Detecting regularities in soccer dynamics: A T-pattern approach

Valentino Zurloni*, Cesare Cavalera*, Barbara Diana*, Massimiliano Elia* y Gudberg Jonsson**

DETECTING REGULARITIES IN SOCCER DYNAMICS: A T-PATTERN APPROACH

KEYWORDS: Soccer dynamics, Observational methodology, T-pattern detection.

ABSTRACT: The dynamics of play during professional soccer matches are complex phenomena that traditional approaches to the quantification of team sports are not fully able to identify. The aim of this study was to detect such dynamics through an analysis of temporal patterns. In particular, the objective was to reveal the hidden yet stable structures which underlie the interactive situations that determine the attack actions of play in soccer. The methodological approach is based on observational design, supported by digital recordings and computer analysis. Data were analyzed with Theme 6 beta software, which detects the temporal and sequential structure of data sets, revealing repeated patterns that may regularly or irregularly occur within a period of observation. Theme detected many temporal patterns (T-patterns) in the soccer matches analyzed. Striking differences were found when won and lost matches were compared. The number of pattern occurrences and the number of different T-patterns detected was greater for lost matches and lower for the won matches, whereas the number of events coded was similar. Theme software and T-pattern enhance research opportunities by moving further than frequency-based analysis of performance, making this method an effective research and support tool for sports analysis. Our results indicate a need for further investigation upon the possible links between temporal structure detection and human observations in soccer performance. This approach could assist both soccer teams’ staff and coaches in obtaining a greater understanding of game dynamics, providing information that current methods may overlook or not detect at all.

Traditional analysis of performance in team sports has examined behavior through reporting cumulative data on discrete actions, proving to be limited in identifying complex structural regularities (Camerino, Chaverri, Anguera and Jonsson, 2012). A quantitative logic, based on Who [did]- What- Where- When fashion, is unable to consider each player’s behavior within a complex team sport dynamic (Vilar, Araújo, Davids and Bar-Yam, 2013). Team sports performance consists of a multiple series of interrelationships between a vast array of variables, such as the location of the player, the interactions with other teammates and opponents and the influence of the match location (Borrie, Jonsson and Magnusson, 2001; Williams and Jamie, 2009). Successful performance in team sports is achieved through a long-term and methodical training process planned to improve the skills and competence required to meet competitive demands (Garganta, 2009).

Due to this complexity, the simple frequency data risk to provide a superficial and partial view of the observed team’s performance (Borrie and Jones, 1998).

Even in the game of soccer, the most part of the data analyzed have been based on a discreet logic, examining the number of passes in a given area, the number of goal scored home or away, or the number of corners, free or penalty kick scored by the observed team (Camerino et al., 2012). Although this kind of analysis of 1 vs. 1 dynamics has provided, and will continue to provide, valuable information, that coaches use to enhance the coaching process, there is a growing demand for data analysis methods or techniques in this field, that can generate more complete, and therefore more complex, quantitative representations of performance (Jonsson et al., 2006). Key aspects include the detection of regularities that are not revealed through visual inference or traditional methods of data analysis, as well as the lack of standard observation instruments and the need to develop powerful, computerized coding systems (Camerino et al., 2012).

Several studies underlined the importance of tactical aspects to the performance in soccer matches (e.g. Vaeyens, Lenoir, Williams, Matthys and Philippaerts, 2010). As in other team

Correspondencia: Valentino Zurloni. Department of Human Sciences for Education. University of Milan Bicocca. Piazza Ateneo Nuovo, 1. 20126 Milan. Italy. valentino.zurloni@unimib.it

1 Este trabajo forma parte de la investigación Observación de la interacción en deporte y actividad física: Avances técnicos y metodológicos en registros automatizados cuantitativos-cuantitativos, que ha sido subvencionado por la Secretaría de Estado de Investigación, Desarrollo e Innovación del Ministerio de Economía y Competitividad (DEP2012-32124), durante el trienio 2012-2015.

Este trabajo se inscribe en el Grupo de Investigación Consolidado de Cataluña GRUPO DE INVESTIGACIÓN E INNOVACIÓN EN DISEÑOS (GRID). Tecnología y aplicación multimedia y digital a los diseños observacionales, que ha sido subvencionado por el Departamento de Universidades, Investigación y Sociedad de la Información de la Generalitat de Catalunya [2009 SGR 829] durante el período 2009-2013.

* Universidad de Milano-Bicocca.
** Iceland University.

— Artículo invitado con revisión.
sports, the behavior of football players can vary, ranging from simple reactive behaviors to complex reasoning and strategizing (Jonsson, Bjarkadottir, Gislason, Borrie and Magnusson, 2003). For example, traditional frequency analysis of passing would have identified the ball reception and subsequent pass from each zone as discrete events but would not have linked the consecutive actions in a complex pattern.

To sum up, the dynamics of play during soccer matches are complex phenomena that traditional quantitative approaches of team sports are not able to identify (Davids, Araújo and Shuttleworth, 2005). Due to the limits of this discrete logic based on a 1 vs. 1 dynamics, a new data analysis method based on T-pattern detection has been recently introduced (Camerino et al., 2012; Castañer, Torrents, Anguera, Dinușová, and Jonsson, 2009; Fernández, Camerino, Anguera and Jonsson, 2009; Lapresa, Ibáñez, Arana, Garzón and Amatria, 2011).

The T-pattern method has been developed, outside of sport, assuming that complex human behaviors have a temporal structure that cannot be fully detected through traditional observational method or mere quantitative statistical logic (Magnusson, 2000). The T-pattern analysis is able to detect structural analogies across highly different levels of organization and provide an important change from quantitative analysis to structural analysis.

Focusing on soccer performance analysis, T-pattern analysis is able to reveal the hidden yet stable structures which underlie the dynamic situations that determine the actions of play. Therefore, detecting hidden patterns could help the coach to better predict both the performer’s behavior and the opponent’s one thanks to an integrated system that allows for an increased depth of analysis.

In a preliminary investigation on soccer, Jonsson and colleagues (2010) have examined T-pattern in five national and nine international soccer matches, detecting complex patterns of attacking movement made by the observed team. In this study, data showed a correlation between the number of patterns identified in each match and the coaches’ ratings of team performance. Moreover, the level of temporal structure between national and international soccer matches was compared. Data showed a more defined temporal structure in international matches than in national ones, suggesting that international soccer matches are characterized by the presence of a more structured game (Jonsson et al., 2010).

In another study, thirteen national and seven international soccer matches, were coded using T-pattern analysis, confirming that the players’ behavior is more synchronized than the human eye can detect and suggesting that high levels of synchrony are correlated with a good evaluation of performance by professional coaches (Jonsson et al., 2003).

Camerino and colleagues (2012) used T-pattern analysis in order to analyze five National League (Liga) matches and five Champions’ League matches from the 2000-2001 season of FC Barcelona. T-patterns detected revealed regularities in the playing styles of the observed team, including ball possession and ball position patterns during the attacking actions.

The present study applies T-pattern method for detecting the dynamics of attack actions of play in professional soccer. Although soccer is by nature a varied game, some authors (Castellano, Hernández-Mendo, Morales-Sánchez and Anguera, 2007) argued that certain aspects tend to follow one another during the game. Since there is a greater degree of predictability when referring to the same team, our first aim is to reveal the hidden yet stable structures which underlie the dynamic situations (Shepherd, Lee and Kerr, 2006) that determine the attack actions of play in different soccer matches of the same team. Secondly, to understand whether and how the temporal and sequential organization of attack actions can affect the outcome of the performance, we compared the T-patterns of attack actions in the won matches and in the lost ones of the same team.

Method

The present study exploits the observational methodology (Anguera, Blanco-Villaseñor, Hernández-Mendo and Losada, 2011; Anguera, Blanco-Villaseñor and Losada, 2001), enhanced by the use of new technology. The observational methodology became more and more widespread within sport research, because of its high flexibility and adaptability (Anguera, 2009).

As the game of soccer has evolved, so have the methods of analyzing performance developed from the simple use of hand notation tracking of players’ movements on scale plans of pitches (Bloomfield, Jonsson, Polman, Houlahan and O’Donoghue, 2005) to the current utilization of digital video recordings and computerized analyses (Borrie, Jonsson, and Magnusson, 2002). Digital recordings and computer analysis (e.g., Borrie, Jonsson and Magnusson, 2002) have been widely used in sports research (e.g., Luo, Wu and Hwang, 2003) and specifically in soccer (e.g., Jonsson et al., 2006), because of the great advantages brought in terms of recording quality, measurement of time and capture of co-occurrences or diachrony.

Design

As proposed by Anguera and colleagues (2001; 2011), we consider three main criteria to give a taxonomic definition of the observational design, as applied to our study. The observational design was nomothetic (as opposed to idiographic, which refers to the number of subjects observed), since we observed different matches; punctual (as opposed to continuous, referring to the number of observations conducted on the same subject); multidimensional (as opposed to unidimensional, referring to the number of levels of response observed).

The adoption of this N/P/M (nomothetic, punctual, multidimensional) design led to a series of decisions being made regarding the structure of the observation instrument, the type of data, data quality control and data analysis.

Participants

This study is part of a broader research project involving the analysis of all games played by a top club during the first leg (19 matches) of the Italian National League Championship (Serie A) over the 2012-2013 season (Anguera, Zurloni, Jonsson, Diana and Elia, 2013). For this study, we chose to take in consideration only won and lost matches, discarding tied games; then, we balanced our datasets for the home/away condition (this fixed criteria was controlled but not considered for the purposes of this study) and for the won/lost one, obtaining 12 games (3 matches per condition were randomly extracted from each of the 4 groups).
Procedure

Recording instrument. We used Lince behavior coder (Gabin, Camerino, Anguera and Castañer, 2012) to analyze and code the events of each match taken in consideration.

Lince software is consistent with the proposed observational design, since it is multidimensional in nature and structured around fixed, mixed and changing criteria (Castellano, Perea and Hernández-Mendo, 2008). Moreover, Lince can calculate Cohen's kappa coefficient for all or some of the criteria by comparing two registered data files.

Observation instrument. The observation instrument combines field format and category systems (Anguera, Magnusson and Jonsson, 2007). The fixed criteria are entered at the beginning of the match, while the mixed criteria apply every time there is a change in the score, number of players and between the first and the second half of the match. The changing criteria are coded throughout the whole match. Each one of these criteria gives rise to respective category systems that fulfill the conditions of exhaustiveness and mutual exclusivity (E/ME).

In this study, we decided to focus on attack actions mainly because of shooting restrictions. Since our video data were obtained from the TV recordings of the matches, the focus of the video was always the active play, in particular the player in ball possession. Since defense tactics comprehend a certain amount of moving and changing positions without ball possession, they were only partially observable using a public TV shooting.

The dimensions considered in the present observation instrument correspond to the following criteria: lateral position, zone (see Figure 1), lateral passing, zone passing, recovery and loss, ball out of play (see Table 1).

<table>
<thead>
<tr>
<th>CRITERIA</th>
<th>CATEGORIES</th>
<th>CODE</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>LATERAL POSITION</td>
<td>Right</td>
<td>RI</td>
<td>Lateral position of play</td>
</tr>
<tr>
<td></td>
<td>Center</td>
<td>CE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>LE</td>
<td></td>
</tr>
<tr>
<td>ZONE</td>
<td>Ultra-defensive</td>
<td>UD</td>
<td>Zone of play, defined according to five pitch areas or spatial strips</td>
</tr>
<tr>
<td></td>
<td>Defensive</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Offensive</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ultra-offensive</td>
<td>UO</td>
<td></td>
</tr>
<tr>
<td>LATERAL PASSING</td>
<td>Right Passing</td>
<td>RP</td>
<td>Lateral direction of pass</td>
</tr>
<tr>
<td></td>
<td>Center Passing</td>
<td>CP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left Passing</td>
<td>LP</td>
<td></td>
</tr>
<tr>
<td>ZONE PASSING</td>
<td>Ultra-defensive Passing</td>
<td>UDP</td>
<td>Zone direction of pass</td>
</tr>
<tr>
<td></td>
<td>Defensive Passing</td>
<td>DP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central Passing</td>
<td>CP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Offensive Passing</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ultra-offensive Passing</td>
<td>UOP</td>
<td></td>
</tr>
<tr>
<td>RECOVERY &amp; LOSS</td>
<td>Momentary Loss</td>
<td>ML</td>
<td>Ball is lost temporarily, the attack action continues. A ball is temporarily lost when there is a Recovery in the UO zone, or in the O zone after no more than 2 consecutive passes of the opposing team during loss time.</td>
</tr>
<tr>
<td></td>
<td>Permanent Loss</td>
<td>PL</td>
<td>Ball is definitely lost, the attack action is over. It is considered permanent loss after 3 consecutive passes from the opposing team or if the ball is recovered in C, D, or UD zone.</td>
</tr>
<tr>
<td></td>
<td>Recovery</td>
<td>R</td>
<td>Ball is recovered in UO zone, or in O zone no more than after 2 consecutive passes from the opposing team during loss time.</td>
</tr>
<tr>
<td></td>
<td>Keeps Ball</td>
<td>KB</td>
<td>The player brings the ball from a zone to another or from a lateral position to another (or both at the same time). It can also happen if the player launches the ball forward and he himself gets to the ball in the next zone. Anything that is not attack action.</td>
</tr>
<tr>
<td></td>
<td>Passive Ball</td>
<td>PB</td>
<td></td>
</tr>
<tr>
<td>BALL OUT OF PLAY</td>
<td>Corner Kick</td>
<td>CK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Throw In</td>
<td>TI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Goal Kick</td>
<td>GK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dropped Ball</td>
<td>DB</td>
<td>When the referee interrupts the game for reasons not due to the other conditions of BALL OUT OF PLAY (for example, medical staff is entering the area because a player is down but there is no foul).</td>
</tr>
<tr>
<td></td>
<td>Goal</td>
<td>G</td>
<td>Any shot on target, or on target but saved by the keeper or a defender.</td>
</tr>
<tr>
<td></td>
<td>Not Goal</td>
<td>NG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Penalty Kick</td>
<td>PK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Free Kick</td>
<td>FK</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Observation Instrument.
We defined an attack action as an action that brings the ball in the ultra-offensive zone and can end with: Goal (G), Not Goal (NG), Permanent Loss (PL), including penalties, corner kicks, free kicks in ultra-offensive zone and throw in in ultra-offensive zone.

Each attack action was coded starting from the first pass that crosses the offensive line. The same happens in case the ball gets to the ultra-offensive zone thanks to a free kick or a throw in conceded to the observed team.

Analysis instrument. Datasets were analyzed with Theme 6 beta. This software detects the temporal and sequential structure of data sets, revealing repeated patterns that may regularly or irregularly occur within a period of observation (T-patterns). A T-pattern is essentially a combination of events where the events occur in the same order with the consecutive time distances between consecutive pattern components remaining relatively invariant (Magnusson, 2005).

### Analysis and results

#### Data quality

Two observers used the Lince software to code all the games selected. The same software calculated Cohen’s kappa coefficient for all the criteria by comparing the two registered data files. The values ranged between .75 and .85, which provides a satisfactory guarantee of data quality. However, when particular disagreements were identified, the specific cases were discussed and agreed on by the two coders.

#### Common T-patterns in all matches

The individual datasets of each match were combined in Theme using a function called “concatenate into a multi-sample file”. This function creates a single file containing all the single-sample files in the project. Thus, the resulting file contains more than one sample and is called a multi-sample file.

Theme detected 167 different types of patterns ($p < .0001$; minimum occurrence = 7) that occurred in over 80% of all the matches analyzed (mean length = 2.45; mean level = 1.43); overall, most T-patterns show a succession of passing, momentary or permanent loss and recovery. The most frequent pattern (see Figure 2), detected 24 times in all the analyzed matches, describes a succession of a momentary loss in ultra-offensive central area, a corner kick from the right side, passing the ball from the ultra-offensive right area to the ultra-offensive central area, ending with a momentary loss in the ultra-offensive central area.

No T-patterns were detected containing successful attack actions (goal or not goal) among all the matches analyzed.

Comparison between won and lost matches

A second analysis was performed ($p < .001$; minimum occurrence = 3) in order to identify the patterns that allow us to distinguish the organization of play between won and lost matches. We used the function in Theme called “selection–multi-sample file selection– statistical”, where the software selects patterns that appear significantly more often in samples selected than in the multi-sample file as a whole.

Won matches presented only 9 common T-patterns ($p < .05$). The most complex T-pattern (see Figure 3) was detected 8 times in won matches whilst only 2 times in lost matches. It describes a succession of passing strategies (from the central area to the offensive right area, then from the ultra-offensive right to the ultra-offensive central area), followed by a momentary loss in the ultra-offensive central area, another pass from the ultra-offensive right to the ultra-offensive central area, ending with a permanent loss in the ultra-offensive central area.

---


Figure 2. Succession of a momentary loss in ultra-offensive central area, a corner kick from the right side, passing the ball from the ultra-offensive right area to the ultra-offensive central area, ending with a momentary loss in the ultra-offensive central area.

Figure 3. The most complex T-pattern detected in won matches.
Theme detected 101 common T-patterns ($p < .05$) in lost matches. The most complex T-pattern (see Figure 4) was detected 7 times in lost matches whilst only one time in won matches. It describes a complex and articulated succession of passing strategies (in particular in central offensive area), a recovery in offensive central area, ending with a permanent loss in ultra-offensive central area.
Discussion and conclusions

Data from the first analysis have shown a remarkable number of common T-patterns in all matches. However, there were no event-types (goal and not goal) related to successful attack action among the detected T-patterns. This suggests that the game strategies that are repeated in all matches (regardless of factors such as the result of the match) do not lead to effectively realize the amount of effort created.

Comparing T-patterns’ structure and distribution between won and lost matches, striking differences were found. The total number of pattern occurrences and the number of different T-patterns detected was greater in lost matches than in won ones, whereas the number of events coded per game was similar. This suggests that, in a winning situation, the team is more likely to continue using the same strategy. Specifically, T-patterns of the won matches show that the action was frequently pushed forward (from central to offensive zone, from offensive to ultra-offensive zone). Furthermore, the team capitalizes the right flank to reach the central ultra-offensive zone. Conversely, in losing situations, the team tries different strategies to gain control of the match, often with a high number of passes in the offensive central zone, which result unproductive.

The results show that T-pattern analysis is an effective tool that supports research in sport performance analysis. Our preliminary data need to be deepened and confirmed by other analyses: a) through observation and comparison between all the matches of the first and second leg (the next 19 games of the championship); b) considering factors other than the result of the matches, such as the field factor (comparison between home and away matches), the precise score of the matches (analyzing the different strategies as the score changes); c) extending the observation not only to attack actions, but also to teams’ defensive behavior. However, the third point requires having ad hoc shootings of the games more than the TV ones, which focus on the ball possession and not on the whole team’s movements.

Nevertheless, our results highlight the potential for T-pattern analysis to make a significant contribution to the study of soccer dynamics. This type of analysis could provide the coach and the soccer teams’ staff with useful information that current analytical methods may not detect or overlook. This could help increasing the team’s performance at different levels.

This results also point towards the need to investigate the potential link between temporal structure detection and coach observations in soccer performance. This approach could assist in obtaining a greater understanding of coach knowledge construction.
Detecting regularities in soccer dynamics: A T-pattern approach

References


