University of Cape Town

STA5093W: Data Science Minor Dissertation

Research Proposal

Quantifying Pressing Effectiveness and Its Impact on Formations in Football

Student:
Kenneth Ssekimpi

Supervisor:
Neil Watson

March 17, 2024
Student Number: SSKKEN001

Contents

1   Title

Quantifying Pressing Effectiveness and Its Impact on Formations

2   Introduction/Background Information

The prominence of data science has increased in association football (or soccer), with the volume of its analytics output fast approaching other sporting codes such as baseball, cricket, and basketball (Decroos, 2020). Football's suitability to analytics has given rise to numerous use cases including the analysis of player performance (for example, passing and positioning) and the strategic decision-making process employed by coaches (Cotton & Miller, 2022; Goes et al., 2021). One fundamental tactic analysed is pressing, a defensive strategy where the team without possession aggressively attempts to regain the ball. It disrupts the opponent's build-up play, forces errors, and facilitates quick transitions to offense, making it a cornerstone of modern football (Chambers, 2022; Morgan, 2018). Pressing strategies involve the application of different formations to optimise defensive effectiveness. For instance, a high pressing strategy might be more suitable for a 4-3-3 formation, whereas zonal pressing fits better with a 4-4-2 setup (The Coaches' Voice, 2023). Formations refer to the arrangement of players on the pitch, influencing both attacking and defensive strategies and manoeuvres (Bauer et al., 2023).

Machine learning techniques have increasingly been used to understand the pressing strategies implemented by successful teams and the potential impact of different formations on the effectiveness of pressing (Bauer & Anzer, 2021; Rico-González et al., 2023). Despite the development of metrics like Passes Allowed Per Defensive Action (PPDA) or Defensive Action Expected Threat (DAxT) to measure aspects of pressing effectiveness (or more generally, the effectiveness of defensive contributions) (Merhej et al., 2021; Trainor, 2014), there is a gap in research regarding the influence of formations on pressing throughout the pitch.

3   Literature Review

3.1   Types of Presses

There are several types of presses used in different situations, often depending on the team's overall tactics and the location of the ball on the pitch (Chambers, 2022). Some common types include:

- **High Press**: The team applies pressure high up on the pitch in the opponent's attacking or midfield third. This is a high-risk, high-reward tactic that can lead to turnovers in dangerous areas but leaves the defence vulnerable if bypassed (Chambers, 2022; Low et al., 2018).

- **Midfield Press**: Pressing occurs around the centre of the pitch, allowing the opponent some space to build up play, but restricting their options and forcing them backwards (Chambers, 2022).

- **Low Block**: The team defends deep in their own half, staying compact and forcing the opponent to break them down (Chambers, 2022; Low et al., 2018).

- **Counter-Press** (or *gegenpress*): The team aims to win the ball back immediately after losing possession, typically in the opponent's attacking half (like the high press). This high-intensity tactic relies on quick transitions and is effective against unprepared teams. However, the counter-press needs to be used strategically to avoid the team losing its shape (Chambers, 2022).

Effective pressing requires co-ordinated pressing triggers (specific actions or cues that signal players to initiate pressing), player co-operation, and individual pressing skills (Modric et al., 2023).

3.2   Formation Analysis and Gameplay Influence

Different types of formations can be employed to optimise various aspects of gameplay, such as:

- **Attacking Play**: Formations impact how teams build attacks with some emphasizing width and others focusing on central control (Bauer et al., 2023).

- **Defensive Play**: Formations dictate defensive shape and pressing strategies, with some formations favouring high pressing fullbacks and others prioritizing defensive solidity (Bauer et al., 2023).

- **Player Roles**: Formations define player responsibilities. Wingers in a 4-2-3-1 focus on attacking the flanks, while a lone striker needs good hold-up play (Bauer et al., 2023).

- **Flexibility**: Modern teams often adapt formations dynamically during the game based on the opponent and situation (Bauer et al., 2023).

3.3   Data Analysis in Football Performance

Data analysis is increasingly influencing football performance analysis, offering new tools and methods for evaluating tactics and player performance (Link, 2018). Pressing and formation analysis are two such nascent applications that have received more attention in recent years. There now exists ample data for:

- **Pressing Analysis**: StatsBomb is a football database and analytics website that aims to provide comprehensive statistics and data on various aspects of football. Their data allows for quantifying pressing intensity, success rates, and player contributions (Morgan, 2018).

- **Formation Analysis**: The tracking of player movements within formations, analysing passing patterns specific to formations and assessing their effectiveness against different opponents. For example, data analysis can reveal if a team's wingers in a 4-3-3 are providing enough width and crosses in attack (Goes et al., 2021).

3.4   Gap in Existing Research: Pressing and Formations

There's a gap in research regarding a **quantitative analysis** of how formations directly influence pressing effectiveness. Current research focuses on analysing pressing or formation analysis separately, with little attention paid to the influence of one on the other (Forcher et al., 2022).

3.5   Proposal Contribution

My research aims to bridge this gap by:

- Analysing pressing metrics (intensity, success rates) for different formations (e.g., high press effectiveness in 4-3-3 vs. 4-4-2).

- Identifying formations that might be better suited for specific pressing strategies (e.g. formations that facilitate quicker counter-pressing).

By addressing this research gap, my work can provide valuable insights for coaches and analysts to optimise pressing tactics based on team formations.

4   Research Objectives

4.1   Research Questions

- How do formations influence pressing success rates?

- Are there specific game situations (e.g., trailing in the second half) where formations influence pressing effectiveness?

- Which player positions within a formation (e.g., defensive midfielders vs. wingers) are most crucial for successful pressing?

4.2   Significance

This research will provide valuable insights for coaches and analysts by:

- Highlighting formations that optimise pressing effectiveness, which could be used to guide player recruitment and training strategies that align with the team's pressing approach.

- Identifying game situations where specific formations should be used to maximize pressing success.

By delving into the interplay between formations and pressing, this research can empower teams to gain a tactical advantage on the pitch.

4.3   Justification

The rationale for this research is supported by studies like (Modric et al., 2023), which emphasize the importance of pressing for success. Additionally, studies such as those from (Rico-González et al., 2023) and (Goes et al., 2021) highlight the gap in research on formations' impact on pressing effectiveness.

This research offers a novel approach to analysing pressing tactics by considering the influence of formations. It has the potential to impact football strategy and performance optimisation.

5   Methodology

5.1   Data Source

The StatsBomb Open Data repository facilitates football research by making some of its proprietary data publicly available its *StatsBombR* R package.

Additionally, the *worldfootballR* package enables users to extract relevant player metrics and match data from renowned football data platforms, such as *FBref* and *Understat*, which both provide granular player- and team-level data. This ensures a comprehensive coverage of the necessary data points.

It should be noted, however, that using free data sources might require additional cleaning and preprocessing steps before analysis. This is especially true if combining data from different data sources.

5.2   Data Description

The following types of data, which are confirmed to be available from the sources listed in Section 5.1, would aid in answering the research questions stipulated in Section 4:

- Event Data:
  - Tackles (successful and unsuccessful)
  - Interceptions
  - Pressures (number of times a player is closely guarded by the opponent)
  - Recoveries (winning the ball back after losing possession)
- Formation Data:
  - Team formation (e.g., 4-4-2, 3-5-2)
  - Player positions within the formation (e.g., central midfielder, left winger)
- Match Context Data:
  - Game state (winning, losing, tied)
  - Match period (first half, second half)

– Scoreline

5.3   Data Analysis Methods

This research aims to bridge the gap by analysing how different formations influence pressing effectiveness and how pressing effectiveness varies across these formations. Various techniques will be used, such as:

- **Classification**: Categorise pressing events as successful (ball recovery) or unsuccessful based on features like tackle success rate, location of the press, and players involved (Forcher et al., 2022).

- **Regression**: Analyse how pressing effectiveness (e.g., percentage of successful pressures) relates to different formations. This will involve modelling the relationship between pressing metrics, formation types, and outcome variables like goals scored or possession won.

- **Clustering**: Group formations based on player positioning and tactics to identify clusters with inherently different pressing styles, allowing for the assessment of the effectiveness of each style (Merhej et al., 2021).

6   Ethical Considerations

6.1   Data Collection

Data providers such as StatsBomb collect statistical data relating to players for use in their own services (*StatsBomb Privacy Policy*, 2022). Identifiable player information, such as physical attributes, is collected from teams, broadcast feeds, and public sources (*StatsBomb Privacy Policy*, 2022). No health or confidential data about players is processed. Personal player data that pertains to in-game situations, such as substitutions due to injury, is included; however, specific details about the injuries are not disclosed (*StatsBomb Privacy Policy*, 2022).

6.2   Data Use and Regulations

To safeguard player identities, the data will be anonymised, and player names will be omitted from the datasets. While it is unlikely that individuals' consent was sought during data collection, professional sports players expect their in-game actions to be analysed, as this data is available to anyone watching a game. Given that millions of people watch these matches, it is reasonable to assume that the actions made by players in games are not confidential information. None of the methodology employed in or the results of this study will discriminate against any individual player or groups/teams of players.

Strict adherence to the data providers' terms of use regarding data access and usage will be observed. Sharing raw data or unauthorised usage will be avoided.

6.3   Additional Considerations

- Data collection methods can introduce bias, such as in the case of focusing on certain leagues.

- Transparency around data collection, anonymisation techniques, and intended uses will be observed.

- This study advocates for responsible data usage in football that benefits the football ecosystem (i.e., players and coaches).

7   Timelines (work plan) and Budget

Phase	Description	Due Date	Key Milestones
Ethics Approval	Gain ethics approval	14 June 2024	Prepare/submit ethics application, address feedback
Literature Review	Analyze existing research	30 June 2024	Complete bibliographic search, identify relevant articles, summarize key findings
Data Collection and Cleaning	Collect, clean and pre-process data	31 July 2024	Download R packages, write scripts to clean and organize data. Document steps
Data Analysis	Analyze data (quantitative methods)	30 September 2024	Explore data, develop analysis plan (statistical tests), perform data analysis, interpret results, visualize findings
Report Writing	Write comprehensive research report	14 November 2024	Structure report, explain research question/s, methodology, data analysis procedures, present results, discuss implications, conclusion
Thesis Submission	Hand in final dissertation	14 December 2024	Adhere to formatting convention/s, proofread and edit work, meet all deadlines, prepare for possible thesis defence

Table 1: Research Timelines

Budget: None

References

Bauer, P., & Anzer, G. (2021). Data-driven detection of counterpressing in professional football: A supervised machine learning task based on synchronized positional and event data with expert-based feature extraction. *Data Mining and Knowledge Discovery*, *35*(5). https://doi.org/10.1007/s10618-021-00763-7

Bauer, P., Anzer, G., & Shaw, L. (2023). Putting team formations in association football into context. *Journal of Sports Analytics*, *9*(1). https://doi.org/10.3233/jsa-220620

Chambers, T. (2022, November 6). *Different Types Of Presses In Football: 4 Variations*. Soccer Source Coaching. https://soccersourcecoaching.com/different-types-of-presses-in-football/

Cotton, R., & Miller, I. L. (2022, November). *How Data Science is Changing Soccer*. Datacamp. https://www.datacamp.com/blog/how-data-science-is-changing-soccer

Decroos, T. (2020). *Soccer Analytics Meets Artificial Intelligence: Learning Value and Style from Soccer Event Stream Data*.

Forcher, L., Altmann, S., Forcher, L., Jekauc, D., & Kempe, M. (2022). The use of player tracking data to analyze defensive play in professional soccer - A scoping review. In *International Journal of Sports Science and Coaching* (Vol. 17, Issue 6). https://doi.org/10.1177/17479541221075734

Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., Elferink-Gemser, M. T., Knobbe, A. J., Cunha, S. A., Torres, R. S., & Lemmink, K. A. P. M. (2021). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. In *European Journal of Sport Science* (Vol. 21, Issue 4). https://doi.org/10.1080/17461391.2020.1747552

Link, D. (2018). Data Analytics in Professional Soccer. *Data Analytics in Professional Soccer*. https://doi.org/10.1007/978-3-658-21177-6

Low, B., Boas, G. V., Meyer, L., Lizaso, E., Hoitz, F., Leite, N., & Gonçalves, B. (2018). Exploring the effects of deep-defending vs high-press on footballers' tactical behaviour, physical and physiological performance: A pilot study. *Motriz. Revista de Educacao Fisica*, *24*(2). https://doi.org/10.1590/S1980-6574201800020009

Merhej, C., Beal, R. J., Matthews, T., & Ramchurn, S. (2021). What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3447548.3467090

Modric, T., Gabrilo, G., & Sekulic, D. (2023). INFLUENCE OF TEAM PRESSING ON MATCH PERFORMANCE IN HIGHEST-LEVEL SOCCER; PRELIMINARY REPORT. *Kinesiologia Slovenica*, *29*(1). https://doi.org/10.52165/kinsi.29.1.138-148

Morgan, W. (2018, May 21). *How StatsBomb Data Helps Measure Counter-Pressing*. StatsBomb. https://statsbomb.com/articles/soccer/how-statsbomb-data-helps-measure-counter-pressing/

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F. M., & Baca, A. (2023). Machine learning application in soccer: a systematic review. *Biology of Sport*, *40*(1). https://doi.org/10.5114/biolsport.2023.112970

StatsBomb Privacy Policy. (2022, November). StatsBomb. https://statsbomb.com/privacy-policy/

The Coaches' Voice. (2023, September). *The 4-3-3: football tactics explained*. The Coaches' Voice. https://www.coachesvoice.com/cv/4-3-3-football-tactics-explained-formation-liverpool-klopp-barcelona-guardiola/#:~:text=The%204%2D3%2D3%20is%20also%20a%20good%20formation%20from,into%20a%20more%20reserved%20block.

Trainor, C. (2014, July 30). *Defensive Metrics: Measuring the Intensity of a High Press*. StatsBomb. https://statsbomb.com/articles/soccer/ defensive-metrics-measuring-the-intensity-of-a-high-press/