# *The relationship of team and individual athlete performances on match quarter outcome in elite women's Australian Rules football*



Cust, Emily, Sweeting, Alice, Ball, Kevin, Anderson, Hamish and Robertson, Samuel (2019) The relationship of team and individual athlete performances on match quarter outcome in elite women's Australian Rules football. Journal of Science and Medicine in Sport. ISSN 1440-2440

# Accepted Manuscript


Title: The relationship of team and individual athlete performances on match quarter outcome in elite women's Australian Rules football

Authors: Emily E. Cust, Alice J. Sweeting, Kevin Ball, Hamish Anderson, Sam Robertson

PII: S1440-2440(19)30075-1
DOI: https://doi.org/10.1016/j.jsams.2019.05.004
Reference: JSAMS 2070

To appear in: *Journal of Science and Medicine in Sport*

Received date: 20 January 2019
Revised date: 17 April 2019
Accepted date: 7 May 2019

Please cite this article as: Cust EE, Sweeting AJ, Ball K, Anderson H, Robertson S, The relationship of team and individual athlete performances on match quarter outcome in elite women's Australian Rules football, *Journal of Science and Medicine in Sport* (2019), https://doi.org/10.1016/j.jsams.2019.05.004

# The relationship of team and individual athlete performances on match quarter outcome in elite women's Australian Rules football

Emily E. Cust[1, 2], Alice J. Sweeting[1, 2], Kevin Ball[1], Hamish Anderson[2], Sam Robertson [1, 2]

**Author details:**

[1]Institute for Health and Sport (IHES), Victoria University, Melbourne, Australia

[2]Western Bulldogs Football Club, Footscray, Melbourne, Australia

**Corresponding author:** Associate Professor Sam Robertson Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001, Australia

Email: sam.robertson@vu.edu.au

## Abstract

**Objectives:** To evaluate the relationships between the athlete distribution of team performance indicators and quarter outcome in elite women's Australian Rules football matches.

**Design:** Retrospective longitudinal cohort analysis

**Method:** Thirteen performance indicators were obtained from 56 matches across the 2017 and 2018 Australian Football League Women's (AFLW) seasons. Absolute and relative values of 13 performance indicators were obtained for each athlete, in each quarter of all matches. Eleven features were further extracted for each performance indicator, resulting in a total of 169 features. Generalised estimating equations (GEE) and regression decision trees were run across the different feature sets and dependent variables, resulting in 22 separate models.

**Results:** The GEE algorithm produced slightly lower mean absolute errors across all dependent variables and feature sets comparative to the regression decision tree models. Quarter outcome was more accurately explained when considered as total points scored comparative to quarter score margin. Team differential and the 75th percentile of individual athlete Inside 50s were the strongest features included in the models.

**Conclusions:** Modelling performance statistics by quarter outcomes provides specific practical information for in-game tactics and coaching in relation to athlete performances each quarter. Within the current elite women's Australian Rules football competition, key high performing individual athletes' skilled performances within matches contribute more to success rather than a collective team effort.

**Keywords**

Sports analytics; AFLW; machine learning; performance analysis

**Introduction**

Match performance analysis in team sports can provide a greater understanding of the physical, technical and tactical characteristics athletes require to produce a successful competition outcome.[1] Analysis may help guide coaching staff on training practices that replicate and prepare athletes for the demands of competition.[2] Determining the form and function of events within the specifics of a sport for teams and individual athletes should inform the variables for quantification of performance and therefore the sport analytics approaches used to facilitate future coaching practise.[3] The relationship between match athlete performance indicators[1] in Australian Rules football (AF) have been investigated heavily in the literature across elite male teams[4–6], individual athlete contributions[7], and recently, elite women's teams.[8]

In 2017, AF established a national elite women's competition, the Australian Football League Women's (AFLW) in addition to the long running elite men's Australian Football League (AFL). The opening two seasons consisted of a seven-round home-and-away competition, incorporating eight teams. As the depth of talent and resources develop, the league has set plans for expansion to the competition. This in turn will provide further opportunities to investigate elite women's football training and match physical, technical and tactical areas. For example, information on athlete match demands may improve club training practices, assess the effectiveness of the rule changes implemented differently to the AFL competition, and inform league directors on the quality of development in the competition.

Research in women's AF is currently limited.[8–10] Recent research on the physical demands, technical performances and activity profiling across field playing positions of match-play in AFLW[9] has provided initial insights into match activity. There were no absolute differences between physical variables, based on match playing position, in the AFLW[9]. Furthermore, no positional group differences were noted for skill measures such as total kicks, handballs, contested possessions, uncontested possessions, and marks.[9] This is in contrast to several physical demands characteristic differences that have been observed across athlete match positions in the AFL.[11,12] The specificity of AFLW positional roles may not yet be established and consequently, athletes may be more homogenous in playing tactics and physical abilities comparative to AFL players.[9] Although there are inherent differences between the AFL and AFLW games such as amount of time and players on ground creating independent constraints between each competition. Currently focussing on the AFLW as an independent competition and quantifying match variables as the league matures may be more beneficial over a direct sport analytics comparison of the AFL and AFLW given the current game constraint differences. Match performance indicator analysis assessed the relationship between team skill involvements and match outcome in the first season of AFLW.[8] Match outcome, defined as win/ loss and score margin, indicated that higher uncontested possessions and Inside 50: goal score ratio were the strongest predictors for winning. Increased kick numbers and contested marks resulted in a higher team ladder position.[8]

Match success in the AFL has been linked to individual athlete skill efficiency rather than their physical activity profile.[13] Specifically, physical activity profiles may increase, yet skill involvements efficiency may decrease when teams lose a quarter.[14] An analysis inclusive of athlete skilled match performances, by individual match quarter and across feature derived performance distributions, is yet to be investigated in AF. A quarter by quarter approach could provide differentiated information about specific technical and tactical foci for coaches. Situational variables such as starting quarter score, quality of opposition, and whether the team is playing at a home or away ground have shown influence on elite women's team sport quarter performances.[15,16] Analysing by quarter could improve relevancy of results, given output may fluctuate across quarters for several reasons.[17] During quarter time breaks, coaches can address athletes directly. Knowledge or information transfer from the coach to the playing group should be of purpose, work in context of the current events and tie in with previously delivered knowledge the coach has provided prior to the match to maximise group understandings of the information.[18] Factors may affect the extent of knowledge transfer to the playing group between the restricted quarter time frame such as the coach's communication style, clarity of information, and a player's prior involvement in the match strategy system development.[18]

Quantifiable information about skill performances, in context of the match, could further justify changes to team playing strategies based on the current situation. With respect to influence on the team match outcome, quantification of individual athlete distributions have been linked to successful match outcome.[7] Specifically, lower 75th, 90th and 95th percentile values for team goals and higher 25th and 50th percentile values for disposals.[7] Measured athlete performance distribution information calculated by individuals rather than a team data as a whole could determine the influential basis for match success in the AFLW. Information may also convey whether success in the current AFLW game constitutes a more collective team-based effort or skewed to a few stronger individual athletes. Findings may inform match team selection to suit the current game style influence or opposition at play. This may be important as several new teams are introduced to the competition over the next few years making key athlete retention or attainment a challenge.

The primary aim of this study was to evaluate the relationship of AFLW athlete skill performance indicator distributions, to explain match quarter outcomes during the 2017 and 2018 seasons. Secondly, this study aimed to compare quarter outcome model error rates from separate machine learning approaches, based on the varied input feature set variables.

**Methods**

All match performance indicators were obtained from the AFL match statistics provider, Champion Data Pty Ltd. (Melbourne, Australia) online portal, Coaches Information Analysis (CIA). Data collection by Champion Data involves human recordings of the statistics by working at each match, as such the inter- and intra-reliability of the data is currently unknown.[7,19] Reliability and validity of the data has been assessed independently to determine the agreement between the Champion Data and author-coded values.[7] Reliability assessment showed very high agreement levels, intra-class correlation coefficient range 0.947 – 1.000. The validity of author's coding showed low absolute error in regards to the Champion Data, RMSE range 0.0 – 4.5,[7] indicating the expected absolute error points between each performance indicator for each game. A total of 56 matches across the 2017 and 2018 AFLW season were obtained and 13 discrete performance indicators were selected.[6–8] The definitions for each indicator are provided in Appendix A. Absolute values from every quarter (n = 224), match (n = 56), athlete (n = 154), and all teams (n = 7), across performance indicators, were extracted into a custom Excel™ spreadsheet. Quarter outcome (as win = 1 or loss = 0 or draw = 2), quarter score margin (points), match outcome (win/ loss) and match score margin (points) were recorded. Score points were recorded as both their absolute values and relative values to the opposition at play. The University's Human Research Ethics Committee approved the study (application number 0000025654).

Each athlete's contribution to their team's total were converted to a relative form, as a percentage of their team total for each match.[7] Features extracted for each performance indicator were the minimum, maximum, mean, standard deviation and percentiles, at 0.05, 0.10, 0.25, 0.50, 0.75, 0.90 and 0.95, resulting in a total of 143 features (11 features x 13 performance indicators).[7] Features were collated with team name, round number (1 – 7), season (2017 or 2018), quarter number (1 – 4), quarter outcomes (loss, win or draw), and match outcome (loss, win or draw). The stability of the data performance profiles [20] was plotted and assessed by visual inspection, and deemed acceptable to model for comparison of analysis methods and reporting of results for practical feedback.

A total of 22 models were developed. Modelling of statistics by machine learning was performed for quarter points scored (absolute), and quarter point margin relative to the opposition (relative). Four features sets were used in separate models: total performance indicator values (n = 13), performance indicator values relative (n = 13) to the opposition, derived feature distribution values for each performance indicator (n = 143), combined performance indicator total, relative and feature distribution values (n = 169).

Regression decision trees were computed with Python version 3.6.6 [21], using the package Scikit-learn.[22] Data was split into a 70% training set and 30% testing set. Each regressor tree was computed with a minimum sample split of 30 (>13.4% of total sample) and a maximum depth of five. Several model parameter combinations were tested to reduce the risk of overfitting whilst minimising error.[23] Regression trees were also computed using the whole training set for the four feature sets as a comparison. Generalised Estimating Equations (GEE) were also constructed separately in R[24] for each dependent variable and feature sets. Team (n = 7) was considered a fixed repeated measure and a greedy feature selection was implemented for feature selection in model construction. Model evaluation was based on the mean absolute error (MAE) computed from the withheld testing set, unless otherwise indicated.

**Results**

The MAE results for each model are presented in Table 1. The GEE produced lower MAE's than the decision trees (Table 1). Across both analysis approaches, the influence of performance indicators was more accurately explained by quarter score points, as opposed to quarter score margin, for all input feature set variables. The mean average difference between score margin and score points MAE results was 2.32 points (Table 1). Modelling performance statistics by quarter score points using the relative values feature set (n = 13) resulted in one of the lower MAE scores for both the GEE (3.83) and the decision tree (5.59). The lowest prediction errors for both models were on larger feature sets. The GEE MAE was 3.60 on the 169-feature set comprised of the combined total, relative and feature distribution values. The decision tree MAE was 5.45 on the 143-feature set comprised of the derived feature distribution values.

Rule outputs from the two regressor decision tree models, with the lowest MAE, are shown in Figures 1 and 2. The relative performance indicator of team differential of Inside 50 values (Figure 1) and feature distribution Inside 50s in the 75th percentile (Figure 2) contributed most strongly to the models. Interpretation involves following the branches down, from the root node representing the outcomes for each test, to the final terminal node to define the regression decision rules for the model. For example, in Figure 1 following down the right side, teams with relative Inside 50s greater than -3.5, relative kicks long greater than 5.5 scored more points per quarter, model prediction of 18.5 points based on 22 samples. Teams with higher contributions from more athletes to their Inside 50 count, short and long kicks, and lower contributions from more athletes to their ineffective kick counts are more successful per quarter (Figure 2). See Figure 1 and Figure 2 for further examples for rule sets. The defined rules represent performance skill fulfilment requirements for teams to achieve a successful quarter score or score margin outcome.

**** INSERT TABLE 1 ABOUT HERE ****

**** INSERT FIGURE 1 ABOUT HERE ****

**** INSERT FIGURE 2 ABOUT HERE ****

**Discussion**

This study assessed the extent to which AFLW athlete skill performance distributions explain match quarter outcome across the first two seasons of the inaugural AFLW national competition. Key results indicate that modelling data by quarter score points total was more accurate compared to quarter score margin. Teams with more successful Inside 50 entries than their opposition likely scored more points in the quarter.

Modelling performance indicator data by quarter and not an entire match may allow for specific information and clearer relationships between the variables and success within different periods of a match.[16] During matches, coaches have the chance to address the playing group and reset tactics at quarter time breaks. Specific quarter-based skill influence information may aid in modifying individual athlete and team tactics, in comparison to the opposition as shown in elite women's basketball.[25] Therefore, breaking performance indicator data into the influence by quarter may provide targeted information for coaches during matches. As the league expands and more data becomes available, longitudinal comparisons would be of interest. In comparison of the two approaches, the GEE produced lower prediction errors across all data input variables. This may indicate that a simpler model approach is more appropriate for the current smaller dataset with relatively low feature dimensionality. However, to provide a practical outcome for coaches, a decision tree model may be more applicable as the output does not consider all features. Rather, decision trees provide a practical, parsimonious rule set for coaches who may be focused on the most influential performance indicators.

Features or variables are representative aspects of data that should be relevant, in that they have an influence on the model result with a function that is not assumed by the rest.[26] Performance indicators that were a direct function of scoring in AFLW, including shots at goal, goal assist, behind assist and goal accuracy were not included in this study. These variables would potentially trivialise the process of determining performance skills which influence match success. Modelling quarter points scored produced the lowest prediction errors on the larger feature sets GEE (n = 169) and decision tree (n = 143) for both algorithms. But this was only a slight improvement from using the smaller relative values feature sets (n = 13). A larger data set could facilitate improved feature extraction and selection engineering for better representation of the data characteristics. More efficient algorithm processing and prediction accuracy[27] may also be increased. Further extracting distribution features, from individual athletes, demonstrates the structure contributions for AFLW teams. Interestingly, results suggest that in contrast to the AFL game, increased match skill performance contributions from key high performing individual athletes is more beneficial for team success. This is suggested by the higher percentile feature distributions contributing most strongly to the decision model (Figure 2). For example, the Inside 50 P75, short kick P95, long kick P90 and ineffective kick P75 values.

Successful outcomes in the AFL involve relatively even performances from athletes across a team.[7] The comparatively higher performance contributions by key individual athletes to team success in the AFLW may be explained by the fact it is a new competition format and across many facets is still developing. As such, the level of game plan seen in the AFL competition [7,28,29] may yet be reasonable in the AFLW due to the variety of AFLW athlete game experience and skill maturity levels being contracted. The skill development of AFLW athletes, who have either recently progressed from junior competitions or transitioned from another sport and hence not marquee or high performing athletes may be also affected by the lower resourced professional support structures and training opportunities currently experienced in the AFLW. As opposed to the well-established AFL, where newly contracted athletes are highly coached, skill acquired and AF experienced before competing in AFL level matches.[30] This may be partly because of lower coaching and sports science resource support in the

AFLW competition relative to elite male AF. These factors could be contributing to individual athlete dominance in the AFLW, potentially preventing collective team contributions towards successful match outcome.

Comparison of the current results to AFLW match skill analysis during the 2017 season only[8] is difficult, due to the differentiating features sets used. In Black et al. (2018), variables with direct functions of scoring were used. In order to build upon this previous analysis,[8] further data feature extraction from a larger sample size and revised statistical modelling was run in the present study. Breaking down the performance indicators to types of the variable, for example, including long, short and ineffective kicks allows for expansion of the key performance measures.

Practically, as the strongest features in the regression decision tree models relate to kick performance indicators, clubs may look to emphasise kick skill development. Inside 50's, hit outs and contested possessions, by key athletes, contribute most to quarter success during matches. AFLW clubs may also look to compile teams with capable skilled kickers and recruit future athletes with current or potential strong kick skills.[5] Game plan development around a kick dominant ball movement strategy, particularly in hit-out clearances and efficient Inside 50 entries may also be of match tactical advantage. Coaches may work specifically with key forward and midfield athletes to develop efficient plays and decision making from centre bounce to Inside 50 entry possession chains, in order to maximise scoring opportunities. Improving an athlete's kick execution skills may also benefit kick delivery and mark success from a team member in contested possessions during matches. Analysis of match performance statistical information can also be applied off field in the athlete recruitment department.[19] As the AFLW is in its infancy, a greater understanding of team and individual contributions to winning may highlight what performance characteristics are beneficial towards maximising team success. Recruiters could make strategic decisions on selecting athletes that currently exhibit or have the potential to develop the key performance characteristics identified.

Future research may look to investigate the contextual variables around match play on the outcome such as travel requirements, days between matches and player interchange rotations per quarter. Specifically, given the current short home-and-away season, increased importance is on the outcome of each match for ladder positioning. Across different team sports, contextual variables influence match outcomes and performance indicators,[25,31] particularly in team field sports who play multi-round home-and-away seasons.[32–34] Furthermore, spatiotemporal data characteristics of players could be analysed to explain team behaviours in match play styles and tactics[29] in the AFLW.

## Conclusion

Quarter success in the AFLW was characterised by greater Inside 50s as a relative to the opposition and key athletes in the 0.75 percentile performing Inside 50s. Results suggest within the current AFLW competition, key athletes' skilled performances are contributing more to match success rather than a collective team effort as opposed to the AFL competition. Using machine learning methods in sport analytics to uncover practical information from athlete match performance statistics allows for analysis on how these athletes are contributing towards team success. Post-hoc reporting of results, in a comprehensible format for coaching staff, may provide a basis for training and match strategic planning.

## Practical Applications

- Identifying key contributing athlete match skills in the AFLW can practically inform coaches on training drills, athlete development, and tactical match approaches per quarter relative to their opposition.
- Current AFLW match success is influenced more by marquee athletes within a team as opposed to an even team contribution, which indicates the need for athlete development support and resources to be improved in women's football.
- Results position kick variables as strong performance indicators hence training plans should place emphasis on individual kick skill development.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Acknowledgements

There was no financial assistance associated with this study.

**References**


1. Hughes MD, Bartlett RM. The use of performance indicators in performance analysis. *J Sport Sci* 2010; 20(10):739–54.
2. Pinder RA, Davids K, Renshaw I, et al. Representative learning design and functionality of research and practice in sport. *J Sport Exerc Psychol* 2011; 33(1):146–55.
3. Hughes M. Sports analysis. In: Essentials of Performance Analysis in Sport Routledge; 2015. p. 141–51.
4. McIntosh S, Kovalchik S, Robertson S. Examination of player role in the Australian Football League using match performance data. *Int J Perform Anal Sport* 2018; 18(3):451–62.
5. Stewart MF, Mitchell H, Stavros C. Moneyball applied: econometrics and the identification and recruitment of elite Australian footballers. *Int J Sport Financ* 2007; 2(4):231–48.
6. Robertson S, Back N, Bartlett JD, et al. Explaining match outcome in elite Australian Rules football using team performance indicators. *J Sports Sci* 2016; 34(7):637–44.
7. Robertson S, Gupta R, McIntosh S. A method to assess the influence of individual player performance distribution on match outcome in team sports. *J Sports Sci* 2016; 34(19):1893–900.
8. Black GM, Gabbett TJ, Johnston RD, et al. A skill profile of the national women's Australian football league (AFLW). *Sci Med Footb* 2018; June 30:1–5.
9. Clarke AC, Ryan S, Couvalias G, et al. Physical demands and technical performance in Australian Football League Women's (AFLW) competition match-play. *J Sci Med Sport* 2018; 21(7):748–52.
10. Black GM, Gabbett TJ, Johnston RD, et al. The influence of physical qualities on activity profiles of female Australian football match play. *Int J Sports Physiol Perform* 2018; 13:524–9.
11. Gray AJ, Jenkins DG. Match analysis and the physiological demands of Australian football. *Sport Med* 2010; 40(4):347–60.
12. Boyd LJ, Ball K, Aughey RJ. Quantifying external load in Australian football matches and training using accelerometers. *Int J Sports Physiol Perform* 2013; 8(1):44–51.
13. Sullivan C, Bilsborough JC, Cianciosi M, et al. Factors affecting match running performance in professional Australian football. *Int J Sports Physiol Perform* 2014; 9(3):561–6.
14. Sullivan C, Bilsborough JC, Cianciosi M, et al. Match score affects activity profile and skill performance in professional Australian Football players. *J Sci Med Sport* 2014; 17(3):326–31.
15. Moreno E, Gomez MA, Lago-Peñas C, et al. Effetcs of starting quarter score, game location, and quality of opposition in quarter score in elite women's basketball. *Kinesiology* 2013; 45(1):48–54.
16. Gómez MÁ, Serna A Dela, Lupo C, et al. Effects of situational variables and starting quarter score in the outcome of elite women's water polo game quarters. *Int J Perform Anal Sport* 2014; 14(1):73–83.
17. Sampaio J, Lago C, Casais L, et al. Effects of starting score-line, game location, and quality of opposition in basketball quarter score. *Eur J Sport Sci* 2010; 10(6):391–6.
18. Joshi KD, Sarker S, Sarker S. Knowledge transfer within information systems development teams: Examining the role of knowledge source attributes. *Decis Support Syst* 2007; 43(2):322–35.
19. McIntosh S, Kovalchik S, Robertson S. Validation of the Australian Football League player ratings. *Int J Sport Sci Coach* 2018; 13(6):1064–71.
20. Hughes M. Notational analysis – a mathematical perspective. *Int J Perform Anal Sport* 2017; 4(2):97–139.
21. Foundation PS. Python programming language [Internet]. 2018.
22. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011; 12:2825–30.
23. Hawkins DM. The problem of overfitting. *J Chem Inf Comput Sci* 2004; 44(1):1–12.
24. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
25. Gómez MA, Lorenzo A, Ibañez SJ, et al. Ball possession effectiveness in men's and women's elite basketball according to situational variables in different game periods. *J Sports Sci* 2013; 31(14):1578–87.

26. L. Ladha, T. Deepa. Feature Selection Methods and Algorithms. *Int J Comput Sci Eng* 2011; 3(5):1787–97.
27. Liu H, Motoda H, Setiono R, et al. Feature selection: an ever evolving frontier in data mining. In: JMLR: Workshop and Conference Proceedings 10: Feature Selection in Data Mining Hyderabad, India; 2010. p. 4–13.
28. Woods CT, Robertson S, Collier NF. Evolution of game-play in the Australian Football League from 2001 to 2015. *J Sports Sci* 2017; 35(19):1879–87.
29. Alexander JP, Spencer B, Mara JK, et al. Collective team behaviour of Australian Rules football during phases of match play. *J Sports Sci* 2019; 37(3):237–43.
30. Haycraft JAZ, Kovalchik S, Pyne DB, et al. Physical characteristics of players within the Australian Football League participation pathways: a systematic review. *Sport Med - Open* 2017; 3(1).
31. Oliveira T, Gómez M, Sampaio J. Effects of game location, period, and quality of opposition in elite handball performances. *Percept Mot Skills* 2012; 114(3):783–94.
32. Taylor J, Mellalieu S, James N, et al. The influence of match location, quality of opposition, and match status on technical performance in professional association football. *J Sports Sci* 2008; 26(9):885–95.
33. García-Rubio J, Gómez MÁ, Lago-Peñas C, et al. Effect of match venue, scoring first and quality of opposition on match outcome in the UEFA champions league. *Int J Perform Anal Sport* 2015; 15(2):527–39.
34. Ruano MA, Serna AD, Lupo C, et al. Effects of game location, quality of opposition, and starting quarter score in the outcome of elite water polo quarters. *J Strength Cond Res* 2016; 30(4):1014–20.

**Figure legends**

**Figure *1*.** Regressor decision tree output of model 14, quarter score points and performance indicator relative values. *diff*, differential, *mse*, mean sample error

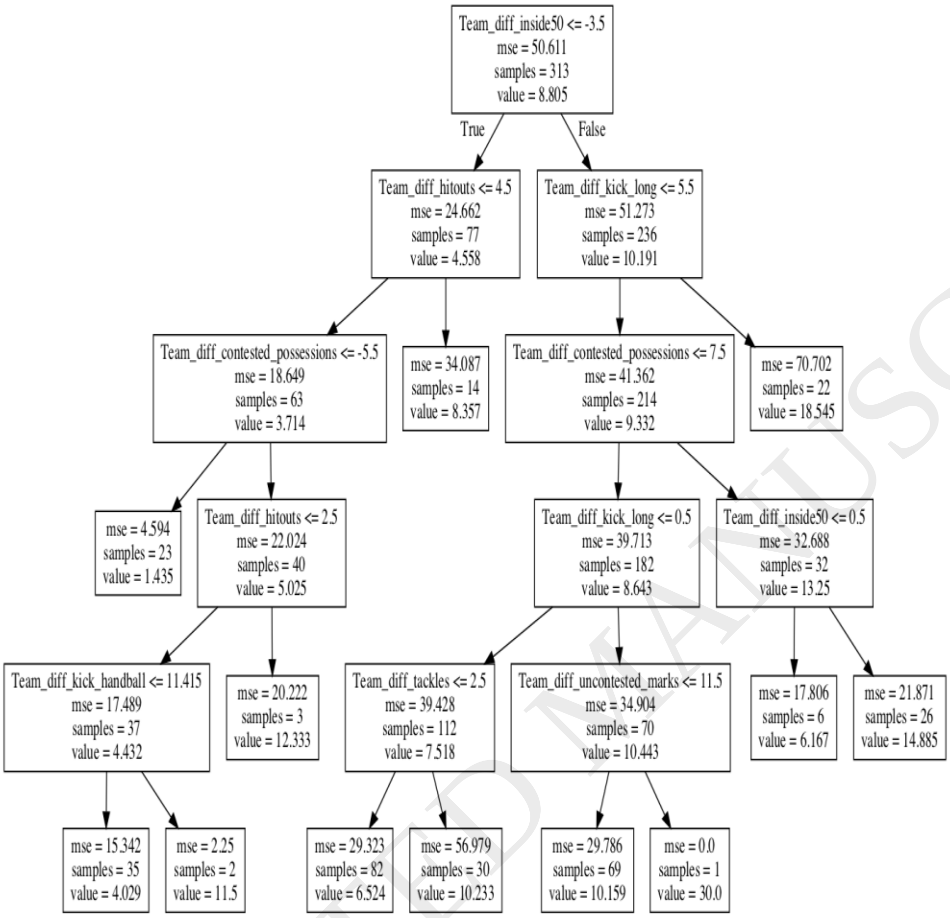


Figure 1

**Figure *2*.** Regressor decision tree output of model 15, quarter score points and performance indicator feature distributions. *mse*, mean sample error; *P25, P50, P75, P90, P95,* percentile level; *stdev*, standard deviation

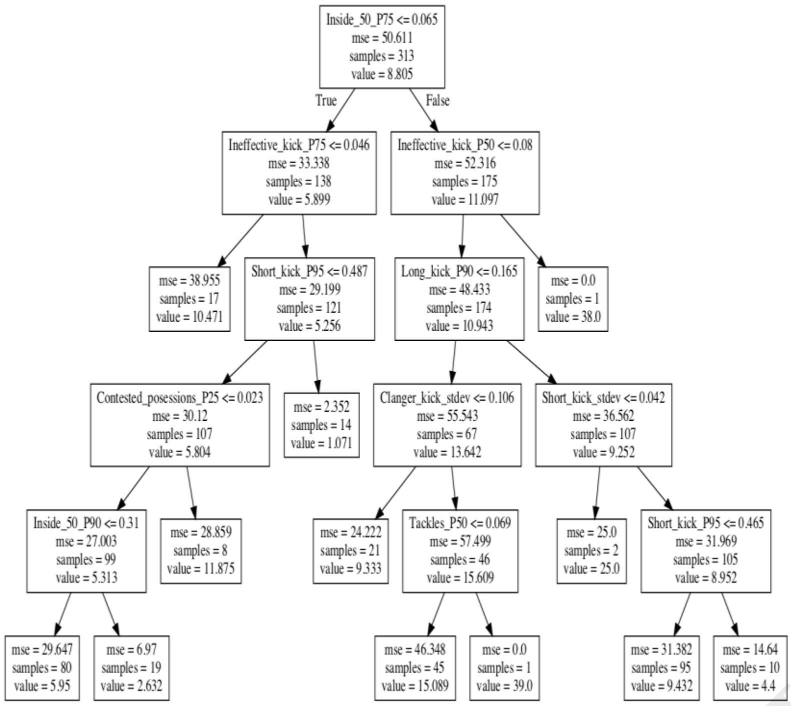
Inside_50_P75 <= 0.065
mse = 50.611
samples = 313
value = 8.805
True
False
Ineffective_kick_P75 <= 0.046
mse = 33.338
samples = 138
value = 5.899
Ineffective_kick_P50 <= 0.08
mse = 52.316
samples = 175
value = 11.097
mse = 38.955
samples = 17
value = 10.471
Short_kick_P95 <= 0.487
mse = 29.199
samples = 121
value = 5.256
Long_kick_P90 <= 0.165
mse = 48.433
samples = 174
value = 10.943
mse = 0.0
samples = 1
value = 38.0
Contested_posessions_P25 <= 0.023
mse = 30.12
samples = 107
value = 5.804
mse = 2.352
samples = 14
value = 1.071
Clanger_kick_stdev <= 0.106
mse = 55.543
samples = 67
value = 13.642
Short_kick_stdev <= 0.042
mse = 36.562
samples = 107
value = 9.252
Inside_50_P90 <= 0.31
mse = 27.003
samples = 99
value = 5.313
mse = 28.859
samples = 8
value = 11.875
mse = 24.222
samples = 21
value = 9.333
Tackles_P50 <= 0.069
mse = 57.499
samples = 46
value = 15.609
mse = 25.0
samples = 2
value = 25.0
Short_kick_P95 <= 0.465
mse = 31.969
samples = 105
value = 8.952
mse = 29.647
samples = 80
value = 5.95
mse = 6.97
samples = 19
value = 2.632
mse = 46.348
samples = 45
value = 15.089
mse = 0.0
samples = 1
value = 39.0
mse = 31.382
samples = 95
value = 9.432
mse = 14.64
samples = 10
value = 4.4

Figure 2

## Tables

**Table 1.** Model results across data variables evaluated by mean absolute error (MAE).

| Algorithm | Parameters | Model | Variables (number of features) | MAE (points) |
|---|---|---|---|---|
| Regressor Decision Tree | Total training data | 1 | Quarter margin and PI totals (13) | 6.65 |
| | | 2 | Quarter margin and PI relative (13) | 5.93 |
| | | 3 | Quarter margin and PI features distributions (143) | 6.81 |
| | | 4 | Quarter margin and combined PI totals, relatives and feature distributions (169) | 5.98 |
| | | 5 | Quarter score and PI totals (13) | 4.47 |
| | | 6 | Quarter score and PI relative (13) | 4.31 |
| | | 7 | Quarter score and PI feature distributions (143) | 4.32 |
| | | 8 | Quarter score and combined PI totals, relatives and feature distributions (169) | 4.10 |
| Regressor Decision Tree | 70% train set and 30% test set | 9 | Quarter margin and PI totals (13) | 8.56 |
| | | 10 | Quarter margin and PI relative (13) | 7.63 |
| | | 11 | Quarter margin and PI features distributions (143) | 9.57 |
| | | 12 | Quarter margin and combined PI totals, relatives and feature distributions (169) | 8.38 |
| | | 13 | Quarter score and PI totals (13) | 5.60 |
| | | 14 | Quarter score and PI relative (13) | 5.59 |
| | | 15 | Quarter score and PI feature distributions (143) | 5.45 |
| | | 16 | Quarter score and combined PI totals, relatives and feature distributions (169) | 5.74 |
| Generalised Estimating Equations | Total training data | 17 | Quarter margin and PI totals (13) | 7.13 |
| | | 18 | Quarter margin and PI relative (13) | 6.18 |
| | | 19 | Quarter margin and PI features distributions (143) | 6.03 |
| | | 20 | Quarter margin and combined PI totals, relatives and feature distributions (169) | 5.12 |
| | | 21 | Quarter score and PI totals (13) | 4.48 |
| | | 22 | Quarter score and PI relative (13) | 4.64 |
| | | 23 | Quarter score and PI feature distributions (143) | 3.83 |
| | | 24 | Quarter score and combined PI totals, relatives and feature distributions (169) | 3.60 |

*PI,* performance indicator