Introduction

Football is a challenging research domain. Each match involves 22 players demonstrating collaborative behaviour that requires specific roles in an adversarial, uncertain, and dynamic environment. The behaviour of players and the decision making processes can range from the most simple reactive behaviours, such as running towards the ball, to complex reasoning that take into account the behaviour and perceived strategies of team-mates and opponents (Jonsson, 1998).

In the pursuit of generating quantitative information on performance sport researchers have traditionally used frequency of event occurrence as their index of performance e.g. the analyst has recorded how many passes have been made from particular playing zones or how many times possession has been lost (Jonsson et al., 2000). In essence the analyst has been answering the question "how many times did 'x' occur?". However frequency of event occurrence has been shown to be an inadequate index of performance that cannot differentiate between effective performances (Borrie and Jones, 1998). If one accepts the argument that sport performance consists of a complex series of interrelationships between a vast array of performance variables then simple frequency data can only ever provide a relatively superficial view of performance.

If performance analysis is to continue to advance understanding of sports performance then it must find better methods of collecting and analyzing match analysis data. The purpose of this paper is to introduce and explain a new data analysis method that has the potential to make a significant contribution to analyses of sports performance. Data from preliminary studies of football performance are also presented to show the potential outcome from the analysis process.

T-pattern detection and analysis

The analysis approach presented is based on a process known as T-pattern detection which allows the detection of the temporal and sequential structure of a data set. The method has been developed, outside of sport, on the assumption that complex streams of human behaviour have a temporal/sequential structure than cannot be fully detected through unaided observation or with the help of standard statistical and behaviour analysis methods. Given that observational records of human behaviour, including sport performance analysis, have both a temporal and sequential structure an analysis tool that can describe this structure will enhance understanding of the behaviour (s) being studied. A generic observational software package called Theme has been specifically developed to operationalise T-pattern detection as an analysis process (Magnusson, 1996, 2000).
A schematic representation of a T-pattern is shown in figure 1. If one assumes that the letters in line 1 correspond to specific performance events (e.g. pass, tackle and shot in football) that appear on the line in proportion to the time of their occurrence then line 1 is a visual representation of the temporal structure of a sports performance.

Within the upper line there are four events (a, b, c, d) that occur in a regular temporal pattern however the pattern has been masked by the surrounding, more random, occurrence of the events w and k. If a performance analyst or coach were simply visually inspecting the data string it is unlikely that the pattern would have been detected. The T-pattern analysis would have identified the pattern because of its consistent temporal structure. The T-pattern detection algorithms allow an analyst to separate out randomly occurring events from temporal patterns even when the random events occur in between elements of the pattern.

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 with respect to an expectation assuming, as a null hypothesis, that each component is independently and randomly distributed over time. As stated by Magnusson 'that is, if A is an earlier and B a later component of the same recurring T-pattern then after an occurrence of A at t, there is an interval that tends to contain at least one occurrence of B more often than would be expected by chance' (Magnusson, 2000, p. 94). The temporal relationship between A and B is defined as a critical interval and this concept lies at the centre of the pattern detection algorithms.

The pattern detection algorithms can analyze both ordinal and temporal data however, for the algorithms to generate the most meaningful analyses the raw data must be time coded i.e. an event must be coded according to time of occurrence as well as event type. The coding of many event-types and corresponding times results in the type of data set shown in figure 2. This figure displays a behaviour record from the second half of a club football match and consists of 250 series of occurrence times (one for each coded event type) ordered according to their first occurrence time.
Detection of real-time patterns in sports interactions in football

Figure 2:
A time series behaviour record from the second half of a football match from the European Champions League 1997. The match was coded from a digitized video recording of approximately 45 minutes duration (time is in seconds).

Only limited aspects of T-pattern detection has been presented here to give some insight in to the theoretical base of the process. A complete explanation of the theoretical roots of the pattern-detection algorithms together with an overview of the wider use of the process has been presented elsewhere (Magnusson, 1996, 2000).

Method
A team sport, football, has been analyzed with the intention of identifying whether T-pattern detection has relevance as an analytical method within performance analysis. The research utilized multiple game analysis with each game being treated as a single case. Twenty football matches, thirteen club and seven international matches, were coded using a combination of the football match analysis system developed at Liverpool John Moores University and ThemeCoder, enabling detailed coding of digitized video files (25 frames per sec.). Coding included data on pitch position, player and match events. Pitch position was classified according to the pitch divisions shown in figure 3. The primary event categories for data collection were: pass; tackle; header; run; dribble; clearance; shot; cross; set-play; lost control; foul. Additional qualifying statements could be tagged to each event category. All data was analyzed using the Theme software package.

Figure 3:
A schematic representation of the zones identified for analysis of ball movement within football.

Table:
<table>
<thead>
<tr>
<th>1</th>
<th>4</th>
<th>7</th>
<th>10</th>
<th>13</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>15</td>
<td>18</td>
</tr>
</tbody>
</table>

Direction of play
Results and discussion

The data show that a high number of temporal patterns exist in football. The number, frequency and complexity of the detected patterns, indicates that sport behaviour is very structured. This synchrony was found to exist on different levels, with highly complex time structures that extended over considerable time spans within performances with patterns occurring in both cyclical and acyclical fashion.

The data taken from football show three discrete examples that identify within-team patterns (e.g. figure 4, ball movement) and interactive patterns involving both teams (figures 5 and 6, goal scoring).

Figures 4a et 4b:

A temporal pattern relating to attacking movement of the ball through the centre of the pitch.

Exemple d’un T-pattern relatif à un mouvement d’attaque depuis le centre du terrain.

a. Data output from Theme analysis software showing temporal and hierarchical representation of a T-pattern. The three boxes in this figure involve the same observation period and the T-pattern relates to ball movement in the centre of the pitch within a football match.

b. Schematic representation of the same data.

1. Player A receives the ball in Zone 8, passes the ball to a team mate and runs forward.
2. Player A receives the ball in Zone 11, passes the ball to a team mate and runs forward.
3. Player A receives the ball in Zone 14, passes the ball to a team mate in Zone 15 (4).

a. Données obtenues avec le logiciel Theme montrant une représentation temporelle et hiérarchique d’un T-pattern. Les 3 cadres de la figure concernent la même période d’observation et le T-pattern correspond au déplacement du ballon sur le terrain.

b. Représentation schématique des mêmes données. 1. Le joueur A reçoit le ballon en zone 8, passe le ballon à un équipier et court. 2. Le joueur A reçoit le ballon en zone 11, passe le ballon à un équipier et court. 3. Le joueur A reçoit le ballon en zone 14, passe le ballon à un équipier dans la zone 15 (4).
A typical within-team event pattern from the football analysis is shown in figures 4a and 4b. This figure displays a detected T-pattern that occurred three times during the first half of a European Championship qualifying match (1998). The three boxes in figure 4a show the same observation period. The upper-left box shows the hierarchical construction of the pattern. The tree structure identifies the simple patterns on its right hand edge and, as the tree builds towards the left edge of the box, shows how the simple patterns are linked together to form the more complex pattern. The upper-right box displays the time point of each event-type in the pattern and their pattern connection based on the critical interval relationship between their occurrence series. The bottom box shows the pattern, as a hierarchical structure, expressed in relation to the observation period i.e. when it occurred during the match (only complete patterns are shown in this box).

The pattern describes how player A moves the ball towards the opponents goal by receiving the ball in, and then passing it out of, pitch zones 8, 11 and then 14 consecutively. Player A then completes the sequence by passing it on to player B who receives it in zone 15. The pattern describes an attacking movement through the middle of the pitch, which opponents would clearly wish to prevent. 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 the four zones. The movement from zone 11 to 14 also occurred on another five occasions during the first half (figure 4a, upper right box) further suggesting that player A was working effectively through the central channel of the pitch. This integrated form of analysis would potentially enhance the information given to the coach.

Figure 5:

1. BCL - Rivaldo passes the ball; 2. PSV - Faber intercepts the ball; 3. BCL - Sergi passes the ball; 4. PSV - Jonk makes a bad pass; 5. BCL - Dugarry successful individual act; 6. BCL - Dugarry attempts a shoot, saved; 7. BCL - Sergi intercepts the ball; 8. PSV - Jonk takes a free-kick; 9. BCL - Luis Enrique scores a goal (two occurrences, p < .005).

Figures 5 and 6 show examples of detected T-patterns from European Cup match between PSV Eindhoven and Barcelona, which finished with a 2-2 draw. An analysis of the match revealed patterns including all four goals that were scored. Barcelona goals in figure 5 and PSV goals in figure 6. The data shows the capacity of the analysis process to identify longer, more complex
patterns that involve extensive interactions between opposing teams. There is a clear consistent temporal pattern preceding all goals scored in this match. The patterns cover an extended period of time within which the two teams exchange possession on at least two occasions. The length of the time periods between the events forming the pattern was such that other match events will have occurred between pattern events. The length and nature of the patterns is such that one could question whether, given the extended time period of the patterns and the changes of possession, the patterns events were causally related. However the consistency in the temporal pattern preceding each goal is quite clear. In this case the information provided by the T-pattern analysis raises questions about the relationship between events that are spread over an extended time period. This doesn’t provide a coach with clear and simple answers about the nature of the goals scored but without the analysis the potential significance of the relationships between events would not have been considered. This analysis may therefore prompt a coach to revisit the video footage of passages of play to identify causally linked elements of performance that might have been missed had the temporal pattern not been identified.

Figure 6:

### Pattern 3:
1. PSV - Petrovic passes the ball;
2. BCL - Hesp, keeper throws in the ball;
3. PSV - Vampeta passes the ball;
4. BCL - Luis Enrique makes a bad pass;
5. BCL - Dugarry attempts a shoot, saved;
6. PSV - scores a goal (two occurrences, p<.005).

In addition to immediate analysis of individual matches the data were also used to look at two additional issues relating to structure within team performance. The first issue investigated related to the potential interrelationship between performance rating by coaches and the degree of structure in team performance. Three experienced football coaches and five amateurs observed several club and international matches and were asked to rate the performance of every player (on both teams) on a simple ten point Likert type scale. For each coach the player ratings for a specific team were averaged to give a team performance rating. Team performance ratings were then correlated (Pearson product-moment correlation) against the number of patterns exhibited by each team. The data (cf. figure 7) show that the coaches’ ratings of team performance were significantly correlated to the number of patterns identified for each team ($r=0.81$, $p<0.05$). Lower correlation was found between the amateurs ratings of team performance and the number of patterns identified for each team ($r=0.53$, $p<0.05$).
The link between performance rating and pattern participation suggests that coaches were recognizing, albeit at a potentially subconscious level, the structure within a team’s play. However the traditional rationale for performance analysis is that coaches cannot observe and remember discrete events within critical event sequences (Franks and Miller, 1986). Yet, in this sample, the fact that coach performance ratings were correlated with pattern participation suggests that coaches were perceiving information about the interrelationships between events. This finding also warrants further investigation since it relates to such a fundamental foundation in the performance analysis literature.

Figure 7:
Correlation between team pattern participation and coaches and novice subjective ratings (p>0.05).

![Correlation between team pattern participation and coaches and novice subjective ratings](image)

The second issue considered was the comparative level of temporal structure within club and international football matches. In a simple data manipulation three randomly selected club and three international matches were compared in terms of the mean number of patterns and the mean number of pattern occurrences identified in each match type. The data (cf. figure 8) show that international football has a more defined temporal structure than club football. This finding may be due to the presence of higher technical abilities in international footballers which help create a more structured game or, alternatively, contextual differences between club and international football e.g. club football is played at a higher pace throughout mitigating against the development of structure within the game. Whatever the reason the clear difference in temporal structure between club and international football merits further investigation.

Figure 8:
Number of events, different pattern types and pattern occurrences between national team matches and club matches. All differences are significant at p>0.001.

![Number of events, different pattern types and pattern occurrences](image)
Discussion

The number, frequency and complexity of detected patterns indicates that the behaviour of football players is more synchronized than the human eye can detect. This synchrony was found to exist on different levels, with highly complex time structures that extended over considerable time spans, often in a cyclical fashion, as well as less complex patterns with a shorter time span. Synchrony of this kind was found to correlate highly with assessment of performance. A stronger correlation was also discovered between performance assessments of professional coaches and team pattern participation than between assessments of amateurs and team participation in patterns. National matches were found to be more structured than club matches.

The results show that pattern analysis can be used to track elements in the game such as timing of events, the passing of the ball, team structure etc. in a novel way, indicating that pattern analysis is useful in enhancing existing methods used in football analysis. New kinds of profiles, for both individuals and teams, can be discovered using the detected behavioural patterns in combination with elementary statistics. Moreover, some answers are already suggested to questions, such as: Are there certain patterns that are related to doing well or bad? What responses seem to be evoked by certain actions or sequences of actions? Coaches could use this kind of structural information when selecting players or when searching for the opponent’s "weak spots".

The preliminary data highlights the potential for T-pattern analysis to make a significant contribution to sport performance analysis. Current analytical methods that focus on simple frequency analysis cannot identify the temporal patterns within a sports performance. Consequently without this form of analysis meaningful information is not being made available to the coach. If this information is not available then it possible that performance is not being optimized.

The data also point towards the need to investigate the potential link between temporal structure in sport performance and the understanding of performance being generated by coach observations. The data suggest that whilst coaches may not be able to accurately recall discrete events they do perceive inter-relationships between events. This analysis approach can assist in generating a greater understanding of coach knowledge construction.

References


