Rein and Memmert *SpringerPlus (2016)5:1410*
DOI 10.1186/s40064-016-3108-2

SpringerPlus

REVIEW Open Access

# Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science

Robert Rein* and Daniel Memmert

**Abstract**

Until recently tactical analysis in elite soccer were based on observational data using variables which discard most contextual information. Analyses of team tactics require however detailed data from various sources including technical skill, individual physiological performance, and team formations among others to represent the complex processes underlying team tactical behavior. Accordingly, little is known about how these different factors influence team tactical behavior in elite soccer. In parts, this has also been due to the lack of available data. Increasingly however, detailed game logs obtained through next-generation tracking technologies in addition to physiological training data collected through novel miniature sensor technologies have become available for research. This leads however to the opposite problem where the shear amount of data becomes an obstacle in itself as methodological guidelines as well as theoretical modelling of tactical decision making in team sports is lacking. The present paper discusses how big data and modern machine learning technologies may help to address these issues and aid in developing a theoretical model for tactical decision making in team sports. As experience from medical applications show, significant organizational obstacles regarding data governance and access to technologies must be overcome first. The present work discusses these issues with respect to tactical analyses in elite soccer and propose a technological stack which aims to introduce big data technologies into elite soccer research. The proposed approach could also serve as a guideline for other sports science domains as increasing data size is becoming a wide-spread phenomenon.

**Keywords:** Big data, Sports performance, Sports analytics, Machine learning, Simulation, Spatiotemporal data, Neural networks, Deep learning, Quantified self

Tactics are a central component for success in modern elite soccer. Yet until recently, there have been few detailed scientific investigations of team tactics. One reason in this regard has been the lack of available, relevant data. With the development of advanced tracking technologies this situation has changed recently. Instead, now the amount of available data is becoming increasingly difficult to manage. In the present article we discuss how recent developments of big data technologies from industrial data analytics domains address these problems. Further, the present work provide an overview how big data technologies may provide new opportunities to study tactical behavior in elite soccer and what future challengers lie ahead.

## Soccer tactics background

According to the Oxford dictionary, tactics describe "an action or strategy carefully planned to achieve a specific end". Regarding competitive soccer, naturally the aim the end of the activity is to win the game. Choosing an appropriate tactic is therefore crucial for every pre-game preparation (Carling et al. 2005b; Kannekens et al. 2011; Sampaio and Macas 2012; Yiannakos and Armatas 2006). Regarding the definition of tactics Gréhaigne and Godbout (1995) introduced a distinction between the strategy and tactics. Here, the team strategy describes the decisions made before the game with respect to how the team

*Correspondence: r.rein@dshs-koeln.de
Institute of Cognition and Team/Racket Sport Research, German Sport University Cologne, Am Sportpark Müngersdorf 6, 50933 Cologne, Germany

wants to play whereas the tactic is the result of the ongoing interactions between the two opposing teams. This approach seems somewhat counter to the basic definition of the term tactics provide above. Furthermore, it is not clear how these two concepts can be clearly delineated from each other as the real-time interactions between the players will be conditioned by the a priori strategy. Following a classical practitioner's approaches the tactic specifies how a team manages space, time, and individual actions to win a game (Fradua et al. 2013; Garganta 2009). In this context, space specifies for example were on the pitch a certain actions takes places or which area a team wants to occupy during the attack and the defense. Time in contrast describes variables like frequency of events and durations (ball possession) or how quick actions are being initiated. For example, a team could decide to have a slow buildup during attack initiate in the defense third where individual players hold the ball for longer times whereas in the attacking third only fast on-touch pass sequences are preferred. Finally, individual actions specify the type of actions which are being performed, for example turnovers, crosses and passes (Garganta 2009). This classification can be further hierarchically organized along the number of participating players into individual tactics, group tactics, team tactics, and match tactics which is also a scheme commonly referred to by soccer practitioners (Bisanz and Gerisch 1980, p.201; Carling et al. 2005a). Individual tactics describe all one-on-one events during offensive and defensive play with and without the ball. For example, the way the ball carrier is approached by a defender can be considered as part of the individual tactic. For example, the defender could immediately attack the ball carrier and put him under pressured or the defender could use a more passive approaches focusing mainly on blocking passing channels. Group tactics describe the cooperation between sub groups within a team for example the defensive block during an offside trap. Team tactics describe preferred offensive and defense team formations (e.g. 4-4-2) and the positioning of the formation on the pitch (Grunz et al. 2012). Finally, game tactics describe the team's playing philosophy such as counter-attack or ball possession play. A recent study investigated for example ball possession regain in the German Bundesliga where the results showed that more successful teams were faster to regain ball possession after losing possession (Shafizadehkenari et al. 2014; Vogelbein et al. 2014). In summary, soccer tactics describe the microscopic and macroscopic organizational principles of the players on the pitch spanning from individual to group decision making processes.

To ensure successful execution at all tactical levels, a coach has to take into account the status of the team, the status of the opposition, as well as external factors like playing at home or even the weather (Gréhaigne and Godbout 1995; Lago 2009; Mackenzie and Cushion 2013; Sarmento et al. 2014) (compare Fig. 1). Therefore, in the following tactics refers to both the a priori decisions as well as the real-time adaptations during a game. As the two competing teams try to out-smart each other, the tactics are not constant but should be adapted according to the interactions between and within the two teams (Balagué and Torrents 2005; Garganta 2009; Grehaigne et al. 1997; Gréhaigne and Godbout 2014). For example, a player substitution by the opposition team may introduce a change in playing tactics which the coach may have to respond to be changing his teams' tactics. Team tactics are therefore governed by a complex process resulting from a network of inter-dependent parameters (Kempe et al. 2014). Although the scheme presented above follow a hierarchical pattern the flow information in reality does go in both directions. Tactics at a higher level condition the tactics at the lower level and vice versa success of individual actions equally conditions success at a higher level (Araújo et al. 2006; Sampaio and Macas 2012). Thus, tactics can be interpreted as complex structure of composed of a new of interwoven dependencies. Accordingly, tactical analysis should reflect this complexity.

Over the years tactical decisions, like preferred playing formations or game tactics, have increased in complexity and coaches' tactical abilities are under constant public scrutiny. Until very recently this stood somewhat in contrast to the amount of scientific investigations studying tactical decisions in elite soccer (Carling et al. 2005c; Garganta 2009; Sampaio and Macas 2012; Sarmento et al. 2014). The reason for this somewhat surprising fact may have been the lack of accessible and/or reliable data required for tactical analysis (Rampinini et al. 2007). The present gold standard to assess tactical behavior and team performance in general in elite soccer is commonly based on individual game observations (Dutt-Mazumder et al. 2011; Mackenzie and Cushion 2013). A domain expert (coach, scout) observes a game and rates the team tactics according to his personal experiences. Although usually a specific coding manual is used a general consensus regarding relevant variables is currently missing (James 2006; Sarmento et al. 2014) and data often lack objectivity and reliability (James et al. 2002). Furthermore, as game interactions are highly dynamic and contextual circumstances change continually it is under debate to what extent reliable measures are attainable in general (Lames and McGarry 2007). In addition, detailed game analyses based on observational approaches are highly time-consuming which limited their application in the past (Carling et al. 2008; James 2006). Consequently, demand for more quantitative oriented (automatic) approaches to analyze tactical behavior in elite soccer is increasing

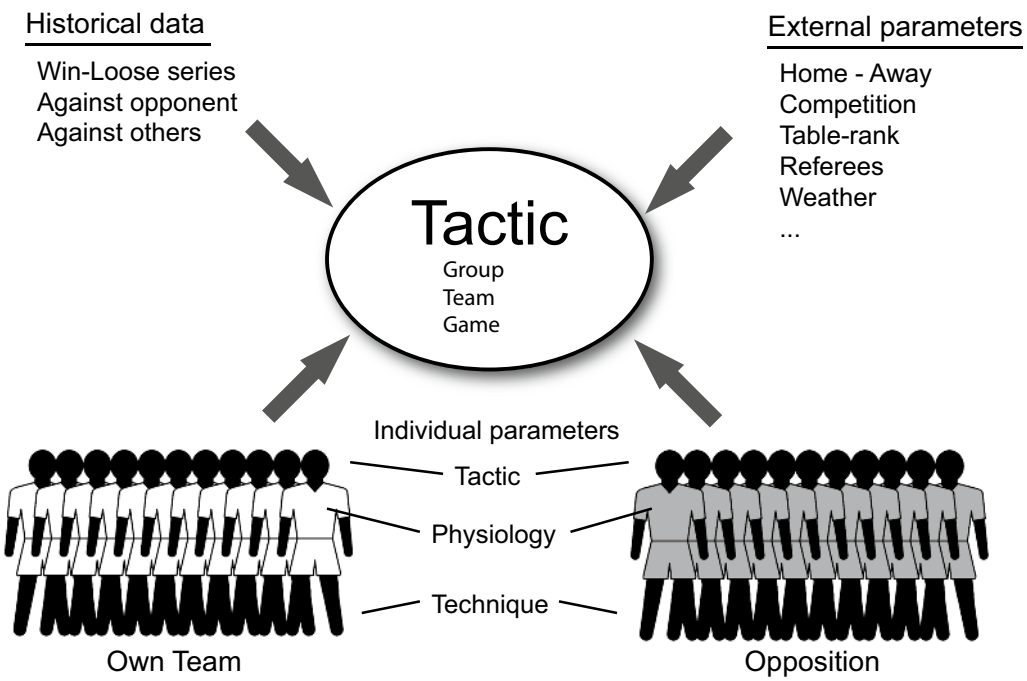


**Fig. 1** Overview of factors influencing tactics in soccer

(Beetz et al. 2005; Carling et al. 2014; Lucey et al. 2013a, b; Wang et al. 2015). Thus, whereas the processes underlying tactics in elite soccer have increased over the years the scientific approaches have not quite evolved with the same speed.

In this regard, fine-grained global reporting of game event statistics for commercial audiences has seen a tremendous rise in recent years and detailed game data are routinely reported (Baca 2008; Baca et al. 2004; Sarmento et al. 2014). The reason for this increased availability of game data is largely due to progress made in player tracking technologies (Baca 2008; Carling et al. 2008; Castellano et al. 2014; D'Orazio and Leo 2010; Lu et al. 2013). Recently FIFA the governing body for international competitive soccer decided to allow the usage of wireless sensors technologies to track player positions and physiological parameters during competitions (di Salvo and Modonutti 2009). This will further increase the availability of detailed performance data from elite soccer. Thereby this has been a results of today's common practices among professional teams to already collect physiological and tracking data during training and friendly matches to manage the training process (Bush et al. 2015; Carling et al. 2008; Ehrmann et al. 2016; Goncalves et al. 2014; Ingebrigtsen et al. 2015). At present, several different tracking systems are available in the market including vision based systems, Global Positioning Systems (GPS), and radio wave based tracking systems (Leser et al. 2011). Although data quality and reliability used to be a problem, in recent years the systems have matured to such an extent that the data is now of sufficient quality to satisfy scientific standards. Several recent overviews addressing the advantages and disadvantages between the different available systems are available in the literature (Barris and Button 2008; Buchheit et al. 2014; Carling et al. 2008; Castellano et al. 2014; D'Orazio and Leo 2010; Harley et al. 2011; Valter et al. 2006). Thus modern tracking data allows the analysis of technical, tactical and physical demands in elite soccer.

In general, a trend seems to emerge where analyses of soccer games in public media outlets are also becoming increasingly data aware. One example in this regard is the increasing number of free internet blogs reporting detailed game analyses. Using observational techniques from TV game broadcasts data as well as publicly available internet soccer databases these blogs provide novel approaches to data driven performance analysis in soccer much in the same spirits as the sabermetrics community has for American baseball during the late 90's (Lewis 2004). Recently, investigations have emerged which used sentiment analysis from twitter feeds to identify for example high impact events during games (Buntain 2014; Yu and Wang 2015) and to predict game outcomes (Godin et al. 2014). In this regard, quantified-self initiatives may also provide future opportunities to generate valuable data for scientific investigations (Appelboom

et al. 2014; Shull et al. 2014). In summary, lack of reliable data to perform tactical analysis in elite soccer is becoming less of a problem and novel data sources are continually being discovered and developed.

## Analysis of soccer tactics

Traditionally, one area which has produced a wealth of studies investigating soccer performance is with respect to the physiological demands in competitive soccer (Carling et al. 2008; Mohr et al. 2005). However, until recently few connections between physiological demands and tactical behavior in elite soccer have been made (Bloomfield et al. 2007; Drust et al. 2007; Moura et al. 2012). As was made clear in the introduction, the success for a tactics depends on the abilities of the individual players to actually implement the required actions. Obviously this requires that the players fulfill the necessary physiological requirements, for example, when playing a ball possession type of play (da Mota et al. 2016). Rampinini et al. (2007) investigated the total running distances and the time spent different running speed categories (standing to sprinting). The results showed a significant influence of the level of the opponents and the playing position (compare also Goncalves et al. 2014). Bush et al. (2015) investigated the changes in physiological performance variables in the English Premier League across several seasons and results indicated significant increases in passing event rates associated with changes in team tactics (Bush et al. 2015). Carling (2011) investigated the influence of opposition formations on physiological and skill-related performance variables and found for example increased running distances when playing against a 4-2-3-1 formation compared to a 4-4-2 formation (Carling 2011). Sampaio et al. (2014) investigated the influence of time unbalance and game pace on physiological demands during a 5-a-side small sided game were one player was dropped in either side to create an inferiority or an superiority condition. The results suggested an effect of team unbalance on the time spent in different hear rate zones suggesting that the inferior team had to work harder (Sampaio et al. 2014). In summary, these results indicate that tactical behavior and physiological variables are linked but more in-depth analyses are missing. Accordingly, at present it is unclear how to combine information about player's physiology from training and competition with team tactics (Castellano et al. 2014) and no connections between individual technical performance and team tactics have been made so far (Hughes and Bartlett 2002).

Traditionally, tactics analyses relied on notational analysis approaches based on average statistics and tallies (Hughes and Bartlett 2002). Indicators include for example passing variables (Hughes and Franks 2005; Liu et al. 2015), ball possession (Collet 2013; Lago 2009), ball recovery (Vogelbein et al. 2014), or playing style (Tenga et al. 2010a, b). The main limitation of the traditional notational approach is that almost all contextual information is discarded, these measures have shown weak explanatory power with limited adoption by practitioners (Glazier 2015; Hughes and Bartlett 2002; Mackenzie and Cushion 2013; Nevill et al. 2008; Sarmento et al. 2014; Tenga et al. 2010a, b). To circumvent this problem increasingly multi-variate approaches are being used to retain contextual information (Fernandez-Navarro et al. 2016; Kempe et al. 2014). Almeida et al. (2016) investigated the effect of different scoring modes on ball-recovery type and location, playing configuration and defensive state in youth players. The results showed that more ball recoveries were made when a central goal was used and that most recoveries were a result of set-play in the defensive third of the pitch. Younger players also produced more elongated shapes in the playing direction whereas the older teams produced more flattened shapes with larger spread in the direction orthogonal to the playing direction (Almeida et al. 2016). Tenga et al. (2010a, b) investigated the effects of a ten different variables on score-box possession based on video data from 163 matches from the Norwegian men's professional league in 2004. The results showed that the odds ratio for producing a score-box possession increased when the attacking team had a long possession, started their attack from the final third, or used penetrative passes against a balanced defense. However, counterattack, possession starting in the final third, long possession, long pass, and penetrative passes showed increased odds ratios against an imbalanced defense. Recently, Fernandez-Navarro et al. (2016) used 19 performance indicators to identify different playing styles. The results showed that several factors like possession directness which correlated with ball possession, sideway passes, and passes from the defensive third into the attacking third were important to identify playing styles (Fernandez-Navarro et al. 2016).

One approach which is increasingly being used to study team tactics is the team centroid method (Folgado et al. 2014; Frencken et al. 2011, 2012; Yue et al. 2008). Here the behavior of the team centroid, the geometric center of the positions of all players from a team, is used to analyze the behavior of the whole team. Results from this line of research indicate a strong coupling between team centroids during game play (Frencken et al. 2011), changes of inter-centroid distances due to pitch size variations (Duarte et al. 2012a, b; Frencken et al. 2013), and key game events like goal shots are accompanied by increased inter-team coupling variability (Frencken et al. 2012). More recently, investigation of centroid behavior has been further extended by calculating the

Approximate Entropy (ApEn) (Pincus and Goldberger 1994), a non-linear time-series measurement techniques, to quantify the regularity in time-series data (Aguiar et al. 2015; Goncalves et al. 2014; Sampaio and Macas 2012). Results using ApEn analysis suggest increased centroid behavior regularity after tactical training in novice players (Duarte et al. 2012a, b; Sampaio and Macas 2012). Goncalves et al. (2014) investigated the coordination during on 11-a-side game between and within the defenders, mid-fielders, and attacker subgroups using ApEn. The results showed that players movements were more regular with respect to the centroid of their respective groups compared to the other groups. Sampaio et al. (2014) further showed that during an inferiority situation during a 5-a-side small sided game the regularity of the distance to the team centroid increased. Goncalves et al. (2016) investigated the influence of numerical imbalances between attacking and defending team in small sided games in professional and amateur players. Player numbers varied between 4 versus 3, 4 versus 5, and 4 versus 7. The results showed that in experts an increase in the number of opponents increased the regularity in team behavior with respect to the opponents. Although the application of ApEn is becoming more prominent, it still remains to be shown what this measure really represents as the regularity behavior of team centroids in itself represent a highly abstract description of team behavior. Nevertheless, team centroid measures increasingly are being used to capture team behavior and many interesting applications have been reported in the literature in recent years.

Another more recent group of approach to study team tactics focuses on the control of space. On such approach uses for example the team surface area as calculated from the convex hull which encloses all players from one team (Frencken et al. 2011; Moura et al. 2012, 2013). Results from this line of research indicates that greater surface areas are covered by the attacking compared to the defensive teams (Frencken et al. 2011; Moura et al. 2012). Similar, more experienced players also cover a greater area compared to less experienced players (Duarte et al. 2012a, b; Olthof et al. 2015). Fradua et al. (2013) investigated the individual player area during 11-a-side matches by calculating the largest rectangle enclosing all field players divided by the number of players. The results showed that individual playing areas become smaller when the ball moved into the central pitch area. Another approach uses Voronoi-diagrams to investigate space control (Nakanishi et al. 2008). Here the controlled space is determined using the location and distances between individual players to determine the controlled space. Results using Voronoi-diagrams show similar results compared to the team surface area approach (Fonseca et al. 2012; Fujimura and Sugihara 2005; Gudmundsson and Wolle 2014; Kim 2004; Taki and Hasegawa 2000) Finally, another approach is based on the determination of numerical superiority in a particular pitch area (Silva et al. 2014). Together these results indicate that space control is a central aspect of soccer tactics and further highlight the interactive nature underlying soccer games (Duarte et al. 2013; Garganta 2009; Grehaigne et al. 1997; Tenga et al. 2010a, b).

Another emerging analysis approach to study team tactics studies investigates team passing behavior using network approaches (Watts and Strogatz 1998). The basic rationale of this approach is to model the players of a team as nodes and the passes occurring between them as weighted vertices where the number of passes between two players determine the weights (Duarte et al. 2012a, b; Passos et al. 2011). This representation of team passing behavior allows to easily identify key players within in a team as they display more connection to other vertices accompanied by greater vertex weights (Gama et al. 2014; Passos et al. 2011). Recent network analyses which included next to the player information also pass position information were able to predict game outcomes and the final ranking of the top teams using a K-Nearest Neighbor classifier (Cintia et al. 2015). Similar, Wang et al. (2015) used Bayesian latent model approach applied to passing network and passing position information from 241 games from the Spanish First (2013–2014). The obtained model was able to automatically identify different tactical patterns across teams. By combining the obtained tactical information with attacking success the authors were further able to show which specific tactical patterns were more efficient across teams. By investigating the contributions by the individual players to each tactical pattern the authors were further able to determine individual contributions by the players to each tactical pattern (Wang et al. 2015). Together these results suggest that players interactions mediated through passing behavior in combination with spatial information provides an interesting new approaches to analyze tactical behavior in elite soccer thereby providing much more information compared to traditional notational analysis approaches.

Increasingly tactical decision making in elite soccer is also investigated using machine learning (ML) algorithms based on game position data (Bialkowski et al. 2014a, b; Fernando et al. 2015; Xinyu et al. 2013). Machine learning algorithms allow to identify specific data patterns in large datasets by building an a priori unknown model from the data (Haykin 2009; Jordan and Mitchell 2015; Waljee and Higgins 2010). Although this approach has been discussed in sports research for some time (Bartlett 2004; Borrie et al. 2002; Nevill et al. 2008) only recently

successful applications become more prevalent (Bartlett 2004; Lucey et al. 2013a, b). For example, application of an expectation maximization algorithm with position data from an entire English Premier League season allowed the automatic identification of team formations (Bialkowski et al. 2014a, b; Lucey et al. 2013a, b). The results further showed that teams used more defensive formations during away games (Bialkowski et al. 2014a, b). The authors used a two-step algorithm where the formations were identified only after each player was assigned a specific role. This approach allowed the authors to circumvent the problem that the player' roles are not constant throughout the game but change according to the context which precludes the possibility to simply use the id of each individual player to identify team formations (Bialkowski et al. 2014a; Lucey et al. 2013a, b). Knauf et al. (2016) used a spatio-temporal kernel algorithm to cluster trajectories which allowed automatic differentiated game initiation and scoring opportunities from position data. Pairwise similarities between trajectories during attacking phases were compared using a specific metric and subsequently a clustering algorithms grouped the trajectories into clusters. Again, one of the underlying features of the algorithm used by the authors is that the comparison between trajectories is invariant to permutations between players (Knauf et al. 2016). Using spatial tracking data, Kihwan et al. (2010) applied a temporal kernel method to predict the location of the ball on the pitch. By calculating a flow-field from the running directions of the players the authors were able to determine convergence points of flow-field which predicted future positions of the ball with good agreement (Kihwan et al. 2010). Hirano and Tsumoto (2005) used a multiscale comparison technique with combined event data type and event location data to automatically identify reoccurring attacking sequences leading to a goal. The multiscale comparison technique allowed to compare event sequences of varying length with each other. For example, in the spatial-kernel method this problem has been resolved by time-normalizing the data (Knauf et al. 2016). Similar, Fernando et al. (2015) were able to differentiate attacking plays across teams using cluster analysis of game sequences (compare also Xinyu et al. 2013). Recently, Montoliu et al. (2015) applied a Bag-of-Words algorithm to coding soccer game video snippets followed by a Random Forest classifier to identify game play patterns. The authors divided the pitch into ten areas and calculated the optical flow representing the moving direction of players during short video sequences extracted from two complete soccer game recording. Thus, the application relied on the pre-segmentation of the raw video data by experts (Montoliu et al. 2015).

A second group of ML approaches featuring prominent in the soccer literature uses neural network modeling (compare Dutt-Mazumder et al. 2011 for a comprehensive overview). Here, in particular Kohonen Feature Maps (KFM) have been used to study tactical patterns (Barton et al. 2006; Bauer and Schöllhorn 1997; Dutt-Mazumder et al. 2011; Kohonen 1990, 2001; Lees and Barton 2003). For example, Grunz et al. (2012) used a Hierarchically Dynamically Controlled Network KFM (Perl 2002, 2004; Perl and Weber 2004) to automatically identify team formations (Grunz et al. 2012; Kempe et al. 2015; Memmert and Perl 2009). In summary, numerous machine learning studies of have used soccer data to study tactical decision making with little guidance for non-experts. Common to these approaches is that mostly a certain facet of team tactics, predominantly team formations, was investigated. Accordingly, information how to combine the information across tactical domains (Fig. 1) is lacking currently (Garganta 2009; Glazier 2015). For example it is not clear how group formations interact with the individual technical and tactical skills of players. As it is clear that different tactical positions within a team have different physiological demands there has been no research addressing how this information can be used in combination with tactical formations used by the attacking and defensive teams (Carling et al. 2008). Furthermore, with respect to the tactics hierarchy introduced in the introduction (compare also Fig. 1) the presented approaches work at the team tactics level. Accordingly, how team formations influence group tactics of subgroups and individual tactics has not been investigated so far. An interesting side-note of the presented studies is the fact that most ML soccer analyses are performed by computer scientist research group with little apparent involvement by sports scientists.

This short overview shows that although many interesting analyses are available what is lacking is a conceptual connection between them. Accordingly, it appears that the main obstacle to study team tactics stems from the lack of a theoretical model (Garganta 2009; Glazier 2015; Mackenzie and Cushion 2013). One model which has been repeatedly proposed in the literature is based on a Dynamic system theoretical framework (Duarte et al. 2012a, b; Duarte et al. 2013; Garganta 2009; McGarry et al. 2002; Reed and Hughes 2006; Ric et al. 2016). However, although this approach merits great potential, at present already the basic definition of a relevant phase space is lacking. In the dynamic systems theoretical approaches, the phase space constitutes a key concept which describes a theoretical abstractions describing mathematically a space where the system resides in and which enable to capture the dynamics of

the system in a meaningful manner (Nevill et al. 2008; Vogel 1999). Current suggestions regarding appropriate phase space variables in team game vary widely (Duarte et al. 2012a, b; Gréhaigne 2011; Grehaigne et al. 1997; Gréhaigne and Godbout 2014; Lames and McGarry 2007). In this regard, a common approach for example is to use the relative phase as a measure to capture coordination phenomena between players (Duarte et al. 2013; Goncalves et al. 2014; Sampaio and Macas 2012). Relative phase approaches stem from the domain of physical dynamical systems were oscillators typically constitute the building blocks of the systems (Pikovsky et al. 2003). Accordingly, the question of whether an oscillator assumption is justified to model team games is an open question at present. Modeling efforts of soccer games as a dynamic system which go beyond a purely phenomenological description are therefore not available at present.

The lack of a higher-order description about soccer team dynamics also prevents the current analytical approaches from making a real impact with practitioners (Carling et al. 2008; Lames and McGarry 2007; Nevill et al. 2008). One of the challenges for tactical match analysis in elite soccer will be to work towards an explanatory theoretical model which is able to integrate information from various domains including tactics, physiology, and motor skills (Garganta 2009; Sarmento et al. 2014) (compare Fig. 1). In this regard, new approaches in Artificial Intelligence (AI) research (Bishop 2013; Gibney 2016; Jones 2014; LeCun et al. 2015) may provide promising avenues towards the development of a theoretical model of tactical decision making in elite soccer. In particular, so-called deep learning networks are becoming increasingly powerful in modeling domains previously considered computational intractable (Hinton and Salakhutdinov 2006; LeCun et al. 2015; Xue-wen and Xiaotong 2014). However, these approaches rely on large training datasets to determine network parameters (Jones 2014; Xue-wen and Xiaotong 2014), which at present have not been used in tactical analyses in soccer. In this regard, recent machine learning models using neural networks have been extended such to allow to incorporate a priori information into the models (Bishop 2013). This might be of great relevance to develop novel approach to model team tactical behaviors as for example insights gained from the studies summarized above might be used to constrain network modeling efforts and at the same time allowing the connection between physiological, tactical and skill related information. Accordingly, modern algorithm from AI might prove highly useful for tactical analysis in elite soccer and fulfill previous proposals (Dutt-Mazumder et al. 2011).

## Big data and soccer tactics

A potential solution with respect to model building and the combination various data sources might present itself through the recent rise of big data technologies which has been already suggested as shaping the future of performance analysis in elite soccer (Cassimally 2012; Kasabian 2014; Lohr 2012; Medeiros 2014; Norton 2014). As the phenomenon of big data is relatively recent first a definition of the relevant concepts will be provided. Surprisingly, no universally agreed definition of big data is available and big data is rather described by its characteristics (Baro et al. 2015; Noor et al. 2015; Romanillos et al. 2016). Accordingly, big data is characterized using the so-called three V's: (1) Volume, (2) Variety and (3) Velocity (Noor et al. 2015; Xue-wen and Xiaotong 2014). Volume describes the magnitude of the data, Variety refers to the heterogeneity of data, and Velocity characterizes the data production rate (Noor et al. 2015). With respect to tactical analytics in soccer these concept can be mapped in the following way: (1) Volume refers to the size of datasets in soccer. For example, a current dataset for positional data typically encoded using Extensible Markup Language (XML) ranges between 86 and 300 megabytes (mb). Thus, storing position, event and video data from a single complete Bundesliga season results in 400 gigabytes of tracking data. Accordingly the data volume increases with the addition of other sources including for example physiological or event data. By itself this is far from the petabyte data sizes commonly associated with big data (Pääkkönen and Pakkala 2015), yet the main problem is to provide structured access to the data. Common solutions using Excel sheets do not scale well with these data. Big data technologies in contrast provide specific solutions for storing such data sets and make them accessible through specific user interfaces and application programming interfaces (API). (2) Variety refers to different data formats and data sources. Variety can be further distinguished into: (a) structured, (b) semi-structured, and (c) unstructured data. Structured data has a clearly predefined schema describing the data. Structured data allows simple navigation and searching through the data where a relational database system is the canonical example. In contrast, unstructured data lacks a definite schema with video data and text messages being typical examples. Accordingly, semi-structured data falls in between these two extremes and consists of data which lacks a pre-defined structure but may has a variable schema which is often part of the data itself (Sint et al. 2009). Current XML data types used for tracking data are examples in this regard (IPTC 2001). Thus, in soccer data variety refers to position, video, fitness, training, skill performance, and notational meta-data next to health records and crowd data from blogs. As data access and

data processing patterns vary across data types, big data technologies provide specific solutions to combine the information distributed across such datasets. (3) Velocity describes the speed with which novel data is being generated. In soccer, the velocity varies widely between real-time streams from physiological and positional data to delayed data from notational analysis during training and competition. Big data technologies specifically address how to process and store high velocity data. In summary, all three key concepts characterizing big data are highly relevant with respect to tactical analysis in elite soccer and big data technological stacks provide specific solutions to address each of these areas.

A candidate big data soccer technological stack for soccer tactics analyses should be organized along several levels (compare Fig. 2). First, the necessary infrastructure to collect the data is required spanning physiological and tracking data in addition to video and observational data. Second, a storage system is required allowing efficient data storage and access. Finally, a processing pipeline has to be established to extract relevant information from the data and to subsequently merge the information to build an explanatory and/or predictive model (Coutts 2014). For all these processing levels reporting and visualization capabilities are needed to monitor the different processing steps and communicate the results. Unfortunately, there is no one-to-one mapping between these different components and available technologies. However an in-depth discussion of specific technological solutions is beyond the scope of the present article and more specialized literature is referred to (Noor et al. 2015; Pääkkönen and Pakkala 2015; Sitto and Presser 2015).

Yet, what immediately becomes clear from Fig. 2 is that a significant amount of expertise is needed in order to

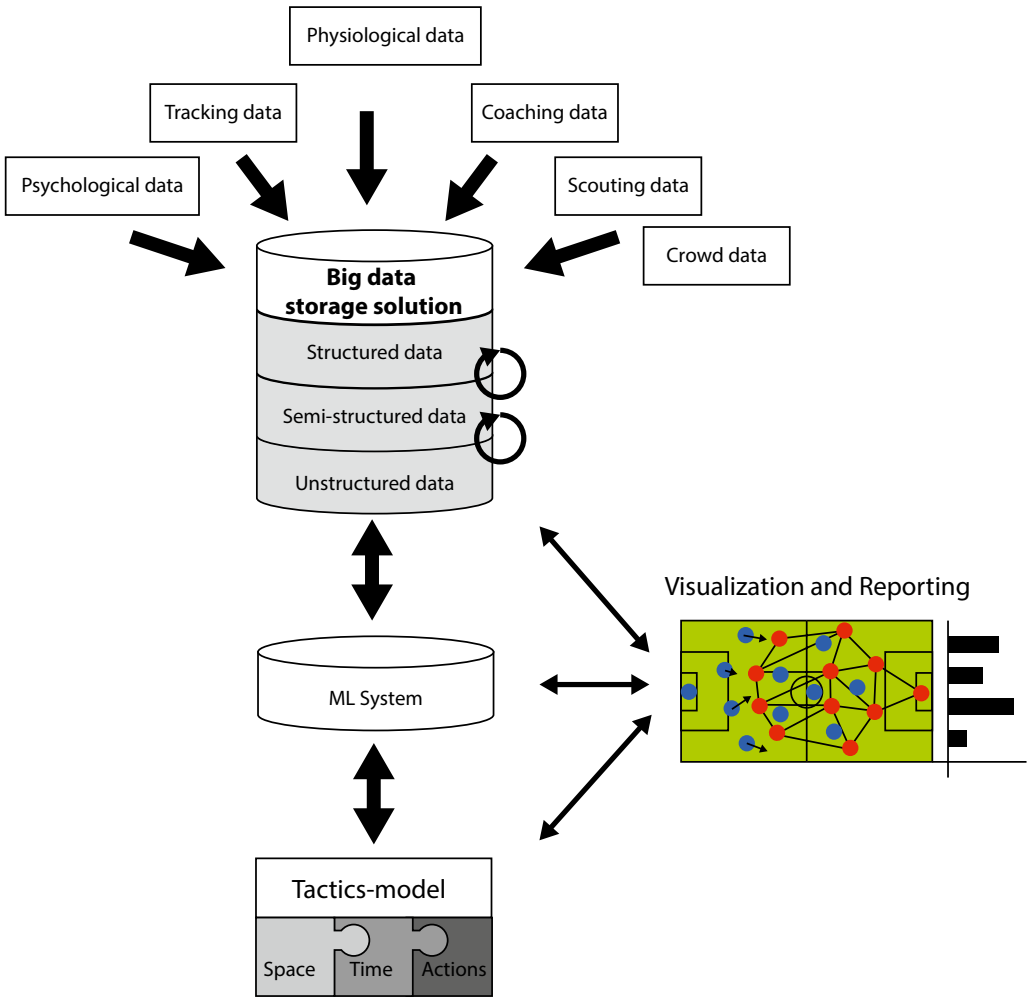


**Fig. 2** Big data technological stack for tactical analysis in elite soccer

establish such a system. One area which is facing similar challenges in this respect is the medical health sector (Noor et al. 2015; Toga et al. 2015; Zhang et al. 2015). In the medical area a so-called personalized (stratified) medicine is increasingly seen is a key are of research to improve current practices (Hood et al. 2015; Kostkova et al. 2016; Zhang et al. 2015). Thereby, for personalized medicine to become realizable big data technologies are needed. One key problem in this area is how data is stored and shared across institutions. At present health data is collected and held by government, commercial and public research institutions. This leads to sever limitations with respect to access and data sharing possibilities across these entities due to privacy and security issues (Costa 2014; Kong and Xiao 2015; Kostkova et al. 2016; Toga and Dinov 2015). This also applies to soccer data where data is collected by commercial institutions, private clubs, and public research institutions. Accordingly, privacy issues have to be addressed as for example detailed profiles about individual players might have significant career implications and professional soccer teams may be reluctant to share data and possibly forfeit competitive advantages. Thus, data governance issues must be resolved before big data approaches may become viable for soccer research potentially. In the medical sector varies solutions are being investigated including standardized open privacy protection mechanisms which encrypts individual data items (Kong and Xiao 2015). Nevertheless, even when access is made available, researchers face the problem that data processing is highly complex and not manageable using common processing pipelines. Experiences from the biomedical sectors shows that in particular smaller research groups lack the required expertise and funding to build the required processing and analysis infrastructures (Bishop 2013; Goecks et al. 2010; Lynch 2008; Marx 2013; Noor et al. 2015; Sitto and Presser 2015). At present, it is also not clear how to ensure that technologies and procedures are made available to researchers lacking the required computer science expertise to build data pipelines of their own. This is already a problem with respect to many of the ML techniques described above.

As computational approaches increasingly become more complex reproducibility issue will also become more important as the development of novel algorithmic approach will become the focus of future publication results (Mesirov 2010). In this regard, efforts from biomedical research like the Galaxy project (Goecks et al. 2010) may provide a model solution for future big data technologies in sports sciences. The Galaxy project is developed through a collaborative effort across several universities and provides a web-based solution to perform genomic research using big data technologies (Goecks et al. 2010; Levine and Hullett 2002; Ohmann et al. 2015). The project aims to provide a standardized way for researchers to access complex processing algorithms which makes it possible for non-expert users to apply cutting edge analysis technologies to their data (Goecks et al. 2010). The system includes a sophisticated documentation solution which allows the storage and presentation of analysis results and documents at the same time the complete processing pipeline ensuring reproducibility of the research results (Goecks et al. 2010). The framework was build to be extensible and allows the inclusion of additional procedures through public repositories efforts (Blankenberg et al. 2014). This approach may be a model for sports sciences to address not only big data approaches for soccer tactics but more general analysis and data processing problems in other domains as well. Inevitable this will lead to increased collaborative efforts between sports and computer scientists as the sports science community at present lacks the required computational background.

## Conclusion

In conclusion, exciting times are emerging for team sports performance analysis as more and more data is going to become available allowing more refined investigations. The adaption of big data technologies for soccer research may therefore provide solutions to some of the key issues outline above. Thus, by providing novel methods to analyze the data and a more comprehensive theoretical model and understanding of tactical team performance in elite soccer may be within reach. This implies however, that future soccer research will have to embrace a stronger multi-disciplinary approach. Performance analysts, exercise scientists, biomechanists as well as practitioners will have to work together to make sense of these complex data sets. As has been pointed out, most of the machine learning approaches presented were performed by computer science research groups. Accordingly, future collaborations between computer and sports scientists may hold the key to apply these complex approaches in a more relevant manner. In turn, relying increasingly on more complex data analysis techniques will also pose new challenges for future sports scientists. Therefore, university curricula will have be augmented to ensure that future students receive the required background training to be able to not only use these techniques but to have at least some understanding of their theoretical and computational underpinnings. The introduction of big data technologies will also require a discussions within the research community of how to share data and techniques across research teams. To make the new insights relevant for practice a tight interchange with practitioners is required. Finally, taking a broader view

on the issue of big data and sports science the proposed model for tactical analyses in elite soccer might also prove beneficial for other sports science domains where data sizes are bound to increase as well and accordingly similar problems will surface.

**Authors' contributions**
Both authors worked equally on all parts of the article. Both authors read and approved the final manuscript.

**Acknowledgements**
This work was supported by a grant from the German Research Foundation (Deutsche Forschungs Gesellschaft, ME 2678/3-3) to the second author.

**Competing interests**
The authors declare that they have no competing interests.

Received: 11 May 2016 Accepted: 19 August 2016
Published online: 24 August 2016

**References**

Aguiar M, Goncalves B, Botelho G, Lemmink K, Sampaio J (2015) Footballers' movement behaviour during 2-, 3-, 4- and 5-a-side small-sided games. J Sports Sci 33(12):1259–1266. doi:10.1080/02640414.2015.1022571

Almeida CH, Duarte R, Volossovitch A, Ferreira AP (2016) Scoring mode and age-related effects on youth soccer teams' defensive performance during small-sided games. J Sports Sci 34(14):1355–1362. doi:10.1080/02640414.2016.1150602

Appelboom G, LoPresti M, Reginster JY, Sander Connolly E, Dumont EP (2014) The quantified patient: a patient participatory culture. Curr Med Res Opin 30(12):2585–2587. doi:10.1185/03007995.2014.954032

Araújo D, Davids K, Hristovski R (2006) The ecological dynamics of decision making in sport. Psychol Sport Exerc 7(6):653–676

Baca A (2008) Tracking motion in sport—trends and limitations. Paper presented at the 9th Australasian conference on mathematics and computers in sport, Math Sport (ANZIAM)

Baca A, Baron R, Leser R, Kain H (2004) A process oriented approach for match analysis in table tennis. In: Lees A, Kahn JF, Maynard IW (eds) Science and racket sports III. Routledge, Abingdon, pp 214–219

Balagué N, Torrents C (2005) Thinking before computing: changing perspectives in sport performance. Int J Comput Sci Sport 4:5–13

Baro E, Degoul S, Beuscart R, Chazard E (2015) Toward a literature-driven definition of big data in healthcare. Biomed Res Int 2015:639021. doi:10.1155/2015/639021

Barris S, Button C (2008) A review of vision-based motion analysis in sport. Sports Med 38(12):1025–1043. doi:10.2165/00007256-200838120-00006

Bartlett R (2004) Artificial intelligence in technique analysis—past, present and future. Int J Perf Anal Sport 4(2):4–19

Barton G, Lees A, Lisboa PJG, Attfield S (2006) Visualisation of gait data with Kohonen self-organising neural maps. Gait Posture 24:46–53

Bauer HU, Schöllhorn W (1997) Self-organizing maps for the analysis of complex movement patterns. Neural Process Lett 5(3):193–199

Beetz M, Kirchlechner B, Lames M (2005) Computerized real-time analysis of football games. IEEE Pervasive Comput 4(3):33–39. doi:10.1109/MPRV.2005.53

Bialkowski A, Lucey P, Carr P, Yue Y, Matthews I (2014a) Win at Home and Draw Away: automatic formation analysis highlighting the differences in home and away team behaviors MIT Sloan Sports Analytics Conference. Boston

Bialkowski A, Lucey P, Carr P, Yue Y, Sridharan S, Matthews I (2014b) Large-scale analysis of soccer matches using spatiotemporal tracking data. In: 2014 IEEE international conference on paper presented at the data mining (ICDM). 14–17 Dec 2014

Bisanz G, Gerisch G (1980) Fußball: Training, Technik, Taktik. Rororo, Hamburg

Bishop CM (2013) Model-based machine learning. Philos Trans A Math Phys Eng Sci 371(1984):20120222. doi:10.1098/rsta.2012.0222

Blankenberg D, Von Kuster G, Bouvier E, Baker D, Afgan E, Stoler N, Nekrutenko A (2014) Dissemination of scientific software with Galaxy ToolShed. Genome Biol 15(2):403. doi:10.1186/gb4161

Bloomfield J, Polman R, O'Donoghue P (2007) Physical demands of different positions in FA Premier League soccer. J Sports Sci Med 6(1):63–70

Borrie A, Jonsson GK, Magnusson MS (2002) Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. J Sports Sci 20(10):845–852. doi:10.1080/026404102320675675

Buchheit M, Allen A, Poon TK, Modonutti M, Gregson W, Di Salvo V (2014) Integrating different tracking systems in football: multiple camera semi-automatic system, local position measurement and GPS technologies. J Sports Sci 32(20):1844–1857. doi:10.1080/02640414.2014.942687

Buntain C (2014) Language-agnostic event detection across sports from twitter and using temporal features. Paper presented at the workshop on large-scale sports analytics (KDD 2014), New York, USA

Bush M, Barnes C, Archer DT, Hogg B, Bradley PS (2015) Evolution of match performance parameters for various playing positions in the English Premier League. Hum Mov Sci 39:1–11. doi:10.1016/j.humov.2014.10.003

Carling C (2011) Influence of opposition team formation on physical and skill-related performance in a professional soccer team. Eur J Sport Sc 11(3):155–164. doi:10.1080/17461391.2010.499972

Carling C, Williams AM, Reilly T (2005a) From technical and tactical performance analysis to training drills Handbook of soccer match analysis: a systematic approach to improving performance. Routledge, London, pp 129–147

Carling C, Williams AM, Reilly T (2005b) Handbook of soccer match analysis. Routledge, London

Carling C, Williams AM, Reilly T (2005c) What match analysis tells us about succeessful strategy and tactics in soccer Handbook of soccer match analysis: a systematic approach to improving performance. Routledge, London, pp 108–128

Carling C, Bloomfield J, Nelsen L, Reilly T (2008) The role of motion analysis in elite soccer: contemporary performance measurement techniques and work rate data. Sports Med 38(10):839–862

Carling C, Wright C, Nelson LJ, Bradley PS (2014) Comment on 'performance analysis in football: a critical review and implications for future research'. J Sports Sci 32(1):2–7. doi:10.1080/02640414.2013.807352

Cassimally KA (2012) Soccer's big data revolution. http://www.nature.com/scitable/blog/labcoat-life/soccers_big_data_revolution

Castellano J, Alvarez-Pastor D, Bradley PS (2014) Evaluation of research using computerised tracking systems (Amisco and Prozone) to analyse physical performance in elite soccer: a systematic review. Sports Med 44(5):701–712. doi:10.1007/s40279-014-0144-3

Cintia P, Pappalardo L, Pedreschi D, Giannotti F, Malvaldi M (2015) The harsh rule of the goals: data-driven performance indicators for football teams. In: IEEE international conference on paper presented at the data science and advanced analytics (DSAA), 2015. 36678 2015. 19–21 Oct 2015

Collet C (2013) The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. J Sports Sci 31(2):123–136. doi:10.1080/02640414.2012.727455

Costa FF (2014) Big data in biomedicine. Drug Discov Today 19(4):433–440. doi:10.1016/j.drudis.2013.10.012

Coutts AJ (2014) Evolution of football match analysis research. J Sports Sci 32(20):1829–1830. doi:10.1080/02640414.2014.985450

D'Orazio T, Leo M (2010) A review of vision-based systems for soccer video analysis. Pattern Recogn 43(8):2911–2926

da Mota GR, Thiengo CR, Gimenes SV, Bradley PS (2016) The effects of ball possession status on physical and technical indicators during the 2014 FIFA World Cup Finals. J Sports Sci 34(6):493–500. doi:10.1080/02640414.2015.1114660

di Salvo V, Modonutti M (2009) Integration of different technology systems for the development of football training. J Sports Sci Med S11:3

Drust B, Atkinson G, Reilly T (2007) Future perspectives in the evaluation of the physiological demands of soccer. Sports Med 37(9):783–805

Duarte R, Araujo D, Correia V, Davids K (2012a) Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. Sports Med 42(8):633–642. doi:10.2165/11632450-000000000-00000

Duarte R, Araujo D, Freire L, Folgado H, Fernandes O, Davids K (2012b) Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. Hum Mov Sci 31(6):1639–1651. doi:10.1016/j.humov.2012.03.001

Duarte R, Araujo D, Correia V, Davids K, Marques P, Richardson MJ (2013) Competing together: assessing the dynamics of team-team and player-team synchrony in professional association football. Hum Mov Sci 32(4):555–566. doi:10.1016/j.humov.2013.01.011

Dutt-Mazumder A, Button C, Robins A, Bartlett R (2011) Neural network modelling and dynamical system theory: are they relevant to study the governing dynamics of association football players? Sports Med 41(12):1003–1017. doi:10.2165/11593950-000000000-00000

Ehrmann FE, Duncan CS, Sindhusake D, Franzsen WN, Greene DA (2016) GPS and injury prevention in professional soccer. J Strength Condit Res 30(2):360–367. doi:10.1519/JSC.0000000000001093

Fernandez-Navarro J, Fradua L, Zubillaga A, Ford PR, McRobert AP (2016) Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. J Sports Sci. doi:10.1080/02640414.2016.1169309

Fernando T, Wei X, Fookes C, Sridharan S, Lucey P (2015) Discovering methods of scoring in soccer using tracking data. Paper presented at the Large-Scale Sports Analytics, Sidney

Folgado H, Lemmink KA, Frencken W, Sampaio J (2014) Length, width and centroid distance as measures of teams tactical performance in youth football. Eur J Sport Sc 14(Suppl 1):S487–S492. doi:10.1080/17461391.2012.730060

Fonseca S, Milho J, Travassos B, Araujo D (2012) Spatial dynamics of team sports exposed by Voronoi diagrams. Hum Mov Sci 31(6):1652–1659. doi:10.1016/j.humov.2012.04.006

Fradua L, Zubillaga A, Caro O, Ivan Fernandez-Garcia A, Ruiz-Ruiz C, Tenga A (2013) Designing small-sided games for training tactical aspects in soccer: extrapolating pitch sizes from full-size professional matches. J Sports Sci 31(6):573–581. doi:10.1080/02640414.2012.746722

Frencken W, Lemmink K, Delleman N, Visscher C (2011) Oscillations of centroid position and surface area of soccer teams in small-sided games. Eur J Sport Sci 11(4):215–223. doi:10.1080/17461391.2010.499967

Frencken W, Poel H, Visscher C, Lemmink K (2012) Variability of inter-team distances associated with match events in elite-standard soccer. J Sports Sci 30(12):1207–1213. doi:10.1080/02640414.2012.703783

Frencken W, Plaats J, Visscher C, Lemmink K (2013) Size matters: pitch dimensions constrain interactive team behaviour in soccer. J Syst Sci Complex 26(1):85–93. doi:10.1007/s11424-013-2284-1

Fujimura A, Sugihara K (2005) Geometric analysis and quantitative evaluation of sport teamwork. Syst Comput Jpn 36(6):49–58. doi:10.1002/scj.20254

Gama J, Passos P, Davids K, Relvas H, Ribeiro J, Vaz V, Dias G (2014) Network analysis and intra-team activity in attacking phases of professional football. Int J Perform Anal Sport 14(3):692–708

Garganta J (2009) Trends of tactical performance analysis in team sports: bridging the gap between research, training and competition. Rev Port Cien Desp 9(1):81–89

Gibney E (2016) Google AI algorithm masters ancient game of Go. Nature 529(7587):445–446. doi:10.1038/529445a

Glazier PS (2015) Towards a grand unified theory of sports performance. Hum Mov Sci. doi:10.1016/j.humov.2015.08.001

Godin F, Zuallaert J, Verndersmissen B, De Neve W, Van der Waller R (2014) Beating the bookmakers: leveraging statistics and Twitter microposts for predicting soccer results. Paper presented at the Workshop on Large-Scale Sports Analytics (KDD 2014), New York, USA

Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11(8):R86. doi:10.1186/gb-2010-11-8-r86

Goncalves B, Figueira BE, Macas V, Sampaio J (2014) Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. J Sports Sci 32(2):191–199. doi:10.1080/02640414.2013.816761

Goncalves B, Marcelino R, Torres-Ronda L, Torrents C, Sampaio J (2016) Effects of emphasising opposition and cooperation on collective movement behaviour during football small-sided games. J Sports Sci 34(14):1346–1354. doi:10.1080/02640414.2016.1143111

Gréhaigne J-F (2011) Jean-paul sartre and team dynamics in collective sport. Sport Ethics Philos 5(1):34–45. doi:10.1080/17511321.2010.536956

Gréhaigne J-F, Godbout P (1995) Tactical knowledge in team sports from a constructivist and cognitivist perspective. Quest 47(4):490–505. doi:10.1080/00336297.1995.10484171

Gréhaigne J-F, Godbout P (2014) Dynamic systems theory and team sport coaching. Quest 66(1):96–116. doi:10.1080/00336297.2013.814577

Grehaigne J-F, Bouthier D, David B (1997) Dynamic-system analysis of opponent relationships in collective actions in soccer. J Sports Sci 15(2):137–149. doi:10.1080/026404197367416

Grunz A, Memmert D, Perl J (2012) Tactical pattern recognition in soccer games by means of special self-organizing maps. Hum Mov Sci 31(2):334–343. doi:10.1016/j.humov.2011.02.008

Gudmundsson J, Wolle T (2014) Football analysis using spatio-temporal tools. Comput Environ Urban Syst 47:16–27

Harley JA, Lovell RJ, Barnes CA, Portas MD, Weston M (2011) The interchangeability of global positioning system and semiautomated video-based performance data during elite soccer match play. J Strength Cond Res 25(8):2334–2336. doi:10.1519/JSC.0b013e3181f0a88f

Haykin S (2009) Neural networks and learning machines, 3rd edn. Pearson, Upper Saddle River

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507. doi:10.1126/science.1127647

Hirano S, Tsumoto S (2005, 6–9 Nov. 2005) Grouping of soccer game records by multiscale comparison technique and rough clustering. Fifth international conference on paper presented at the hybrid intelligent systems, 2005. HIS '05

Hood L, Lovejoy JC, Price ND (2015) Integrating big data and actionable health coaching to optimize wellness. BMC Med 13:4. doi:10.1186/s12916-014-0238-7

Hughes MD, Bartlett RM (2002) The use of performance indicators in performance analysis. J Sports Sci 20(10):739–754. doi:10.1080/026404102320675602

Hughes MD, Franks I (2005) Analysis of passing sequences, shots and goals in soccer. J Sports Sci 23(5):509–514. doi:10.1080/02640410410001716779

Ingebrigtsen J, Dalen T, Hjelde GH, Drust B, Wisloff U (2015) Acceleration and sprint profiles of a professional elite football team in match play. Eur J Sport Sc 15(2):101–110. doi:10.1080/17461391.2014.933879

IPTC (2001) SportsML. http://dev.iptc.org/SportsML

James N (2006) The role of notational analysis in soccer coaching. Int J Sports Sci Coach 1(2):185–198. doi:10.1260/174795406777641294

James N, Mellalieu SD, Hollely C (2002) Analysis of strategies in soccer as a function of European and domestic competition. Int J Perform Anal Sport 2(1):85–103

Jones N (2014) Computer science: the learning machines. Nature 505(7482):146–148. doi:10.1038/505146a

Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349(6245):255–260. doi:10.1126/science.aaa8415

Kannekens R, Elferink-Gemser MT, Visscher C (2011) Positioning and deciding: key factors for talent development in soccer. Scand J Med Sci Sports 21(6):846–852. doi:10.1111/j.1600-0838.2010.01104.x

Kasabian R (2014) World cup: assist goes to big data. Information week. http://www.informationweek.com/big-data/big-data-analytics/world-cup-assist-goes-to-big-data/a/d-id/1278822

Kempe M, Vogelbein M, Memmert D, Nopp S (2014) Possession vs. direct play: evaluating tactical behavior in elite soccer. Int J Sport Sci 4(6A):35–41

Kempe M, Grunz A, Memmert D (2015) Detecting tactical patterns in basketball: comparison of merge self-organising maps and dynamic controlled neural networks. Eur J Sport Sci 15(4):249–255. doi:10.1080/17461391.2014.933882

Kihwan K, Grundmann M, Shamir A, Matthews I, Hodgins J, Essa I (2010) Motion fields to predict play evolution in dynamic sport scenes. Paper presented at the IEEE CVPR. 13–18 June 2010

Kim S (2004) Voronoi analysis of a soccer game. Nonlinear Anal Model Control 9(3):233–240

Knauf K, Memmert D, Brefeld U (2016) Spatio-temporal convolution kernels. Mach Learn 102(2):247–273. doi:10.1007/s10994-015-5520-1

Kohonen T (1990) The self-organizing map. Proc IEEE 78(9):1464–1480

Kohonen T (2001) Self-organizing maps, 3rd edn. Springer, Berlin

Kong G, Xiao Z (2015) Protecting privacy in a clinical data warehouse. Health Inf J 21(2):93–106. doi:10.1177/1460458213504204

Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, Tooke J (2016) Who owns the data? Open data for healthcare. Front Public Health 4:7. doi:10.3389/fpubh.2016.00007

Lago C (2009) The influence of match location, quality of opposition, and match status on possession strategies in professional association football. J Sport Sci 27(13):1463–1469. doi:10.1080/02640410903131681

Lames M, McGarry T (2007) On the search for reliable performance indicators in game sports. Int J Perform Anal Sport 7(1):62–79

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. doi:10.1038/nature14539

Lees A, Barton G (2003) A characterization of technique in the soccer kick using a Kohonen neural network analysis. J Sports Sci 22:491–492

Leser R, Baca A, Ogris G (2011) Local positioning systems in (game) sports. Sensors (Basel) 11(10):9778–9797. doi:10.3390/s111009778

Levine TR, Hullett CR (2002) Eta squared, partial eta squared, and misreporting of effect size in communication research. Hum Commun Res 28(4):612–625. doi:10.1111/j.1468-2958.2002.tb00828.x

Lewis M (2004) Money ball: the art of winning an unfair game. Norton & Company, Newyork

Liu H, Gomez MA, Lago-Penas C, Sampaio J (2015) Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. J Sports Sci 33(12):1205–1213. doi:10.1080/02640414.2015.1022578

Lohr S (2012) The age of big data. New York Times. http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=0

Lu WL, Ting JA, Little JJ, Murphy KP (2013) Learning to track and identify players from broadcast sports videos. IEEE Trans Pattern Anal Mach Intell 35(7):1704–1716. doi:10.1109/TPAMI.2012.242

Lucey P, Bialkowski A, Carr P, Morgan S, Matthews I, Sheikh Y (2013) Representing and discovering adversarial team behaviors using player roles. Paper presented at the IEEE CVPR. 23–28 June 2013

Lucey P, Oliver D, Carr P, Roth J, Matthews I (2013) Assessing team strategy using spatiotemporal data. Paper presented at the 19th ACM SIGKDD, Chicago, Illinois, USA

Lynch C (2008) Big data: How do your data grow? Nature 455(7209):28–29

Mackenzie R, Cushion C (2013) Performance analysis in football: a critical review and implications for future research. J Sports Sci 31(6):639–676. doi:10.1080/02640414.2012.746720

Marx V (2013) Biology: the big challenges of big data. Nature 498(7453):255–260. doi:10.1038/498255a

McGarry T, Anderson DI, Wallace SA, Hughes M, Franks IM (2002) Sport competition as a dynamical self-organizing system. J Sports Sci 20:771–781

Medeiros J (2014) The winning formula: data analytics has become the latest tool keeping football teams one step ahead. Wired. http://www.wired.co.uk/magazine/archive/2014/01/features/the-winning-formula

Memmert D, Perl J (2009) Analysis and simulation of creativity learning by means of artificial neural networks. Hum Mov Sci 28(2):263–282. doi:10.1016/j.humov.2008.07.006

Mesirov JP (2010) Computer science. Accessible reproducible research. Science 327(5964):415–416. doi:10.1126/science.1179653

Mohr M, Krustrup P, Bangsbo J (2005) Fatigue in soccer: a brief review. J Sports Sci 23(6):593–599. doi:10.1080/02640410400021286

Montoliu R, Martin-Felez R, Torres-Sospedra J, Martinez-Uso A (2015) Team activity recognition in Association Football using a Bag-of-Words-based method. Hum Mov Sci 41:165–178. doi:10.1016/j.humov.2015.03.007

Moura FA, Martins LE, Anido Rde O, de Barros RM, Cunha SA (2012) Quantitative analysis of Brazilian football players' organisation on the pitch. Sports Biomech 11(1):85–96. doi:10.1080/14763141.2011.637123

Moura FA, Martins LE, Anido RO, Ruffino PR, Barros RM, Cunha SA (2013) A spectral analysis of team dynamics and tactics in Brazilian football. J Sports Sci 31(14):1568–1577. doi:10.1080/02640414.2013.789920

Nakanishi R, Murakami K, Naruse T (2008) Dynamic positioning method based on dominant region diagram to realize successful cooperative play. In: Visser U, Ribeiro F, Ohashi T, Dellaert F (eds) Robo cup 2007: Robot Soccer World Cup XI, Vol 5001. Springer, Berlin, pp 488–495

Nevill A, Atkinson G, Hughes MD (2008) Twenty-five years of sport performance research in the Journal of Sports Sciences. J Sport Sci 26(4):413–426. doi:10.1080/02640410701714589

Noor AM, Holmberg L, Gillett C, Grigoriadis A (2015) Big data: the challenge for small research groups in the era of cancer genomics. Br J Cancer 113(10):1405–1412. doi:10.1038/bjc.2015.341

Norton S (2014) Germany's 12th man at the World Cup: Big Data. CIO Journal. http://blogs.wsj.com/cio/2014/07/10/germanys-12th-man-at-the-world-cup-big-data/

Ohmann C, Canham S, Danielyan E, Robertshaw S, Legre Y, Clivio L, Demotes J (2015) 'Cloud computing' and clinical trials: report from an ECRIN workshop. Trials 16:318. doi:10.1186/s13063-015-0835-6

Olthof SB, Frencken WG, Lemmink KA (2015) The older, the wider: on-field tactical behavior of elite-standard youth soccer players in small-sided games. Hum Mov Sci 41:92–102. doi:10.1016/j.humov.2015.02.004

Pääkkönen P, Pakkala D (2015) Reference architecture and classification of technologies, products and services for big data systems. Big Data Res 2(4):166–186

Passos P, Davids K, Araujo D, Paz N, Minguens J, Mendes J (2011) Networks as a novel tool for studying team ball sports as complex social systems. J Sci Med Sport 14(2):170–176. doi:10.1016/j.jsams.2010.10.459

Perl J (2002) Game analysis and control by means of continuously learning networks. Int J Perform Anal Sport 2(1):21–35

Perl J (2004) A neural network approach to movement pattern analysis. Hum Mov Sci 23:605–620

Perl J, Weber K (2004) A neural network approach to pattern learning in sport. Int J Comput Sci Sport 3(1):67–70

Pikovsky A, Rosenblum M, Kurths J (2003) Synchronization: a universal concept in nonlinear sciences. Cambridge University Press, Cambridge

Pincus SM, Goldberger AL (1994) Physiological time-series analysis: what does regularity quantify. Am J Physiol 266:1643–1656

Rampinini E, Coutts AJ, Castagna C, Sassi R, Impellizzeri FM (2007) Variation in top level soccer match performance. Int J Sports Med 28(12):1018–1024. doi:10.1055/s-2007-965158

Reed D, Hughes MD (2006) An exploration of team sport as a dynamical system. Int J Perform Anal Sport 6(2):114–125

Ric A, Hristovski R, Goncalves B, Torres L, Sampaio J, Torrents C (2016) Time-scales for exploratory tactical behaviour in football small-sided games. J Sports Sci. doi:10.1080/02640414.2015.1136068

Romanillos G, Zaltz Austwick M, Ettema D, De Kruijf J (2016) Big data and cycling. Trans Rev 36(1):114–133. doi:10.1080/01441647.2015.1084067

Sampaio J, Macas V (2012) Measuring tactical behaviour in football. Int J Sports Med 33(5):395–401. doi:10.1055/s-0031-1301320

Sampaio J, Lago C, Goncalves B, Macas VM, Leite N (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. J Sci Med Sport 17(2):229–233. doi:10.1016/j.jsams.2013.04.005

Sarmento H, Marcelino R, Anguera MT, CampaniCo J, Matos N, LeitAo JC (2014) Match analysis in football: a systematic review. J Sports Sci 32(20):1831–1843. doi:10.1080/02640414.2014.898852

Shafizadehkenari M, Lago-Penas C, Gridley A, Platt GK (2014) Temporal analysis of losing possession of the ball leading to conceding a goal : a study of the incidence of perturbation in soccer. Int J Sports Sci Coach 9(4):363–627

Shull PB, Jirattigalachote W, Hunt MA, Cutkosky MR, Delp SL (2014) Quantified self and human movement: a review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. Gait Posture 40(1):11–19. doi:10.1016/j.gaitpost.2014.03.189

Silva P, Travassos B, Vilar L, Aguiar P, Davids K, Araujo D, Garganta J (2014) Numerical relations and skill level constrain co-adaptive behaviors of agents in sports teams. PLoS One 9(9):e107112. doi:10.1371/journal.pone.0107112

Sint R, Stroka S, Schaffert S, Ferstl R (2009) Combining unstructured, fully structured and semi-structured information in semantic wikis. Paper presented at the Semantic Wikis

Sitto K, Presser M (2015) Field guide to hadoop: an introduction to hadoop, its ecosystem, and aligned technologies. O'Reilly and Associates, Sebastopol

Taki T, Hasegawa J (2000) Visualization of dominant region in team games and its application to teamwork analysis. Proceedings of the paper presented at the computer graphics international, 2000

Tenga A, Holme I, Ronglan LT, Bahr R (2010a) Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. J Sports Sci 28(3):245–255. doi:10.1080/02640410903502766

Tenga A, Ronglan LT, Bahr R (2010b) Measuring the effectiveness of offensive match-play in professional soccer. Eur J Sport Sci 10(4):269–277. doi:10.1080/17461390903515170

Toga AW, Dinov ID (2015) Sharing big biomedical data. J Big Data. doi:10.1186/s40537-015-0016-1

Toga AW, Foster I, Kesselman C, Madduri R, Chard K, Deutsch EW, Hood L (2015) Big biomedical data as the key resource for discovery science. J Am Med Inform Assoc 22(6):1126–1131. doi:10.1093/jamia/ocv077

Valter DS, Adam C, Barry M, Marco C (2006) Validation of Prozone ®: a new video-based performance analysis system. Int J Perform Anal Sport 6(1):108–119

Vogel H (1999) Gerthsen Physik, 20th edn. Springer, Berlin

Vogelbein M, Nopp S, Hokelmann A (2014) Defensive transition in soccer—are prompt possession regains a measure of success? A quantitative analysis of German Fussball-Bundesliga 2010/2011. J Sports Sci 32(11):1076–1083. doi:10.1080/02640414.2013.879671

Waljee AK, Higgins PD (2010) Machine learning in medicine: a primer for physicians. Am J Gastroenterol 105(6):1224–1226. doi:10.1038/ajg.2010.173

Wang Q, Zhu H, Hu W, Shen Z, Yao Y (2015) Discerning tactical patterns for professional soccer teams: an enhanced topic model with applications. Paper presented at the Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442. doi:10.1038/30918

Xinyu W, Long S, Lucey P, Morgan S, Sridharan S (2013, 26–28 Nov. 2013) Large-scale analysis of formations in Soccer. In: 2013 international conference on paper presented at the digital image computing: techniques and applications (DICTA)

Xue-wen C, Xiaotong L (2014) Big data deep learning: challenges and perspectives. Access IEEE 2:514–525. doi:10.1109/ACCESS.2014.2325029

Yiannakos A, Armatas V (2006) Evaluation of the goal scoring patterns in European Championship in Portugal 2004. Int J Perform Anal Sport 6(1):178–188

Yu Y, Wang X (2015) World cup 2014 in the twitter world. Comput Hum Behav 48(C):392–400. doi:10.1016/j.chb.2015.01.075

Yue Z, Broich H, Seifriz F, Mester J (2008) Mathematical analysis of a Soccer game. Part I: individual and collective behaviors. Stud Appl Math 121(3):223–243. doi:10.1111/j.1467-9590.2008.00413.x

Zhang Y, Zhu Q, Liu H (2015) Next generation informatics for big data in precision medicine era. Bio Data Min 8:34. doi:10.1186/s13040-015-0064-2