson, D., 1999. The insignificance of statistical significance testing. *The Journal of Wildlife Management*, 63, pp.763–772.

Jones, P., James, N. & Mellalieu, S.D., 2004. Possession as a performance indicator in soccer. *International Journal of Performance Analysis of Sport*, 4(1), pp.98–102.

Kan, A. et al., 2004. The movement of players and the team: Comparing two games, Japan versus UAE and a J-League. *Journal of Sports Sciences*, 22(6), pp.508–509.

Kelley, J., Higham, D. & Wheat, J., 2016. Technology for studying interpersonal coordination in social collectives. In P. Passos, K. Davids, & J. Y. Chow, eds. *Interpersonal coordination and performance in social systems*. Oxon: Routledge, pp. 305–322.

Kingston, G.B., Maier, H.R. & Lambert, M.F., 2006. A probabilistic method for assisting knowledge extraction from artificial neural networks used for hydrological prediction. *Mathematical and Computer Modelling*, 44(5–6), pp.499–512.

Kirk, R.G., 1996. Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), pp.746–759.

Kohonen, T., 1997. *Self-organising maps*, New York: Springer.

Kolar, M. & Liu, H., 2013. Feature Selection in High-Dimensional Classification. *Proc. of the 30th International Conference on Machine Learning*, 28, pp.329–337. Available at: http://jmlr.org/proceedings/papers/v28/kolar13.html.

Konarski, J., 2010. Characteristics of chosen parameters of external and internal loads in Eastern European high level field hockey players. *Journal* *of Human Sport & Exercise*, 5(1), pp.43–58.

Krause, J. & Pim, R.L., 2002. *Coaching basketball*, Chicago, IL, USA: Contemporary Books.

Kubatko, J. et al., 2007. A Starting Point for Analyzing Basketball Statistics. *Journal of Quantitative Analysis in Sports*, 3(3), pp.1–22.

Lago, C. et al., 2010. The effects of situational variables on distance covered at various speeds in elite soccer. *European Journal of Sport Science*, 10(2), pp.103–109.

Laird, P. & Sutherland, P., 2003. Penalty Corners in Field Hockey : A guide to success. *International Journal of Performance Analysis in Sport*, 3, pp.19–26.

Lames, M., 1991. *Leistungsdiagnostik durch Computersimulation: Ein Beitrag zur Theorie der Sportspiele am Beispiel Tennis*, Frankfurt: Deutsch Harri GmbH.

Lidor, R. & Ziv, G., 2015. On-field performances of female and male field hockey players – A review. *International Journal of Performance Analysis in Sport*, 15, pp.20–38.

Liu, H., Hopkins, W.G. & Gómez, M.-Á., 2015. Modelling relationships between match events and match outcome in elite football. *European Journal of Sport Science*, 16(6), pp.516–525. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84937597843&partnerID=tZOtx3y1.

Lo, A. et al., 2015. Why significant variables aren't automatically good predictors. *Proceedings of the National Academy of Sciences*, 112(45), pp.13892–13897. Available at: http://www.pnas.org/content/112/45/13892%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26504198%5Cnhttp://www.pnas.org/content/112/45/13892.full.

Lo, S.-H. et al., 2008. Discovering interactions among BRCA1 and other candidate genes associated with sporadic breast cancer. *Proc Natl Acad Sci U S A*, 105(34), pp.12387–12392.

Lo, S.-H. & Zheng, T., 2004. A demonstration and findings of a statistical approach through reanalysis of inflammatory bowel disease data. *Proceedings of the National Academy of Sciences of the United States of America*, 101(28), pp.10386–91. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=478581&tool=pmcentrez&rendertype=abstract.

Lo, S.-H. & Zheng, T., 2002. Backward haplotype transmission association (BHTA) algorithm - A fast multiple-marker screening method. *Human Heredity*, 53(4), pp.197–215.

Loh, W., 2012. Variable selection for classification and regression in large p, small n problems. *Probability Approximations and Beyond*, 205, pp.133–157. Available at: http://link.springer.com/chapter/10.1007/978-1-4614-1966-2_10.

Lorenzo, A. et al., 2010. Game related statistics which discriminate between winning and losing under-16 male basketball games. *Journal of Sports Science and Medicine*, 9, pp.664–668.

Lucey, P. et al., 2012. Representing and discovering adversarial team behaviors using as player roles. In *Proceedings o fthe IEEE Conference on Computer Vision and pattern Recognition*.

Lupo, C. et al., 2009. A physiological, time-motion, and technical comparison of youth water polo and Acquagoal. *Journal of Sports Sciences*, 27(June 28), pp.823–831.

Lupo, C., Minganti, C., et al., 2012. Effects of competition level on the centre forward role of men’s water polo. *Journal of Sports Sciences*, 30(9), pp.889–897.

Lupo, C. et al., 2011. Notational analysis of American women’s collegiate water polo matches. *Journal of Strength and Conditioning Research*, 25(3), pp.753–757.

Lupo, C., Condello, G. & Tessitore, A., 2012. Notational analysis of elite men's water polo related to specific margins of victory. *Journal of Sports Science and Medicine*, 11, pp.516–525.

MacLeod, H. et al., 2009. The validity of a non-differential global positioning system for assessing player movement patterns in field hockey. *Journal of sports sciences*, 27(July 2013), pp.121–128.

Massuça, L., Fragoso, I. & Teles, J., 2014. Attributes of top elite team-handball players. *Journal of Strength and Conditioning Research*, 28(1), pp.178–186. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84892973793&partnerID=40&md5=eefd8a6dc3dd0c00842caaa3d83c217f.

Maszczyk, A., Zając, A. & Ryguła, I., 2010. A neural network model approach to athlete selection. *Sports Engineering*, 13(2), pp.83–93. Available at: http://www.springerlink.com/index/10.1007/s12283-010-0055-y%5Cnpapers2://publication/doi/10.1007/s12283-010-0055-y.

MathWorks, 2015. MATLAB R2015b.

MathWorks, 2016. MATLAB R2016a.

Mauthner, T. et al., 2008. Visual Tracking of Athletes in Beach Volleyball Using a Single Camera. *International Journal of Computer Science in Sport*, 6(2), pp.21–34.

Memmert, D., Lemmink, K.A. & Sampaio, J.E., 2017. Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, 47(1), pp.1–10.

Meyer, D., Forbes, D. & Clarke, S.R., 2006. Statistical analysis of notational AFL data using continuous time Markov chains. *Journal of Sports Science and Medicine*, 5, pp.525–532.

Mitchell, J.H. et al., 2005. Task force 8: Classification of sports. *Journal of the American College of Cardiology*, 45(8), pp.1364–1367.

Moody, J.E., 1992. The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems. In J. E. Moody, S. J. Hanson, & R. P. Lippmann, eds. *Advances in Neural Information Processing Systems 4*. San Mateo, CA: Morgan Kaufman, pp. 847–854.

Morgan, S., Williams, M.D. & Barnes, C., 2013. Applying decision tree induction for identification of important attributes in one-versus-one player interactions: A hockey exemplar. *Journal of Sports Sciences*, 31(10), pp.1031–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23409787.

Mosquera, R., Molineuvo, J. & Román, I., 2007. Differences between international men´s and women´s teams in the strategic action of the penalty corner in field hockey. *International Journal of Performance Analysis in Sport*, 7, pp.67–83.

MotionAnalysis, 2016. MotionAnalysis. Available at: http://www.motionanalysis.com/ [Accessed November 11, 2016].

Moura, F.A. et al., 2013. A spectral analysis of team dynamics and tactics in Brazilian football. *Journal of Sports Sciences*, 31(14), pp.1568–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23631771.

Moura, F.A. et al., 2016. Coordination analysis of players’ distribution in football using cross-correlation and vector coding techniques. *Journal of Sports Sciences*. Available at: http://www.tandfonline.com/doi/full/10.1080/02640414.2016.1173222.

Moura, F.A. et al., 2012. Quantitative analysis of Brazilian football players’ organisation on the pitch. *Sports Biomechanics*, 11(1), pp.85–96.

Moura, F.A., Martins, L.E.B. & Cunha, S.A., 2014. Analysis of football game-related statistics using multivariate techniques. *Journal of Sports Sciences*, 32(20), pp.1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24742152.

Müller, R. & Büttner, P., 1994. A critical discussion of intraclass correlation coefficients. *Statistics in medicine*, 13(23–24), pp.2465–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7701147%5Cnhttp://onlinelibrary.wiley.com.libproxy1.nus.edu.sg/doi/10.1002/sim.4780132310/abstract%5Cnhttp://onlinelibrary.wiley.com.libproxy1.nus.edu.sg/doi/10.1002/sim.4780132310/abstract;jsessionid=2AE6ABBF76A8E4327AE50FC.

Nakaya, A. & Isobe, S.N., 2012. Will genomic selection be a practical method for plant breeding? *Annals of Botany*, 110(6), pp.1303–1316.

Neal, R.M., 1995. *Bayesian Learning for Neural Networks*. University of Toronto. Available at: http://www.jstor.org/stable/2965731?origin=crossref.

Nevill, A.M., Atkinson, G. & Hughes, M., 2008. Twenty-five years of sport performance research in the Journal of Sports Sciences. *Journal of Sports Sciences*, 26(4), pp.413–426.

Nevo, D. & Ritov, Y., 2013. Around the goal: Examining the effect of the first goal on the second goal in soccer using survival analysis methods. *Journal of Quantitative Analysis in Sports*, 9(2), pp.165–177.

O'Donoghue, P., 2008. Principal Components Analysis in the selection of Key Performance Indicators in Sport. *International Journal of Performance Analysis in Sport*, 8(3), pp.145–155. Available at: http://www.ingentaconnect.com/content/uwic/ujpa/2008/00000008/00000003/art00015.

O'Donoghue, P.G., 2007. Reliability Issues in Performance Analysis. *International Journal of Performance Analysis in Sport*, 7(1), pp.35–48.

O'Donoghue, P.G., 2010. *Research Methods for Sports Performance Analysis*, Oxon: Routledge.

Okihara, K. et al., 2004. Compactness as a strategy in a soccer match in relation to a change in offence and defence. *Journal of Sports Sciences*, 22(6), p.515.

Olden, J. D., Joy, M. K., & Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178(3–4), pp.389–397.

Olden, J.D. & Jackson, D.A., 2002. Illuminating the “black box”: a ramdomization approach for understanding variable contributions in artificial neuronal networks. *Ecological Modelling*, 154, pp.135–150.

Olsen, E., Larsen, O. & Semb, N.J., 1994. *Effektiv fotball*, Oslo: Gyldendal Norsk Forlag.

Orejan, J., 2011. *Football/Soccer*, Jefferson, NC, USA: McFarland.

Ortega, E., Villarejo, D. & Palao, J.M., 2009. Differences in game statistics between winning and losing rugby teams in the six nations tournament. *Journal of Sports Science and Medicine*, 8, pp.523–527.

Passos, P., Araújo, D. & Davids, K., 2013. Self-organization processes in field-invasion team sports: Implications for leadership. *Sports Medicine*, 43, pp.1–7.

Pearson, K., 1901. On lines and planes of closest fit to a system of points in space. *Philosophy Magazine*, 2, pp.557–572.

Peinado, A.B. et al., 2011. Discriminant analysis of the speciality of elite cyclists. *Journal of Human Sport and Exercise*, 6(3), pp.480–489.

Peng, H.C., Long, F.H. & Ding, C., 2005. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), pp.1226–1238.

Perl, J., Grunz, A. & Memmert, D., 2013. Tactics analysis in soccer - An advanced approach. *International Journal of Computer Science in Sport*, 12(1), pp.33–44.

Perneger, T. V., 1998. What’s wrong wirth Bonferroni adjustments. *British Medical Journal*, 316(1236).

Perš, J. et al., 2002. Observation and analysis of large-scale human motion. *Human movement science*, 21(2), pp.295–311. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12167304.

Pfeiffer, M. & Perl, J., 2006. Analysis of Tactical Structures in Team Handball by Means of Artificial Neural Networks. *International Journal of Computer Science in Sport*, 5, pp.4–14.

Platanou, T., 2004. Time-motion analysis of international level water polo players. *Journal of Human Movement Studies*, 46, pp.319–331.

Podgórski, T. & Pawlak, M., 2011. A half century of scientific research in field hockey. *Human Movement*, 12(2), pp.108–123.

Pollard, R., 1985. Goal-Scoring and the Negative Binomial Distribution. *Mathematical Gazette*, 69(447), pp.45–47.

Pollard, R. & Reep, C., 1997a. Measuring the Effectiveness of Playing Strategies at Soccer. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(4), pp.541–550. Available at: http://www.jstor.org/stable/2988603.

Pollard, R. & Reep, C., 1997b. Measuring the Effectiveness of Playing Strategies at Soccer Measuring the effectiveness of playing strategies at soccer. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(4), pp.541–550.

Pratas, J., Volossovitch, A. & Ferreira, A.P., 2012. The effect of situational variables on teams' performance in offensive sequences ending in a shot on goal. A case study. *The Open Sports Sciences Journal*, 5, pp.193–199.

Prechelt, L., 2012. Early stopping - But when? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7700 LECTU, pp.53–67.

ProZone, 2016. ProZone. Available at: www.prozonesports.com [Accessed November 11, 2016].

Rampinini, E. et al., 2009. Technical performance during soccer matches of the Italian Serie A league: Effect of fatigue and competitive level. *Journal of Science and Medicine in Sport*, 12, pp.227–233.

Read, B. & Edwards, P., 1997. *Teaching children to play: A resource for primary teachers 5-11*, Leeds: White Line Publishing.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), pp.53–65.

Saavedra, J.M., Escalante, Y. & Rodríguez, F. a, 2010. A multivariate analysis of performance in young swimmers. *Pediatric exercise science*, 22, pp.135–151.

Sainz De Baranda, P., Ortega, E. & Palao, J.M., 2008. Analysis of goalkeepers' defence in the World Cup in Korea and Japan in 2002. *European Journal of Sport Science*, 8(3), pp.127–134.

Di Salvo, V. et al., 2007. Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine*, 28(3), pp.222–227.

Sampaio, J.E. et al., 2006. Discriminative game-related statistics between basketball starters and nonstarters when related to team quality and game outcome. *Perceptual and Motor Skills*, 103, pp.486–494.

Sampaio, J.E. et al., 2014. Effects of pacing, status and unbalance in time motion variables, heart rate and tactical behaviour when playing 5-a-side football small-sided games. *Journal of Science and Medicine in Sport*, 17, pp.229–233.

Sampaio, J.E. & Maçãs, V.M., 2012. Measuring tactical behaviour in football. *International Journal of Sports Medicine*, 33, pp.395–401.

Sarle, W.S., 2000. How to measure the importance of inputs? *SAS Institute Inc.* Available at: ftp://ftp.sas.com/pub/neural/importance.html [Accessed

November 30, 2016].

Sarle, W.S., 1995. Stopped Training and Other Remedies for Overfitting. In *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*. Pittsburgh, PA, pp. 352–360. Available at: ftp://ftp.sas.com/pub/neural/inter95.ps.Z.

Sarle, W.S., 2007. What is overfitting and how can I avoid it? *SAS Institute Inc.: Usenet FAQs: comp.ai.neural-nets FAQ*. Available at: http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-3.html [Accessed November 30, 2016].

Satizábal, H.F. & Pérez-Uribe, A., 2007. Relevance metrics to reduce input dimensions in artificial neural networks. In *ICANN'07 Proceedings of the 17th international conference on Artificial neural networks*. Porto, Portugal: Springer-Verlag, pp. 39–48.

Saxén, H. & Pettersson, F., 2006. Method for the selection of inputs and structure of feedforward neural networks. *Computers and Chemical Engineering*, 30(6–7), pp.1038–1045.

Semmlow, J.L., 2004. Multivariate analysis: Principal Component Analysis and Independent Component Analysis. In *Biosignal and Medical Image Processing*. New York: Marcel Dekker, Inc., pp. 243–270.

Serrano, J., Shahidian, S. & Fernandes, O., 2014. Validation of a Manual Position Tracking Software (TACTO) to Quantify the Football Movements. *International Journal of Sports Science*, 4(6A), pp.28–34.

Sewell, D., Watkins, P. & Griffin, M., 2013. *Sport and Exercise Science: An Introduction* 2nd ed., London: Routledge.

Shrout, P.E. & Fleiss, J.L., 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), pp.420–428.

Silva, A.J. et al., 2007. The use of neural network technology to model swimming performance. *Journal of sports science & medicine*, 6(1), pp.117–25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3778687&tool=pmcentrez&rendertype=abstract.

Smola, A.J. & Schölkopf, B., 2004. A tutorial on support vector regression. *Statistics and Computing*, 14(3), pp.199–222.

Sofwan, N. et al., 2012. Distinguishing playing pattern between winning and losing field hockey team in Delhi FIH Road to London 2012 tournament. *World Academy of Science, Engineering and Technology, International Science Index 70*, 6(10), pp.182–185. Available at: http://waset.org/Publications/distinguishing-playing-pattern-between-winning-and-losing-field-hockey-team-in-delhi-fih-road-to-london-2012-tournament/14751.

Spearritt, D., 2013. *The impact of field vision on performance within an English Premier league academy soccer team: A case study*. University of Chester. Available at: http://hdl.handle.net/10034/320791.

SportUV, 2016. SportUV. Available at: http://www.stats.com/sportvu-basketball-media/ [Accessed November 11, 2016].

Stergiou, N., 2004. *Innovative analyses in human movement*, Champaign, IL: Human Kinetics.

Stevenson, O.G. & Brewer, B.J., 2017. Bayesian survival analysis of batsmen in Test cricket. *Journal of Quantitative Analysis in Sports*, 13(1).

Steyerberg, E.W., Eijkemans, M.J.C. & Habbema, J.D.F., 2001. Application of Shrinkage Techniques in Logistic Regression Analysis: A Case Study. *Statistica Neerlandica*, 55(1), pp.76–88. Available at: http://doi.wiley.com/10.1111/1467-9574.00157.

Stöckl, M. & Morgan, S., 2013. Visualization and analysis of spatial characteristics of attacks in field hockey. *International Journal of Performance Analysis in Sport*, 13(1), pp.160–178.

Sullivan, C. et al., 2014. Match score affects activity profile and skill performance in professional Australian Football players. *Journal of Science and Medicine in Sport*, 17, pp.326–331. Available at: http://dx.doi.org/10.1016/j.jsams.2013.05.001.

Sunderland, C. et al., 2005. Notational analysis of goals scored from open play in international field hockey. *Journal of Sports Sciences*, 23, pp.1300–1301.

Sunderland, C. et al., 2006. Patterns of play and goals scored in international standard women's field-hockey. *International Journal of Performance Analysis in Sport*, 6, pp.13–29.

Suzuki, K. & Nishijima, T., 2005. Measurement of a soccer defending skill using game performances. In T. Reilly, J. Cabri, & D. Araújo, eds. *Science and football V: The proceedings of the Fifth World Congress on Science and Football*. Oxon: Routledge, pp. 253–261.

Suzuki, K. & Nishijima, T., 2004. Measurement of soccer defending skill based on game performance. *Journal of Sports Sciences*, 22(6), p.518.

Suzuki, K. & Nishijima, T., 2007. Sensitivity of the Soccer Defending Skill Scale: A comparison between teams. *European Journal of Sport Science*, 7(1), pp.35–45.

Tabachnick, B.G. & Fidell, L.S., 2001. *Using multivariate statistics* 4th ed., New York: Harper & Row.

Taki, T. & Hasegawa, J., 2000. Visualization of dominant region in team games and its application to teamwork analysis. *Proceedings Computer Graphics International 2000*, pp.227–235. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=852338.

Taylor, B.N. & Kuyatt, C.E., 2001. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results: Appendix D1. Terminology. *National Institute for Standards and Technology, Gaithersburg, MD*, p.D.1.1.2. Available at: http://physics.nist.gov/TN1297 [Accessed September 5, 2016].

Tenga, A.P. et al., 2010. Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. *Journal of Sports Sciences*, 28(3), pp.245–255.

Tenga, A.P., 2010. *Reliability and validity of match performance analysis in soccer: A multidimensional qualitative evaluation of opponent interaction*. Norwegian School of Sport Sciences. Available at: http://hdl.handle.net/11250/171287.

Teramoto, M. & Cross, C.L., 2010. Relative Importance of Performance Factors in Winning NBA Games in Regular Season versus Playoffs. *Journal of Quantitative Analysis in Sports*, 6(3).

Thomas, A.C., 2007. Inter-arrival Times of Goals in Ice Hockey. *Journal of Quantitative Analysis in Sports*, 3(3).

Travassos, B. et al., 2012. Informational constraints shape emergent functional behaviours during performance of interceptive actions in team sports. *Psychology of Sport and Exercise*, 13(2), pp.216–223. Available at: http://dx.doi.org/10.1016/j.psychsport.2011.11.009.

Travassos, B. et al., 2011. Interpersonal coordination and ball dynamics in futsal (indoor football). *Human Movement Science*, 30, pp.1245–1259.

Travassos, B. et al., 2013. Performance analysis in team sports: Advances from an ecological dynamics approach. *International Journal of Performance Analysis in Sport*, 13(1), pp.83–95.

Travassos, B. et al., 2012. Spatiotemporal coordination behaviors in futsal (indoor football) are guided by informational game constraints. *Human Movement Science*, 31(4), pp.932–945. Available at: http://dx.doi.org/10.1016/j.humov.2011.10.004.

Tromp, M. & Holmes, L., 2011. The effect of free-hit rule changes on match variables and patterns of play in international standard women ’ s field hockey. *International Journal of Performance Analysis in Sport*, 11(1995), pp.376–391.

Turner, A.P., Allison, P.C. & Pissanos, B.W., 2001. Constructing a Concept of Skilfulness in Invasion Games Within a Games for Understanding Context. *European Journal of Physical Education*, 6(1), pp.38–54. Available at: http://www.tandfonline.com/doi/abs/10.1080/1740898010060105.

Vaz, S. et al., 2013. The Case for Using the Repeatability Coefficient When Calculating Test-Retest Reliability. *PLoS ONE*, 8(9), pp.1–7.

Vicon, 2016. Vicon. Available at: https://www.vicon.com/ [Accessed November 11, 2016].

VideoLAN, 2015. VLC media player. Available at: http://www.videolan.org/vlc/index.html.

Vila, H. et al., 2011. Tactical and shooting variables that determine win or loss in top-Level in water polo. *International Journal of Performance Analysis in Sport*, 11(3), pp.486–498. Available at: http://www.ingentaconnect.com/content/uwic/ujpa/2011/00000011/00000003/art00009.

Vilar, L., Araújo, D., Davids, K. & Travassos, B., 2012. Constraints on competitive performance of attacker–defender dyads in team sports. *Journal of Sports Sciences*, 30(5), pp.459–469.

Vilar, L. et al., 2014. Coordination tendencies are shaped by attacker and defender interactions with the goal and the ball in futsal. *Human Movement Science*, 33, pp.14–24.

Vilar, L. et al., 2013. Spatial-temporal constraints on decision-making during shooting performance in the team sport of futsal. *Journal of Sports Sciences*, 31(8), pp.840–846. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23244400.

Vilar, L., Araújo, D., Davids, K. & Button, C., 2012. The role of ecological dynamics in analysing perfermance in team sports. *Sports Medicine*, 42(1), pp.1–10. Available at: http://www.isrj.net/PublishArticles/1052.pdf.

Vinson, D. et al., 2013. Penalty corner routines in elite women's indoor field hockey: prediction of outcomes based on tactical decisions. *Journal of sports sciences*, 31(April), pp.887–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23305340.

Vogelbein, M., Nopp, S. & Hökelmann, A., 2014. Defensive transition in soccer - are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. *Journal of Sports Sciences*, 32(11), pp.1076–1083. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24506111.

Walton, J.S., 1981. *Close-range cine-phtogrammetry: A generalise technique for quantifying gross human motion.* Pennsylvania State University.

Wang, H. et al., 2012. Interaction-based feature selection and classification for high-dimensional biological data. *Bioinformatics*, 28(21), pp.2834–2842.

Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass corrleation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19(1), pp.231–240.

Welsh, A.H. & Knight, E.J., 2015. "Magnitude-based Inference": A statistical review. *Medicine and Science in Sports and Exercise*, 47(4), pp.874–884.

Wheeler, K.W., Askew, C.D. & Sayers, M.G., 2010. Effective attacking strategies in rugby union. *European Journal of Sport Science*, 10(4), pp.237–242.

Wheeler, K.W., Mills, D. & Lyons, K., 2013. Effective defensive strategies at the ruck contest in rugby union. *International Journal of Sports Science & Coaching*, 8(3), pp.481–492.

Whittaker, J.C., Thompson, R. & Denham, M.C., 2000. Marker-assisted selection using ridge regression. *Genetics Research*, 75(2), pp.249–252.

Wu, F. et al., 2007. An Investigation of an Integrated Low-cost GPS, INS and

Magnetometer System for Sport Applications. In *Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007)*. Fort Worth, TX: The Institute of Navigation, pp. 113–120.

Yue, Z. et al., 2008. Mathematical analysis of a soccer game. Part I: Individual and collective behaviors. *Studies in Applied Mathematics*, 121, pp.223–243.

Zou, H., 2005. Regularization and Variable Selection via the Elastic Net. , pp.301–320. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.9735&rep=rep1&type=pdf.

Zubillaga, A. et al., 2013. Influence of ball position on playing space in Spanish elite women's football match-play. *International Journal of Sports Science & Coaching*, 8(4), pp.713–722. Available at: http://linksource.ebsco.com/ls.941a0f3a-fc44-412f-81a4-365a06658f2f.false/linking.aspx?&sid=OVID:psycdb&id=pmid:&id=doi:&issn=1747-9541&isbn=&volume=8&issue=4&spage=713&pages=713-722&date=2013&title=International+Journal+of+Sports+Science+&+Coaching&atitl.

Appendices

Appendix A - Parameter estimate for the Backward Dropping Algorithm

A.1 The *I*-score

The *I*-score measures the influence of an explanatory variable or set of explanatory variables on an outcome variable. It was proposed by Chernoff et al. (2009) as a generalisation of Lo and Zheng's (2002) haplotype transmission disequilibrium statistic. Before explaining the *I*-score, an understanding of partition elements is required.

A partition element is a unique expression of a set of discrete variables' values. For example, a set of three binary variables has $2^3 = 8$ partition elements because there are eight possible combinations of the variables' values (Table A.1).

Table A.1 Every partition element for a set of three binary variables.

	Variable values		
Partition element	**Variable 1**	**Variable 2**	**Variable 3**
1	0	0	0
2	0	0	1
3	0	1	0
4	1	0	0
5	0	1	1
6	1	1	0
7	1	0	1
8	1	1	1

Each partition element represents a specific interaction between the variables. The [0, 0, 0] partition element represents the all-zero interaction and [0, 1, 0] the partition element represents one of the two-zero interactions. Each of these partition elements is exclusively associated with a portion of observations from the dataset, i.e. some observations were [0, 0, 0] and some were [0, 1, 0]. Not all possible partition elements might be expressed in a dataset. For example, there might not be any case were all three inputs recorded a zero value for the same observation.

The *I*-score evaluates the interaction between an outcome variable and the partition elements of explanatory variables. Consider Y to be an outcome variable, with n observations, that is associated with a dataset of three binary explanatory variables, also with n observations. Let j be one of the partition elements and n_j be the number of observations associated with the j^{th} partition element. The general form of the *I*-score is

$$I = \sum_{j \in m} n_j^2 \left(\bar{Y}_j - \bar{Y}\right)^2 \qquad \text{[A.1]}$$

where $\bar{Y}_j$ is the arithmetic mean of Y for the observations within the j^{th} partition element of m possible partition elements, and $\bar{Y}$ is the arithmetic mean for the entire set of n observations. In this general form, Y is not necessarily discrete but the work of Chernoff et al. (2009), Wang et al. (2012) and Lo and Zheng's (2002) all consider Y to be discrete. This is because the nature of the outcome variable in genetic analytics tends to be a binary classification indicating expression or non-expression of a genetic trait. The discrete form of the *I*-score with a binary outcome variable, $Y = 1$ or 0, is

$$I = \sum_{j \in m} \left(n_{(Y=1 \mid j)} - \left(n_j \cdot \frac{n_{(Y=1)}}{n} \right) \right)^2 \qquad \text{[A.2]}$$

where $n_{(Y=1 \mid j)}$ is the number of outcomes in the j^{th} partition element that equal 1, and $n_{(Y=1)}$ is the overall number of outcomes that equal 1. The second term in equation A.2 represents the expected number of outcomes in the j^{th} partition element that equal 1 under the null hypothesis that the set of explanatory variables has no association with Y. As such, the ***I***-score is the squared-sum of deviations between actual and expected $Y = 1$ observations over all m partition elements.

Using the ***I***-score with the Backward Dropping Algorithm requires three parameters to be defined:

1. The size of the metric sample set, k.
2. The size of the nearly completely covered return set, r.
3. The number of iterations, $\widehat{2B}$.

A.2 Further notes on the number of partition elements, $\boldsymbol{m}$

It was stated in the previous section that each partition element is exclusively associated with a portion of observations from the dataset. This means that the number of partition elements, m, affects the expected mean number of observations per partition element, i. Under the assumption that observations are uniformly distributed across all partition elements, each partition element will have $\frac{n}{m}$ observations, where n is the number of observations. Such a uniform distribution is not likely so the greater the number of partition elements, the more likely it is that some will have few or no associated observations. However, a uniform distribution is the least biased given that there is no prior information as to the distribution of i. The consequence of the relationship between m and i is that the number of partition elements affects the power of the analysis by affecting the volume of supporting data for each partition element. Ideally, every partition element would have many associated observations just as any statistical test benefits from greater availability of supporting data. Wang et al. (2012, p 2838) suggested $i \geq 4$ based on the similarity between the ***I***-score and Pearson's χ^2 test. This assertion provides a lower bound for i.

Wang et al. (2012) provide a relationship between m, i and n based on the assumption that i is distributed uniformly across all partition elements:

$$\frac{n^i}{i! \cdot (m_{k-1})^{i-1}} \geq 1 \quad \text{, for } i \geq 2 \qquad \text{[A.3]}$$

Equation A.3 is a generalisation of equation 2 in Wang et al. (2012, see Suppl. 3.2.1). The m_{k-1} term refers to the number of partition elements resulting from a metric sample set of size k-1 (this is an important point that will be revisited in the next section). Making m_{k-1} the subject of the equation gives

$$m_{k-1} \leq \sqrt[i-1]{\frac{n^i}{i!}} \qquad \text{, for } i \geq 2 \qquad \text{[A.4]}$$

Equation A.4 provides an upper bound for m_{k-1} when $i = 4$, as per Wang et al.'s (2012) suggestion. The exact number of partition elements will be determined by using the largest integer value of k satisfying equation A.4, given the range of possible values of the variables selected.

A.3 Size of metric sample set, *k*

Each iteration of the Backward Dropping Algorithm will start by selecting a metric sample set of size k from the entire set of metrics available. The size of k affects the number of partition elements, m, based on the possible range of values of the metrics selected. A larger value of k increases the likelihood of including a truly influential metric in a selected metric sample set but will reduce

i via k's relationship with m. Recall that i refers to the number of observations associated with a partition element so larger values of i are preferred.

The founding work with partition elements considered only binary explanatory variables (Chernoff et al., 2009). The number of partition elements was $m = 2^k$, where k is the number of variables being considered. In general, when all variables have the same number of discrete values, c, then $m = c^k$. If the dataset contains metrics with a variety of discrete values, then the number of partition elements will vary according the metric sample set under investigation.

Let $\boldsymbol{C}$ be an ordered set of all variables' ranges of values such that $\boldsymbol{C} = \{c_1, c_2, \dots, c_S\}$, where c_1 is the largest range, c_2 is the second largest, etc., and S is the total number of variables. The minimum size of k given m, $k_{min|m}$, would be the number of successive products of $\boldsymbol{C}$ until m was reached. This $k_{\min|m}$ is the conservative value for k because it represents the extreme scenario where a metric sample set favours large ranges of metric values. The maximum size of k given m, $k_{max|m}$, represents the extreme scenario where a metric sample set favours small ranges of metric values. Specifically, for the current work, this is the scenario where all selected metrics could be one of two possible discrete values. This would be calculated similarly to $k_{min|m}$ by getting the number of successive products of $\boldsymbol{C}$ until m was reached, but starting from the end of $\boldsymbol{C}$ with c_S rather than c_1. It might be wise to set the size of the metric sample set at $k = k_{\min|m}$ to favour larger values for i rather than favouring an increased likelihood of including a truly influential metric in a selected metric sample set.

A.4 Size of nearly completely covered return set, r

To use the Backward Dropping Algorithm, the user must decide what order of interaction they want to cover, ranging from the two-way interactions between pairs up to the single S-way interaction between all metrics. It is not practical to examine all possible orders of interaction but it might be possible to generalise a metric’s interaction effect by assessing its many low-order interactions. Wang et al. (2012) suggest that a value r is chosen to represent the maximum level of interaction to be assessed thoroughly. That is to say, that r represents the size of subset whose interactions are covered completely in the sampling. Subsets of size $r - x$ and r will all be sampled but subsets of size $r + x$ will not, for $0 < x < r, x \in \mathbb{N}$.

The value of r could be based on domain knowledge of the expected order of interaction amongst the metrics but no such insight exists for the current work. Wang et al. (2012) suggest that r should be at least the lower bound of k that results in partition elements with an expected arithmetic mean number of observations $i \geq 4$. An adaptation of Wang et al.'s (2012, §3.2.2) discussion suggests a value of

$$r = \left\lfloor \log_4 \left(\frac{n}{4} \right) \right\rfloor \quad \text{[A.5]}$$

where n = the number of observations.

The value of r describes the size of the return set that is only *nearly completely* covered. This is because any metric sample set selected during the Backward Dropping Algorithm can only cover the r-size sets within it. Thus, each metric sample set selected will cover a cluster of r-size sets within the space of all the possible r-size sets in the entire dataset. For example, the metric sample set $\{a, b, c, d\}$ can only cover metric combinations that contain metrics a, b, c and d but cannot cover any metric combinations that involve other metrics. It is therefore important to perform the Backward Dropping Algorithm multiple times to improve coverage of r-size sets. The more metric sample sets taken at random, the more likely it is that all metrics will have been sampled in many combinations.

A.5 Number of iterations, $2\hat{B}$

Each iteration of the Backward Dropping Algorithm randomly selects a metric sample set from the entire collection of metrics. A $b = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ number of iterations is required to ensure that all possible combinations of metrics are considered but many redundant computations would be made. Wang et al. (2012) propose a $\hat{B}$-number of iterations, such that

$$\hat{B} \approx \frac{\binom{S}{r}}{\binom{k}{r}} \log_e \binom{S}{r} \qquad , \text{for } r \leq k \qquad \text{[A.6]}$$

where S = the total number of metrics available, k = the size of the metric sample set selected, and r = the size the nearly completely covered return set. Equation A.6 is derived by considering the selection of metric sample sets to follow a Poisson process (Wang et al., 2012, Suppl. 3.2.2). Practically, this means that each selection of a metric sample set is independent and random. Theoretically, fewer iterations would be required if only unique metric sample sets were submitted to the Backward Dropping Algorithm but this would violate the independence assumption assumed for calculating $\hat{B}$.

A $\hat{B}$-number of iterations will not completely cover all r-size sets because of the clustering characteristic highlighted at the end of the previous section. Such clustering increases the likelihood of not covering all r-size sets. Wang et al. (2012) proposed $2\hat{B}$ as a practical upper bound to cover sufficiently enough r-size sets to find influential variables.

Appendix B - Institutional, ethical approval

Faculty of Health and Wellbeing Research Ethics Committee **Ref No: HWB-S&E-11**
Sport and Exercise Research Ethics Review Group
Report Form

Principal Investigator: Ciaran McInerney

Title: Investigating open play in field hockey using spatial and tactical metrics.

Recommendation:

Approved	x
Approved (see comments)	
Resubmit (see comments)	

Comments:

Signature : Date: 05.05.15
Donna Woodhouse
Chair, Sport and Exercise Research Ethics Review Group

Note: *Approval applies until the anticipated date of completion unless there are changes to the procedures, in which case another application should be made.*

Name of Supervisor : Simon Goodwill

Appendix C - Email correspondence with Prof. Inchi Hu

From: Inchi Hu
To: McInerney, Ciaran
Subject: Re: Query - 2nd difference in Wang et al. (2012)
Date: 30 June 2016 08:43:24

Dear Dr. Mclnerney,

The reason why we chose 7 as the cutoff value and not 14 is that we pick the first big gap after the main body of the histogram, which is 7 and not 14.

I appreciate your persistency because there are so many things that require my attention at the moment even though it is my intention to answer questions about research related my paper.

Best, Inchi Hu

Sent from my iPhone

On 28 Jun 2016, at 19:32, McInerney, Ciaran <C.McInerney@shu.ac.uk> wrote:

> Hi Prof. Hu,
>
> I hope things are well with you. I am writing to follow up on an email that sent on the 3rd June about one of your publications. Please respond when you find the time.
>
> Ciarán
>
> ---
>
> **From:** McInerney, Ciaran
>
> **Sent:** 16 June 2016 20:31
>
> **To:** imichu@ust.hk
>
> **Subject:** FW: Query - 2nd difference in Wang et al. (2012)
>
> Hi Prof. Hu,
>
> I am following up on an email I sent on the 3rd June. Please reply when you find the time.
>
> Regards, Ciarán
>
> ---
>
> **From:** McInerney, Ciaran
>
> **Sent:** 10 June 2016 10:04

To: 'imichu@ust.hk'

Subject: FW: Query - 2nd difference in Wang et al. (2012)

Hi Prof. Hu,

I hope you are well. I appreciate that you are busy. If you find the time, I would value your reply to my previous email. Thank you again for your help to date.

Regards, Ciarán

From: McInerney, Ciaran

Sent: 03 June 2016 15:06

To: 'Inchi Hu'

Subject: RE: Query - 2nd difference in Wang et al. (2012)

Hi Prof. Hu,

Thank you once again for taking the time to respond.

I understand your explanation. Thankfully, my data is responding in a similar way to what you describe. Is the justification for using the 1st difference of retention frequencies also based on your empirical evidence? If not, please explain.
I have been trying to apply the 2nd difference method to determine a cut-off for return sets obtained during an application of Chernoff et al.'s (2009) resuscitation algorithm. I was hoping to use it as an alternative for the nr method that I spoke about in earlier email messages. Unfortunately, the 2nd differences of the I-scores for return sets from the Backward Dropping Algorithm do not behave as you describe. No problem, it was just an idea I was testing.

If you have the time, I have question about the false positive removal algorithm. The false positive removal procedure is explained in section 3.2.3 of the Wang et al.'s (2012) supplementary information (attached for your ease). Figure S1 shows an example histogram of the number of forward-one sets. Could you please explain the rationale for choosing the threshold of 7 forward-one sets – why not

14? I ask because there is no clear gap or outliers in the histogram for my dataset. Again, I am looking for theoretical justification but I understand if the choices made in the publication were empirically based.

Regards,
Ciarán

From: Inchi Hu [mailto:imichu@ust.hk]

Sent: 03 June 2016 14:38

To: McInerney, Ciaran

Subject: Re: Query - 2nd difference in Wang et al. (2012)

Hi Dr. McInerney,

I was occupied by a few urgent tasks in the last few days.

Here are the empirical and heuristic evidence of the 2nd difference method. In many data sets that I have analyzed, all of them have 2nd differences drop rapidly in the beginning and then settle down around zero.

Discard those variables with 2nd differences near or around zero usually reduces prediction error and overfitting.

If one lowers the cutoff value Just a little where 2nd difference is near zero, a lot more variables (than a cutoff with a higher 2nd difference value) be included. These variables included are believed to be mostly noisy ones, carrying very little signal.

The rigorous theoretical justification of the 2nd difference method is still an open issue.

Regards,
Inchi

On 2 Jun 2016, at 22:57, McInerney, Ciaran <C.McInerney@shu.ac.uk> wrote:

> Hi Prof. Hu,
>
> I hope this email finds you well. I'm following up on an email I sent earlier in the week (see below). I would greatly appreciate your help, if you could provide it.
>
> Ciarán

> **From:** McInerney, Ciaran
>
> **Sent:** 30 May 2016 10:02
>
> **To:** Inchi Hu
>
> **Subject:** RE: Query - 2nd difference in Wang et al. (2012)
>
> Prof. Hu,
>
> Thank you for your reply. I have a few more questions about the methods in the paper. Please let me know if you would agree to answer them.
>
> In the meantime, I will ask only one: why is a threshold indicated simply because two successive differences in I-score are similar? The second difference can be zero or near zero whenever I-scores differs by approximately the same amount between successive pairs. For example, if the first three I-scores differ by an equal amount (large or small) then

the first 2nd difference will be zero. Why is this a good threshold? My understanding of the rationale provided in the paper is that the

threshold is based on the shape of the 2nd difference curve rather than a property of the differences in successive values. Please explain or direct me to an appropriate source.

Regards,
Ciarán McInerney

From: Inchi Hu [imichu@ust.hk]

Sent: 27 May 2016 10:34

To: McInerney, Ciaran

Subject: Re: Query - 2nd difference in Wang et al. (2012)

Dear Dr. McInerney,

Thanks for your interest in the 2nd difference method. If the tenth difference is near zero, I would retain the top 10,000 I-scores. My reasoning is based on the mean value theorem, roughly placing each difference in the middle of the corresponding range of scores.

When formulating the 2nd difference method, I did not consider the nr method in Chernoff et al. Maybe they are related but I did not pursue that direction.

Regards,
Inchi Hu

On 25 May, 2016, at 21:53, McInerney, Ciaran <C.McInerney@shu.ac.uk> wrote:

> Hi Prof. Hu,
>
> I am a PhD researcher studying in the Centre for Sports Engineering Research. I am using the method explained in your 2012 publication entitled “Interaction-based feature selection and classification for high-dimensional biological data”. I have a question that I hope you can answer.
>
> In section 3.1 you explain the 2nd difference method for deciding a cut-off during variable screening. You say that the cut-off correspond to the point where the 2nd difference is near zero for the first time. My question is: **How many I-scores does the *i*th 2nd difference refer to**, where *i* is the index of the 2nd difference value? In other

words, if the tenth 2nd difference is near zero, do I carry forward the top 11,000 I-scores or the top 10,000 I-scores?

I thank you kindly for your time, in advance. I hope youcan help.

p.s.
Is this 2nd difference method an improvement on the <image003.png> method used in the Chernoff et al. 2009 paper "Discovering influential variables: A method of partitions", which your co-authors worked on?

Ciarán McInerney

Appendix D - Chernoff et al.'s (2009) resuscitation algorithm

The resuscitation algorithm is as follows:

1. Decide the length of the *final list*.
2. Decide the length of the *resuscitated list*. The length of the *final list* minus the length of the *resuscitated list* is the *reduced list*.
3. For a *reduced list* of length L_d, fill it with the top L_d initially, highly ranked metrics. For example, if the length of the *reduced list* is 20 then assign the top 20 metrics to the *reduced list*.
4. For a *remaining list* of length L_m, fill it with the top L_m resuscitation candidates. NB: metrics in the *remaining list* must not duplicate those in the *reduced list*.
5. Run the Backward Dropping Algorithm analysis with a proportion, p, of each metric sample set taken from the *reduced list* and the remainder randomly sampled from the *remaining list*.
6. Rank the metrics used in the outputted return sets. For the current work, the retention frequency method was used to rank the metrics (cf. Chernoff et al. (2009) for other options).
7. Discard any metrics that are already in the *reduced list*.
8. For a *resuscitated list* of length L_s, append $\frac{L_s}{2}$ of these unique top-ranked metrics to the *reduced list*. This augmented *reduced list* is now the new *reduced list*.
9. Define a new *remaining list* with L_m resuscitation candidates. NB: metrics in the new *remaining list* must not duplicate those in the new *reduced list*.
10. Run the Backward Dropping Algorithm analysis again with a proportion, p, of each metric sample set taken from the new *reduced list* and the remainder randomly sampled from the new *remaining list*.
11. Rank the metrics used in the new return sets using the retention frequency method.
12. Discard any metrics that are already in the new *reduced list*.
13. For a *resuscitated list* of length L_s, append $\frac{L_s}{2}$ of the unique top-ranked metrics to the *reduced list*. This augmented *reduced list* is now the *final list*.

In steps 7 and 12, resuscitated metrics that are already in the *reduced list* are discarded so that only unique metrics are appended. It might be the case that fewer than $\frac{L_S}{2}$ metrics are appended to the *reduced list* because not enough unique metrics were resuscitated. As alluded to in chapter 7, this is how the *final list* might end up being shorter than expected.