Memmert D[1], Rein R[1]

REVIEW

ACCEPTED: February 2018
PUBLISHED ONLINE: March 2018
DOI: 10.5960/dzsm.2018.322
Memmert D, Rein R. Match analysis, Big Data and tactics: current trends in elite soccer. Dtsch Z Sportmed. 2018; 69: 65-72.

Match Analysis, Big Data and Tactics: Current Trends in Elite Soccer

Spielanalyse, Big Data und Taktik: Aktuelle Trends im Profi-Fußball

1. DEUTSCHE SPORTHOCHSCHULE KÖLN, *Institut für Trainingswissenschaft und Sportinformatik, Köln*

Summary

- **In the past few years,** there has been a real revolution in the field of match analysis. New approaches to evaluate performance have been developed especially in commercial football. This has been driven by novel developments in sensor technology and a changing coverage of sports games in media. Compared to analyses that depend on video material, positional data of individual players and the ball allow a much more in-depth analysis of sports games.
- **Previously,** performance analysis relied mainly on frequency distributions of certain game events. In contrast, the novel approaches allow calculating more complex metrics. This helps to measure and identify the performance of teams and individual players and especially how teams interact. However, the actual significance of many of these newer performance indicators has often not been sufficiently scientifically validated.
- **To show how this can possibly be done,** the present work introduces some new performance indicators in football where evidence has already been established. Furthermore, performance analysis is increasingly connected to the field of Big Data. Therefore, in addition this present paper introduces a model to connect Big Data and match analysis and illustrates the resulting implications regarding future training practices.

KEY WORDS:
Technological Developments, Football, Sport Games, Positional Data

Zusammenfassung

- **Im Bereich der Spielanalyse** hat in den letzten Jahren eine echte Revolution stattgefunden. Basierend auf Weiterentwicklungen in der Sensortechnologie, vor allem im Bereich des kommerziellen Fußballs, gepaart mit Veränderung in der medialen Aufbereitung von Sportspielen, haben sich in den letzten Jahren neue Arten der Leistungsbewertung etabliert. Ausgehend von den Positionsdaten der einzelnen Spieler und des Balls sind nun deutlich spezifischere Analysen möglich als auf der Basis von Videomaterial.
- **Während früher** die Analyse von Häufigkeiten bestimmter Spielereignisse im Mittelpunkt stand, ist jetzt die Berechnung von komplexen Leistungsparametern möglich und auch schon in der Anwendung. Diese ermöglichen die Spielleistung von Mannschaften und individuellen Spielern und vor allem das Interaktionsverhalten zwischen Mannschaften deutlich differenzierter zu bestimmen. Allerdings ist die tatsächliche Bedeutsamkeit für die Spielleistung vieler dieser neuen Performance-Indikatoren oftmals noch unzureichend wissenschaftlich abgesichert.
- **In der vorliegenden Arbeit** werden daher verschiedene Indikatoren im Fußball vorgestellt für die bereits Evidenz besteht, um mögliche methodische Herangehensweisen aufzuzeigen. In diesem Zusammenhang wird auch ein Modell beschrieben, um die immer wichtiger werdende Verknüpfung zwischen der Spielanalyse und dem Bereich Big Data näher zu spezifizieren und die Möglichkeiten erläutert, die dies in den nächsten Jahren für die Spielanalyse hinsichtlich der Trainingssteuerung bieten.

SCHLÜSSELWÖRTER:
Technologische Entwicklung, Fußball, Sportspiele, Positionsdaten

QR-Code scannen und Artikel online lesen.

CORRESPONDING ADDRESS:
Prof. Dr. Daniel Memmert
Deutsche Sporthochschule Köln
Institut für Trainingswissenschaft
und Sportinformatik
Am Sportpark Müngersdorf 6, 50933 Köln
: memmert@dshs-koeln.de

Introduction

The digital revolution has already reached professional soccer some time ago and the same applies to traditional match observation and performance analysis. These days, the athletes' every action on the field is recorded – sometimes manually, sometimes using (semi-) automatic procedures. Their statistical evaluation and interpretation play an outstanding role in all age groups (but especially in seniors) and at every performance level (in particular at the professional level) (39). While the usual trivial analyses (such as pass frequencies) or presentation formats (e.g. heatmaps) often enable only limited insight into soccer at the elite level, advanced performance indicators (Key Performance Indicators, abbreviation: KPI) model particular aspects of modern soccer and provide a basis for scientific, data-driven analyses (40). To give match analyses of the top clubs physiological (6), technical (1) and tactical (35) criteria as a valid instrument and to determine and control game performance, broad tests in practice and

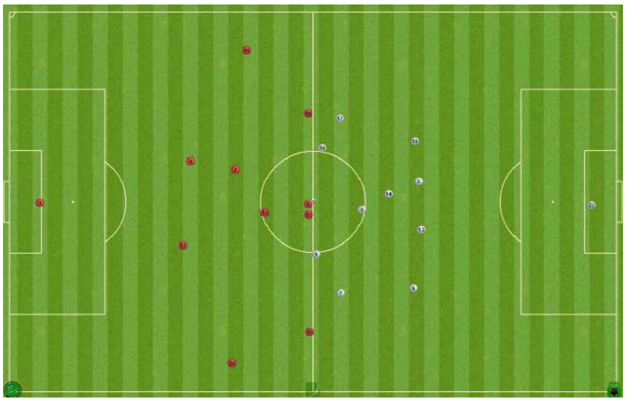

Figure 2

Graphic processing of the positional data in the soccer-WM 2006, Finale: Italy-France 6:4 This was one of the first games for which positional data were available.

evaluation studies are of central importance to confirm the efficacy of these parameters (32).

At present, in soccer several performance factors can be directly investigated (17, 18, 34, 36, 37, 38, 45, 48), due to the availability of their required data in digital form. Hereby, the computer starts to replace the analysts with respect to the evaluation of game situations, yielding more objective records of soccer-specific events. The advantage over traditional video-based procedures is that the identification of individual game sequences in soccer games must no longer be performed manually, but can be (semi-)automatically performed using algorithms supported through modern Big Data methods. Using neural nets, analyses of positional data can be performed within seconds with a high success rate (more than 85% agreement with the expert's call (17, 18)) for complex events such as various group tactical behaviors like game opening, wing play, or standard situations (further differentiated in throw-in, free kick and corner). This enables classification of large data quantities within few minutes using event specific differences and commonality metrics. Professional soccer therefore has crossed the threshold into the digital age and currently is going through a transition from traditional, rather qualitative analysis methods to modern, data-driven game analysis techniques (see Figure 1).

Modern match analysis consists thereby of much more than the evaluation of classical game characteristics like ball possession, pass behaviors, and man-on-man quota and kilometers run. Recent studies showed that these often isolated data have very little predictive power regarding the difference between winning or losing (4, 14, 24, 26). Rather, positional data are increasingly used to obtain new knowledge in elite soccer. This trend is made possible by increasingly more reliable tracking procedures. These enable very precise recording of the players' positions on the field within the model uncertainty for the parameter player position. Retrospection-free optical systems with the help of special cameras and, nowadays, sensor-based techniques are used. The latter moreover enable the recording of height-coordinates of the ball, in cases where the ball is equipped with the appropriate sensors.

The analysis of positional data (see Figure 2), which are nowadays commonly recorded in all top European leagues, has led to the establishment of an independent discipline within match analysis. The decision of the International Football Association Board (IFAB) to permit wearable tracking units even in regular competitive games means another increase in the quantity of data which will be available in future and the expectation of an attendant increase in analysis interest. The precise recording of players' and ball positions enables thereby to perform considerably more complex, dynamic analyses which include also the interactions between the teams using novel performance indicators (39). It can be assumed that the development and validation of ever more-advanced KPIs will be performed in future to obtain objective, data-driven information about the complex individual (e.g. individual pressing behavior) and group-tactical (e.g. space control of a team in the opponents 16m space) player performances.

A recognizable research deficit at present, however, is the lack of broad field tests. Few studies so far have used large numbers of games when applying these novel KPIs (49). In a survey article (35), two initial KPIs were listed which are currently being developed and tested. However, neither of them has yet been investigated or validated using a Big-Data approach. For this reason, a selection of KPIs will be discussed below which have been tested and validated for the criterion win/loss on the basis of a large quantity of data.

New KPIs Based on Positional Data in the Bundesliga

A total of 50 national-league matches (N=186 halftimes) from the 2014/15 season were analyzed in one of the broadest Big-Data field study performed to date. In total, 11,160 performance values were calculated and subsequently evaluated (40). Central to the analysis was the application of the software tool SOCCER (46, 47) which has been developed in a collaborative effort over several years. SOCCER enables the combination of conventional data analyses, dynamic status-event modelling and artificial neural nets. All calculations are based solely on positional data, for which sufficient high reliability and validity were proven (5, 7, 10). This also means that events like ball possession, gain or loss are automatically calculated from player and ball positions. The following information was deduced from the processed positional data concerning player-ball relationships: ball contact, possession gain, ball reception, ball possession, pass, and possession loss. SOCCER allows to evaluate performance of individual players, for tactical groups, and for the entire team. Furthermore, the software allows to assess space-control either for arbitrary positions on the field or critical areas (e.g. 16-m space) with respect to the frequency at any time point and any time duration (e.g. 1st /2nd halftime). Results can be presented as distributions. For the present purpose, three KPIs were the focus of interest:

KPI 1 – Space Control

Voronoi diagrams were used for quantification of the space controlled by a player or a team (15, 22, 54). To this end, the playing field is divided into individual segments and the exact proportion of segments each player can reach before all other players – and thus are controlled by this player – can be determined at any point in time (compare Figure 3).

KPI 2 – Outplayed Opponents

This indicator records the pass effectiveness (especially in the assessment of vertical passes) of a team. A virtual line parallel to the opponents' goal line which runs exactly through the ball carrying player is generated. Subsequently, the number of opponents between this line and the goal line at pass initiation and at pass completion is determined. This number is not only a measure for the average number of defending players, but by calculating the difference during these instances information of how many opponents were outplayed is obtained.

KPI 3 – Pressing Index

The average speed of all players with respect to the ball is calculated to measure the pressing behavior or the defensive team during the transition phase after losing ball possession. Measurements are made at three different timepoints: 1) immediately after loss of the ball. 2) several seconds after the team lost control of the ball. 3) after the ball had reached a certain distance from the point of loss. Thus both temporal and spatial aspects of the switch-off process are taken into account.

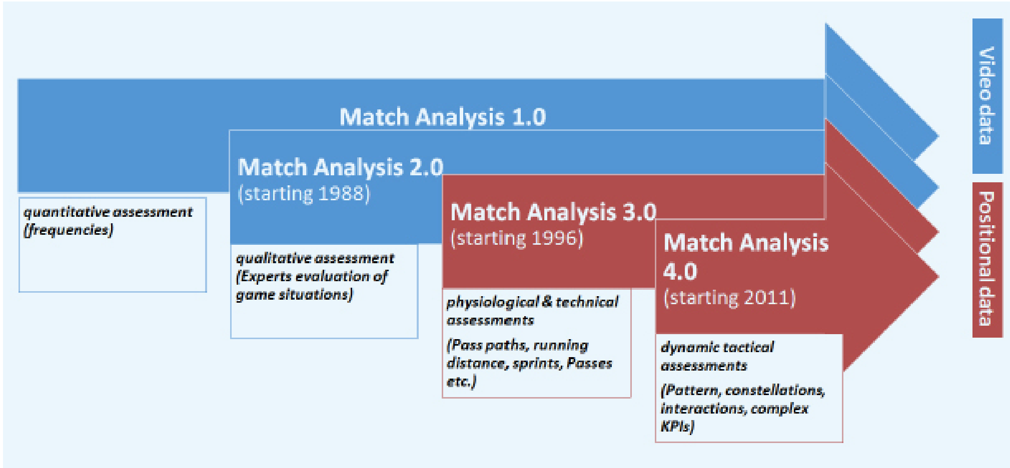

Figure 1

From game analysis 1.0 to game analysis 4.0 (Memmert & Raabe, 2017).

In the field study presented with exploratory character, the following central results pattern could be identified:

KP 1: the best teams (upper vs. lower third of the final standings table) control the most space, especially in their own half (W=2257, p<0.05), and in the critical zones of the playing field (30m area before the goal, penalty area, W=2267, p<0.05). This also applies to clear wins (=difference of at least two goals) (W=239, p<0.05).

KP 2: Successful teams overall outplayed more opponents than losing teams (W=2107, p<0.05). In particular, during clear wins (W=260, p<0.05). Compared to the losers, the successful teams on average face fewer opponents when attacking; even for vertical passes made in the offensive area (W=1263, p<0.05).

KP 3: Independent of playing strength, the losing team significantly presses more often during games won with a difference of at least two goals (W=271, p<0.05). Possibly, because they increase pressure to attempt to gain ball control or the superior team may reduce pressure somewhat as they are in the lead. However, if one compares the best with the worst teams (upper third of the final standings table), the better teams show clearly better pressing values across a whole season compared to teams at the bottom third of the table (W=1379, p<0.05).

Big Data-Technology in Professional Soccer

As already mentioned above, increasingly Big Data technologies are under discussion, especially already in the media, to be used in elite soccer (7, 18, 23, 29, 40). However, there has been no sports science based discussion about the extent to which Big Data technologies are really relevant for sports match analysis or could possibly become relevant. In order to discuss this appropriately however, several clarifications regarding various terms associated with Big Data are necessary.

However, a more in-depth discussion of Big Data technologies is beyond the scope of the present article (52). Surprisingly, there is not yet any universally-recognized definition of the term “Big Data”. Rather, the article Big Data is described based on certain characteristics (42, 50). Central to this characterization are the so-called three Vs: 1) Volume, 2) Variety and 3) Velocity (42, 57). Volume describes the scope of the data, Variety the heterogeneity of the data, while velocity describes the data production rate (42).

With reference to soccer analyses, these concepts can be illustrated as follows: 1) Volume: A positional dataset for a single soccer game is currently coded typically using Extensible Markup Language (XML) and fluctuates between few up to and several hundred megabytes. This number does not appear excessively large at first but the increase is linear with the number of games and additional training sessions. If an entire Bundesliga season is analyzed (306 games) up to 90 gigabytes of positional data may be obtained. If additional data from training sessions, physiological markers, event data (passes, fouls, etc.) and video-data are integrated, the data quantity easily increases to several terrabytes. Obviously, this represents a qualitative leap ›

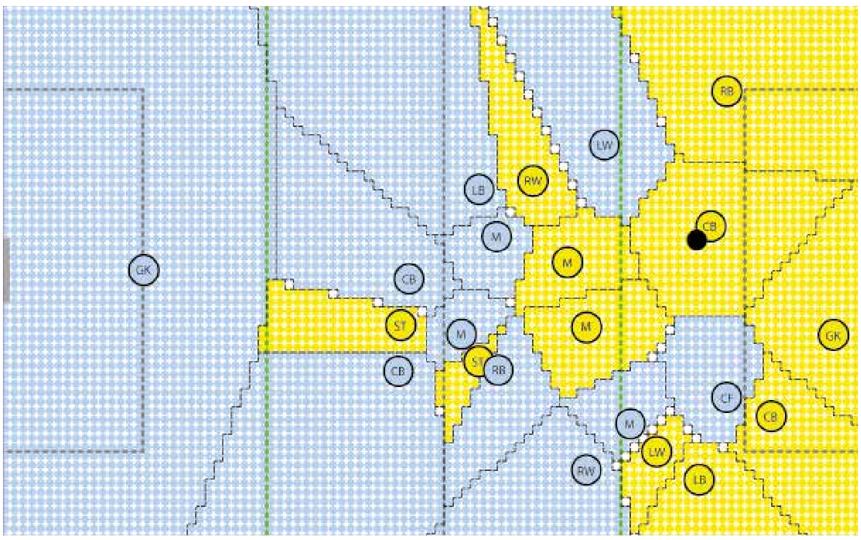

Figure 3

Graphic processing of the positional data in the soccer-WM 2006, Finale: Italy-France 6:4 This was one of the first games for which positional data were available.

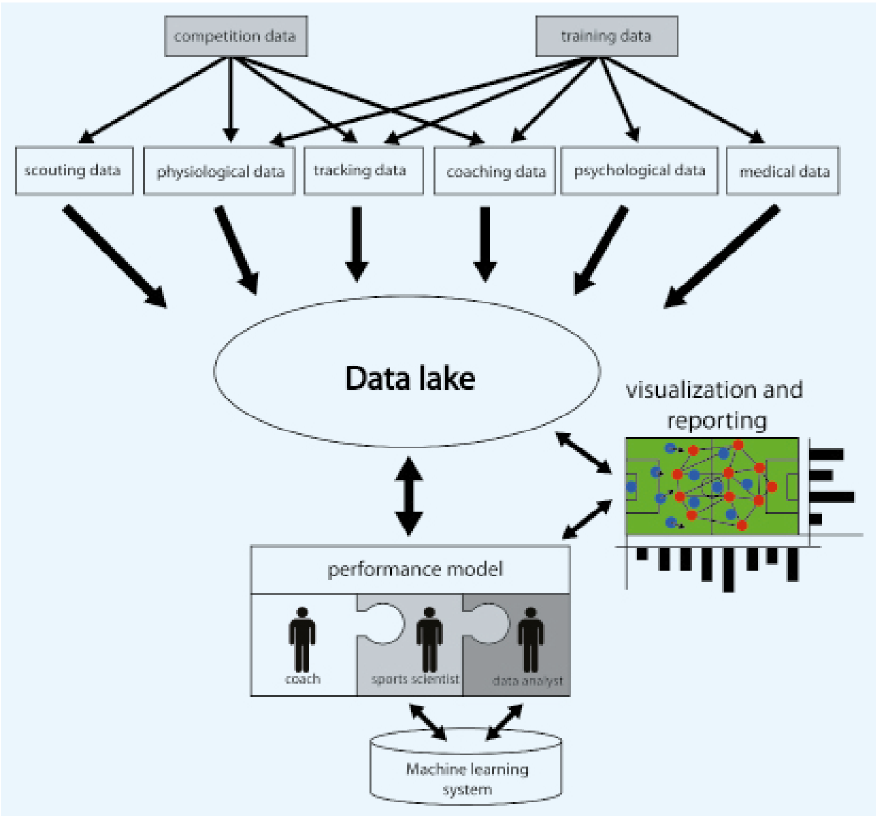

Figure 4

A possible model for Big-Data analyses in professional elite soccer. (adapted from Rein & Memmert, 2016).

compared to the data quantities collected in the notation procedures used thus far which usually can be fitted on simple excel sheets.

2) Variety: As described above, a variety of different data formats and data resources are available in soccer. For example, positional data, video recordings, meta-data in written form and sports-medical data recorded during the competition. These can be expanded as needed by additional training data based on skill tests, psychological tests, and additional performance data (e.g. observations made by the trainer, tactic tests, health reports), in the form of questionnaires, interviews and other methods. The analyst thus receives a bouquet of qualitative and quantitative data in a wide variety of quality and data densities. In data science, data types are divided into: a) structured, b) semi-structured, and c) not structured data types. This classification takes into account that different storage and analytical methods are used depending on the data type (51).

Structured data are characterized through clearly-defined and fixed data schemas. The schema describes which data contents are present and their relationship to one another. The presence of a fixed data pattern enables the user to index the data in a simple way and perform searches using simple queries (e.g. using the SQL database language) or perform further analyses. Examples are classical relational database systems, like the open-source MySQL Database.

Unstructured data, by contrast, have no defined pattern and are thus considerably more difficult to search and process for specific information. Typical examples here are video-data and written reports. Semi-structured data lie in-between these two extremes and may have schemas which for example, only describe a portion of the data (51). A typical example of semi-structured data are the XML data formats currently used for positional and event files (21).

Since the nature of the queries and subsequent data processing procedures differ between these three data types, dealing with such datasets is accordingly highly demanding for the analyst. Moreover, the greatest advantage of Big Data analyses lies not solely in the scope of the underlying data, but through the linking of the information across sources and data types more interesting insights can be obtained.

3) Finally, velocity describes the speed at which new data are generated. In soccer, the velocity of various datasets varies widely and fluctuates between real-time data, like physiological data and positional data, and time-delayed data, such as in notation analyses during training and matches. However, as mentioned earlier, a development can be observed nowadays that earlier time-consuming hand notations are increasingly performed by means of automatized routines (18, 19, 28).

All three Big Data technologies are characterized by key concepts and are thus very important for the future analysis of data in soccer and other kinds of sports. This means that modern match analysts must be knowledge in computer science as well as in sports science in order to not only be able to deal with this trend but also to make innovative contributions in the future as it seems reasonable that within the next couple of years tactical analyses in elite soccer will be increasingly performed using Big Data technologies (49). Although it is evident that there are already research groups working on individual solutions (33), there is as of now no structured approach that recommends how Big Data technologies can be used to perform analyses which are science based and of practical relevance. For this reason, it appears very important that additional efforts be made in this area so that scientific research remains relevant for the user. An analysis model for soccer should incorporate a variety of data sources reflecting the recent developments in data recording technologies. The central purpose of such a model is to combine information from the various areas so that conclusions can be drawn about game performance – individual as well as team performance (see Figure 4).

According to this overview, two challenges arise for Big Data solution in soccer research: First in providing the required infrastructure and second regarding development of appropriate algorithms and processing procedures. To begin, the required data processing infrastructure must be created which enables simple storage of and subsequent access to the collected data (data entry). As described above, the data consist in equal proportions of raw data and already-processed data and could be stored, for example in a centrally-accessible unit in the form of a so-called "data lake", to make queries as simple as possible. Care should be taken that links between the available data, which are initially isolated (e.g. different measurement apparatuses and measuring methods) are stored together with the data to guarantee that contextual information is across data sources is maintained and remains available in later processing stages. The loss of contextual information has been identified as one of the main criticisms of the traditional analysis methods used to date (30). Building on this, a processing layer must be constructed to first extract relevant pieces of information from the data and then combine (processing) and visualize them (reporting).

A possible solution, to somewhat reduce the resultant analysis complexity, could be achieved by using machine learning methods to generate data-driven models (13). Modern machine learning approaches enable to incorporate specific constraints and external information into the underlying analysis model (2). Such an approach has the advantage that previous research results as well as expert entries from practice can be taken equally into account when creating specific data models. This should also be useful in weakening possible criticisms from practitioners by incorporating their input directly into the analysis process. Overall, however, it must be noted that one should not be misled by new technologies like Big Data and Machine Learning. The final determining performance variable in soccer – a goal or winning a match – will always remain largely a product of chance (28). Thus prediction of individual match events will always remain a challenge.

An area which is facing similar hurdles with respect to data analysis is medical or biotechnological research (42,55,58). In particular, the area of precision medicine provides a possible model for sports sciences in this respect (19, 58). Here, experience has shown, that smaller research groups are most prone to face great difficulties to participate in these transformations of how research is being done. Reasons being that the technical expertise as well as the financial support required for creating the needed processing infrastructure are often lacking (29, 42,31,52, 2, 16). This applies to sports science as well and gives rise to the question of how newly-developed technologies can be made accessible to other research groups. This suggests that solutions should be sought based cooperation across research institutions inasfar as possible. Consequently, future collaborations between sport sciences and computer science will gain further importance to introduce the necessary expertise into the sport sciences domain. In this respect, multinational and international sports science majors educational program with a computer science component must be offered to keep up with the general trend toward digitalization in society.

Conclusion

The real-time character and the resulting instant feedback opportunities through the use of positional tracking data technologies has the potential to complete transform how training and matches are organized in elite sports contexts. Therefore, positional data can be seen as a significant addition to traditional data recording approaches in elite sports.

At present however, the full potential of objective performance assessment through digital data is a long way from being achieved and experience has shown that current offers for scientifically sound performance analyses, especially using positional data, fall often short with respect to the needs of practitioners in elite sports settings. Consequently, on the one hand, new educational programs are required (e.g. continued education "Match Analysis" at the German Sport University Cologne) to provide the required personnel and on the other hand much further work from sports science research is required to obtain reliable insight into the processes underlying successful soccer performance drawing from from computer science and statistics. The central questions in these efforts therefore is how to model soccer in such a way that relevant conclusions can be drawn supporting modern coaching approaches.

We believe that Big Data based on positional data tracking methods will be a key component to support this program of research in the future. Based on current discussions within the sport sciences (7, 30, 56), some tactical variables can already be used for managing training and competition in youth and professional soccer. But in the future, additional theoretical and conceptional considerations and discussions between sport science and youth and performance soccer will certainly be necessary.

The trench between "Theory and Practice" cannot presently be considered overcome at present, but there are numerous ongoing cooperations between science, the leagues, and the professional clubs, which are encouraging for future Big Data projects in team games. Finding the right approach to successfully use Big Data will be of central importance for future success.

Acknowledgements

This research was supported by a grant from the Deutschen Fußball Liga (DFL) und der Deutschen Forschungsgesellschaft (DFG; ME 2678/3-3) to the first author, respectively.

Conflict of Interest

The authors have no conflict of interest.

References

(1) **ALI A.** Measuring soccer skill performance: a review. Scand J Med Sci Sports. 2011; 21: 170-183. doi:10.1111/j.1600-0838.2010.01256.x

(2) **BISHOP CM.** Model-based machine learning. Philos Trans A Math Phys Eng Sci. 2013; 371: 20120222.

(3) **BLANKENBERG D, VON KUSTER G, BOUVIER E, BAKER D, AFGAN E, STOLER N; GALAXY TEAM, TAYLOR J, NEKRUTENKO A.** Dissemination of scientific software with Galaxy ToolShed. Genome Biol. 2014; 15: 403. doi:10.1186/gb4161

(4) **BRADLEY PS, LAGO-PEÑAS C, REY E, GOMEZ DIAZ A.** The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. J Sports Sci. 2013; 31: 1261-1270. doi:10.1080/02640414.2013.786185

(5) **BUCHHEIT M, ALLEN A, POON TK, MODONUTTI M, GREGSON W, DI SALVO V.** Integrating different tracking systems in football: multiple camera semi-automatic system, local position measurement and GPS technologies. J Sports Sci. 2014; 32: 1844-1857. doi:10.1080/02640414.2014.942687

(6) **BUSH M, BARNES C, ARCHER DT, HOGG B, BRADLEY PS.** Evolution of match performance parameters for various playing positions in the English Premier League. Hum Mov Sci. 2015; 39: 1-11. doi:10.1016/j.humov.2014.10.003

(7) **CARLING C, BLOOMFIELD J, NELSEN L, REILLY T.** The role of motion analysis in elite soccer. Sports Med. 2008; 38: 839-862. doi:10.2165/00007256-200838100-00004

(8) **CARLING C, WRIGHT C, NELSON LJ, BRADLEY PS.** Comment on 'Performance analysis in football: A critical review and implications for future research'. J Sports Sci. 2014; 32: 2-7. doi:10.1080/02640414.2013.807352

(9) **CASSIMALLY KA.** Soccer's big data revolution. 2012. http://www.nature.com/scitable/blog/labcoat-life/soccers_big_data_revolution. [22nd May 2017].

(10) **CASTELLANO J, ALVAREZ-PASTOR D, BRADLEY PS.** Evaluation of research using computerised tracking systems (Amisco® and Prozone®) to analyse physical performance in elite soccer: A systematic review. Sports Med. 2014; 44: 701-712. doi:10.1007/s40279-014-0144-3

(11) **COCK PJ, ANTAO T, CHANG JT, CHAPMAN BA, COX CJ, DALKE A, FRIEDBERG I, HAMELRYCK T, KAUFF F, WILCZYNSKI B, DE HOON MJ.** Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009; 25: 1422-1423. doi:10.1093/bioinformatics/btp163

(12) **COSTA FF.** Big data in biomedicine. Drug Discov Today. 2014; 19: 433-440. doi:10.1016/j.drudis.2013.10.012

(13) **COUTTS AJ.** Evolution of football match analysis research. J Sports Sci. 2014; 32(20): 1829-1830. doi:10.1080/02640414.2014.985450

(14) **DELLAL A, CHAMARI K, WONG DP, AHMAIDI S, KELLER D, BARROS R, BISCIOTTI GN, CARLING C.** Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. Eur J Sport Sci. 2011; 11: 51-59. doi:10.1080/17461391.2010.481334

(15) **FUJIMURA A, SUGIHARA K.** Geometric analysis and quantitative evaluation of sport teamwork. Syst Comput Jpn. 2005; 36: 49-58. doi:10.1002/scj.20254.

(16) **GOECKS J, NEKRUTENKO A, TAYLOR J, GALAXY T.** Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010; 11: R86. doi:10.1186/gb-2010-11-8-r86

(17) **GRUNZ A, MEMMERT D, PERL J.** Analysis and Simulation of Actions in Games by Means of Special Self-Organizing Maps. Int J Comp Sci Sport. 2009; 8: 22-36.

(18) **GRUNZ A, MEMMERT D, PERL J.** Tactical pattern recognition in soccer games by means of special self-organizing maps. Hum Mov Sci. 2012; 31: 334-343. doi:10.1016/j.humov.2011.02.008

(19) **HOERNIG M, LINK D, HERRMANN M, RADIG B, LAMES M.** Detection of Individual Ball Possession in Soccer. In: Chung P., Soltoggio A., Dawson C., Meng Q., Pain M. (eds) Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS). Advances in Intelligent Systems and Computing. 2016; vol 392. Springer, Cham.

(20) **HOOD L, LOVEJOY JC, PRICE ND.** Integrating big data and actionable health coaching to optimize wellness. BMC Med. 2015; 13: 4. doi:10.1186/s12916-014-0238-7

(21) **IPTC SPORTSML DEVELOPER SITE.** http://dev.iptc.org/SportsML. [20th July 2017].

(22) **KASABIAN R.** World Cup: Assist goes to big data. 2014. http://www.informationweek.com/big-data/big-data-analytics/world-cup-assist-goes-to-big-data/a/d-id/1278822 [22nd May 2017].

(23) **KIM S.** Voronoi analysis of a soccer game. Nonlinear Anal. Model. Control. 2004; 9: 233-240.

(24) **LAGO C, CASAIS L, DOMINGUEZ E, SAMPAIO J.** The effects of situational variables on distance covered at various speeds in elite soccer. Eur J Sport Sci. 2010; 10: 103-109. doi:10.1080/17461390903273994

(25) **LEVINE TR, HULLETT CR.** Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. Hum Commun Res. 2002; 28: 612-625. doi:10.1111/j.1468-2958.2002.tb00828.x

(26) **LIU H, GOMEZ MÁ, LAGO-PEÑAS C, SAMPAIO J.** Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. J Sports Sci. 2015; 33: 1205-1213. doi:10.1080/02640414.2015.1022578.

(27) **LOHR S.** The age of big data. New York Times. Feb 12, 2012: SR1.

(28) **LUCEY P, BIALKOWSKI A, CARR P, MORGAN S, MATTHEWS I, SHEIKH Y.** Representing and discovering adversarial team behaviors using player roles. IEEE CVPR. 2013; 2013: 2706-2713.

(29) **LYNCH C.** Big data: How do your data grow? Nature. 2008; 455: 28-29. doi:10.1038/455028a

(30) **MACKENZIE R, CUSHION C.** Performance analysis in football: a critical review and implications for future research. J Sports Sci. 2013; 31: 639-676. doi:10.1080/02640414.2012.746720

(31) **MARX V.** Biology: The big challenges of big data. Nature. 2013; 498: 255-260. doi:10.1038/498255a

(32) **MCGARRY T, O'DONOGHUE P, SAMPAIO J.** Routledge Handbook of Sports Performance Analysis. Abingdon, GBR: Taylor & Francis Group; 2015.

(33) **MEDEIROS J.** The winning formula: data analytics has become the latest tool keeping football teams one step ahead. Wired. 2014. http://www.wired.co.uk/magazine/archive/2014/01/features/the-winning-formula [20th July 2017].

(34) **MEMMERT D, BISCHOF J, ENDLER S, GRUNZ A, SCHMID M, SCHMIDT A, PERL J.** World-level analysis in top level football, Analysis and simulation of football specific group tactics by means of adaptive neural networks. In: Hui CLP, ed. Artificial Neural Networks – Application. InTech; 2011: 3-12. ISBN: 978-953-307-188-6.

(35) **MEMMERT D, LEMMINK KAPM, SAMPAIO J.** Current Approaches to Tactical Performance Analyses in Soccer using Position Data. Sports Med. 2017; 47: 1-10. doi:10.1007/s40279-016-0562-5

(36) **MEMMERT D, PERL J.** Analysis of Game Creativity Development by Means of Continuously Learning Neural Networks. In: Moritz EF, Haake S, ed. The Engineering of Sport 6. New York, NY: Springer; 2006: 261-266.

(37) **MEMMERT D, PERL J.** Analysis and Simulation of Creativity Learning by Means of Artificial Neural Networks. Hum Mov Sci. 2009; 28: 263-282. doi:10.1016/j.humov.2008.07.006

(38) **MEMMERT D, PERL J.** Game Creativity Analysis by Means of Neural Networks. J Sports Sci. 2009; 27: 139-149. doi:10.1080/02640410802442007

(39) **MEMMERT D, RAABE D.** Revolution im Profifußball. Mit Big Data zur Spielanalyse 4.0. Berlin: Springer; 2017.

(40) **MEMMERT D, RAABE, D, KNYAZEV, A, FRANZEN A, ZEKAS L, REIN R, PERL J, WEBER H.** Big Data im Profi-Fußball. Analyse von Positionsdaten der Fußball-Bundesliga mit neuen innovativen Key Performance Indikatoren. Leistungssport. 2016a; 46: 1-13.

(41) **MESIROV JP.** Computer science. Accessible reproducible research. Science. 2010; 327: 415-416. doi:10.1126/science.1179653

(42) **NOOR AM, HOLMBERG L, GILLETT C, GRIGORIADIS A.** Big Data: the challenge for small research groups in the era of cancer genomics. Br J Cancer. 2015; 113: 1405-1412. doi:10.1038/bjc.2015.341

(43) **NORTON S.** Germany's 12th man at the World Cup: Big Data. CIO Journal 2014. http://blogs.wsj.com/cio/2014/07/10/germanys-12th-man-at-the-world-cup-big-data/ [20th July 2017].

(44) **OHMANN C, CANHAM S, DANIELYAN E, ROBERTSHAW S, LEGRE Y, CLIVIO L, DEMATES J.** ‚Cloud computing' and clinical trials: report from an ECRIN workshop. Trials. 2015; 16: 318. doi:10.1186/s13063-015-0835-6

(45) **PERL J, MEMMERT D.** Net-Based Game Analysis by Means of the Software Tool SOCCER. IJCSS. 2011; 10: 77-84.

(46) **PERL J, GRUNZ A, MEMMERT D.** Tactics in soccer: an advanced approach. IJCSS. 2013; 12: 33-44.

(47) **PERL J, MEMMERT D, BACA A, ENDLER S, GRUNZ A, REBEL M, SCHMIDT A.** Sensors, Monitoring, and Model-Based Data Analysis in Sports, Exercise and Rehabilitation. In: Lai DTH, Begg RK, Palaniswami M, ed. Sensor Networks – challenges towards practical application. Boca Raton, FL: Taylor and Francis: CRC Press; 2011: 237-247.

(48) **PERL J, TILP M, BACA A, MEMMERT D.** Neural Networks for Analysing Sports Games. In: McGarry T, O'Donoghue P, Sampaio J, ed. Routledge Handbook of Sports Performance Analysis. Abingdon: Routledge; 2013: 237-247.

(49) **REIN R, MEMMERT D.** Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. Springerplus. 2016; 5: 1410. doi:10.1186/s40064-016-3108-2

(50) **ROMANILLOS G, ZALTZ AUSTWICK M, ETTEMA D, DE KRUIJF J.** Big Data and Cycling. Transp Rev. 2016; 36: 114-133. doi:10.1080/01441647.2015.1084067

(51) **SINT R, STROKA S, SCHAFFERT S, FERSTL R.** Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis. In: Lange C, Schaffert S, Skaf-Molli H, Völkel M, ed. SemWikis. Publisher: CEUR-WS.org; 2009.

(52) **PRESSER M, SITTO K.** Field Guide to Hadoop, An Introduction to Hadoop, Its Ecosystem, and Aligned Technologies. O'Reilly Media; 2015.

(53) **STAJICH JE, BLOCK D, BOULEZ K, BRENNER SE, CHERVITZ SA, DAGDIGIAN C, FUELLEN G, GILBERT JG, KORF I, LAPP H, LEHVÄSLAIHO H, MATSALLA C, MUNGALL CJ, OSBORNE BI, POCOCK MR, SCHATTNER P, SENGER M, STEIN LD, STUPKA E, WILKINSON MD, BIRNEY E.** The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 2002; 12: 1611-1618. doi:10.1101/gr.361602

(54) **TAKI T, HASEGAWA J.** Visualization of dominant region in team games and its application to teamwork analysis. Comput Graphics Int. Proceedings (SIGGRAPH/Eurographics Workshop Graph Hardw). 2000: 227-235.

(55) **TOGA AW, FOSTER I, KESSELMAN C, MADDURI R, CHARD K, DEUTSCH EW, PRICE ND, GLUSMAN G, HEAVNER BD, DINOV ID, AMES J, VAN HORN J, KRAMER R, HOOD L.** Big biomedical data as the key resource for discovery science. J Am Med Inform Assoc. 2015; 22: 1126-1131.

(56) **WEIMAR D, WICKER P.** Moneyball Revisited: Effort and Team Performance in Professional Soccer. J Sports Econ. 2017; 18: 140-161. doi:10.1177/1527002514561789

(57) **XUE-WEN C, XIAOTONG L.** Big Data Deep Learning: Challenges and Perspectives. IEEE Access. 2014; 2: 514-525. doi:10.1109/ACCESS.2014.2325029

(58) **ZHANG Y, ZHU Q, LIU H.** Next generation informatics for big data in precision medicine era. BioData Min. 2015; 8: 34. doi:10.1186/s13040-015-0064-2