Literature Review

Data-Driven Football Analytics and Set Pieces

The adoption of football data has been on the rise recently–from scouting (e.g. using data to uncover hidden talents)(Data Sports Group, 2020) to fan engagement (e.g. analyzing ticket sales)(Breaking The Lines, 2026) to tactical analysis (e.g. understanding team patterns to provide adjustments and recommendations)(Sportblog, 2025).

Data-driven tactical analysis in football leverages machine learning using tracking and event data to optimise on-field decisions, with set pieces contributing 30-40% of goals across major leagues (Fernández-Navarro et al., 2019). Tracking data providers such as SkillCorner capture spatiotemporal data, showing the coordinates of each player throughout every frame of the match, while event data providers such as StatsBomb, annotate on-ball actions (e.g. shots, passes, dribbles, etc.), used in over 170 professional competitions worldwide for performance analysis (Hudl StatsBomb, n.d.). These systems evolved from occupancy maps and pass networks (Lucey et al., 2014) to multi-agent models distinguishing team formations and playing styles (Bialkowski et al., 2016). However, goal-kick specific studies remain underexplored.

Goal kicks demonstrate set pieces' tactical importance, restarting play from deep while facing pressure from the opponent, constraining build-up options (Billings, 2024). Success in these cases may be defined by how well a team escaped pressure or how well they create dangerous chances from build-up. This can be measured by progression (entries into the final third of the field), xG Chain (a.k.a Expected Goals, xG is the probability of a goal from a given position) or possession value models (PSV) such as On Ball Value (OBV) (StatsBomb, 2021). Early analytics works focused on plotting distributions, such as heatmaps and passing networks, and counting outcomes such as shots, xG from set-pieces etc. (Smith, 2020). With the advancement in machine learning and use of tracking data, we can now uncover latent patterns like pressing styles, coordinated runs, or space creation via unsupervised learning (e.g. clustering techniques), generating phase-specific insights (Decroos et al., 2019).

Despite these advances, challenges still persist. Studies that average team tactics across whole seasons overlook restart moments such as goal kicks (Plakias et al., 2023), while ball focused metrics such as Expected Threat (xT), disregard pressing team shapes. In addition, PSV models also rarely account for opponent defensive setups (Pulis & Bajada, 2022). Transparent models are still essential for coaches to adopt and trust analytics, using tools to explain why specific patterns matter–for example, why compact front lines reduce opponents' OBV and other Possession based metrics.

This study addresses these gaps by analysing goal-kick phases using hybrid data, theoretically clustering pressing styles while leveraging industry labels, and predicting OBV outcomes for tactical recommendations.

Tracking Data, Team Structures, and Pressing Styles

Tracking data captures player and ball positions at high frequency across GPS coordinates covering the full pitch. Providers such as Prozone and SkillCorner (Catapult Sports, 2024) supply tracking data primarily for major leagues like the Premier League and LaLiga.

This spatiotemporal resolution enables analysis of movement patterns beyond traditional event data. However, it does face challenges such as dynamic role interchanges mid-phase, misalignments between tracking frames and events, and the inherent complexity of coordinating 22 players plus the ball.

As football positions become more fluid, players are constantly interchanging positions mid-phase, causing the structure and formation of the teams to become unclear. To address this issue, Bialkowski et al. (2014) developed the EM-Hungarian algorithm, which discovers formations and player roles based on trajectories from tracking data. This produces heatmaps that have achieved 75% agreement with expert labels–transforming analysis from generic dots to tactically meaningful positions.

Plakias et al. (2023) reviewed pressing styles amongst Europe’s top competition, UEFA Champions League, by classifying high press teams versus mid/low block teams. Passes per Defensive Action (PPDA) was the metric used to assess team pressing styles. In short, PPDA is defined by how many passes opponents can complete before your team tackles, intercepts, or fouls. High pressing teams would have a lower PPDA, typically below 8, while mid/low block teams averaged higher PPDA, usually above 12.

During goal kicks, pressing is expressed through defensive structure–characterized by line compactness (the distance between players), pressing height (y-coordinates of the highest attacker), and forward aggression–which this study examines within short windows from the start of the goal kick.

Possession State Value Models

Possession State Value Models (PSV) quantify action quality beyond raw counts. Evolving from shot-based metrics to comprehensive state evaluations critical for build-up analysis like goal kicks.

Expected Goals (xG) models estimate shot quality, predicting the probability of a goal happening based on distance, angle, type and other variables. xGChain and xGBuildup models distribute credit across all possession actions contributing to the shot, showing when and how xG is influenced based on possession (Hudl Statsbomb, 2025).

Expected Threat (xT) models, such as the one developed by SoccerActions, applies a grid-based model where passes or carries into higher-threat zones generate value. It typically uses a Markov decision process to learn the value of specific zones on the pitch. However, its ball-centric nature overlooks pressure and defensive structure, and only focuses on whether the ball has entered a certain area (Decroos et al., 2019).

Valuing Actions by Estimating Probabilities (VAEP) (Decroos et al., 2019), is PSV model from KU Leuven, that quantifies actions by the change in the probability of scoring minus the change in probability of conceding ($\Delta P\_score - \Delta P\_concede$) over subsequent events.

On-Ball Value (OBV), StatsBomb's model, trains separate goal-for and goals-against models over extended possessions, integrating context variables from set pieces and opponent team pressure (Hudl Statsbomb, 2021). Pulis and Bajada (2022) compared OBV against VAEP, finding similar performance, but with OBV having an advantage in contextual scenarios such as pressing resistance.

For goal-kick analysis, OBV seems like the most suitable model, as it values progression and build-up actions while taking into account contextual factors and opponent pressing structures, making it the primary target metric for this study alongside VAEP and xT comparisons.

Goal-Kick Restarts and Pressing: Tactical and Applied Work

Tactical coaching approaches emphasize precise pressing triggers during goal kicks. For example, FIFA Training Centre drills teach 4-4-2 pressing shapes, a typical football formation which sets up the team with 4 defenders, 4 midfielders and 2 forwards, totaling 10 outfield players, excluding the goal keeper. Such a formation is used to press the team in possession, minimizing their passing options and forcing them to play a long goal-kick, instead of build up from the goalkeeper (FIFA Training Center, 2024). Short goal kicks trigger immediate 2v1 situations on the receiver and goalkeeper, while long kicks force teams to drop into a mid/low block, exploiting space behind aggressive defensive lines (Modern Soccer Coach, 2024). This creates a trade-off between aggressive pressing and allowing the goalkeeper to play short passes and build-up from the back, where aggressive pressing allows the defensive team to win the ball back quicker and in more dangerous positions, but also leaves the vulnerable to long passes behind their defensive line.

In terms of academic works on set-pieces, the focus is almost always directed towards corners and free kicks (delivery types, zonal vs man-marking, etc.) rather than goal-kicks, leaving this essential routine underexamined, despite it occurring countless times throughout a game (Casal et al., 2015). Industry analytics like StatsBomb have tracked goalkeeper distribution trends since recent rule changes, where they found an increase in short goal kicks and decrease in long corner kicks since 2019. This study was mainly descriptive, focusing on end-locations of goal kicks, shown through heatmaps, without the use of OBV models or taking into account pressure from the opposing team (Statsbomb, 2024).

No studies have focused on pressing structures and PSV metrics to evaluate goal-kick decision quality while under pressure. This study fills the gap for tactical relevance.

Summary of Gaps and Contributions

Existing analytics examines formations and pressing styles at match or season levels (Plakias et al., 2023; Bialkowski et al., 2014), rarely isolating short restart windows like goal kicks. Hybrid event-tracking datasets enable tactical insights

(Kwiatkowski, 2020; SciTePress, 2024), yet none target goal-kick pressing interactions specifically.

Possession value models such as xT, VAEP, and OBV quantify actions effectively (Decroos et al., 2019; Statsbomb, 2021; Pulis & Bajada, 2022), but rarely condition outcomes on contemporaneous defensive structures. Industry tools visualise goalkeeper distributions descriptively (Statsbomb, 2024), without press-aware decision support for restart scenarios.

This study therefore asks: To what extent do different out-of-possession pressing structures during goal kicks predict build-up success measured by OBV and progression metrics?

References

Bialkowski, A., Lucey, P., Carr, P., Foote-Ennever, Y., Harlick, A., Dennis, J., & Matthews, I. (2014). Representing team behaviours from noisy data using player role. *MIT Sloan Sports Analytics Conference*.

Smith, J. (2020, January 15). *Visually analysing direct set pieces in football using StatsBomb data, R and Tableau*. SportsChord.

https://www.sportschord.com/blogs/blog/visually-analysing-direct-set-pieces-in-football-using-statsbomb-data-r-and-tableau

Breaking The Lines. (2026, January 15). *Why sports investment trends are starting to influence football analytics*.

https://breakingthelines.com/opinion/why-sports-investment-trends-are-starting-to-influence-football-analytics/

Catapult Sports. (2024). *Advanced football analytics for match strategies*.

https://www.catapult.com/blog/advanced-football-analytics

Data Sports Group. (2020, May 19). *How football scouting is changed and improved by big data analytics*.

https://datasportsgroup.com/news-article/117722/how-football-scouting-is-changed-and-improved-by-big-data-analytics/

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

FIFA Training Centre. (2024). *8v8 attack v. defence: Pressing the goal kick*.

https://www.fifatrainingcentre.com/en/practice/talent-coach-programme/press-and-regain/8v8-attack-v-defence-pressing-the-goal-ki

Fernández-Navarro, J., Fradua, L., Zubillaga, A., & McRobert, A. P. (2019). Attacking and defensive styles of play in soccer: Analysis of Spanish and English elite teams. *Journal of Sports Sciences*, *37*(24), 2820–2827.

https://doi.org/10.1080/02640414.2019.1652615

Hudl Statsbomb. (n.d.). *Hudl Statsbomb - The world's most advanced football data*.

https://www.hudl.com/en_gb/products/statsbomb

Kwiatkowski, W. (2020). *The right way to synchronise event and tracking data*.

https://kwiatkowski.io/sync.soccer

Modern Soccer Coach. (2024, July 7). *What's the best pressing shape from goal-kicks?*

https://www.modernsoccercoach.com/post/what-s-the-best-pressing-shape-from-goal-kicks

Plakias, S., Armatas, V., & Mitrotasios, M. (2023). Identifying soccer players' playing styles: A systematic review. *International Journal of Sports Science & Coaching*.

https://doi.org/10.1177/17479541231190579

Pulis, M., & Bajada, J. (2022). *Reinforcement learning for football player decision making*. StatsBomb Conference.

SciTePress. (2024). *Estimation of overlapped tactical actions from soccer tracking and event data*.

https://www.scitepress.org/Papers/2024/130663/130663.pdf

Sportblog. (2025, July 23). *The data revolution in the Premier League: How tactical analysts changed the game*.

https://www.sportblog-online.de/en/premier-league-data-revolution-tactical-analysts/

Sports Data Campus. (n.d.). *Data analysis transforms tactics and scouting in football*. UCAM English Programs.

https://english-programs.sportsdatacampus.com/data-analysis-transforms-tactics-scouting-football/

Statsbomb. (2021). *Introducing on-ball value (OBV)*.

https://blogarchive.statsbomb.com/news/introducing-on-ball-value-obv/

Statsbomb. (2024, October 1). *How football is changing*.

https://statsbomb.com/articles/soccer/how-football-is-changing/