

Kicking for Goal or Touch? An Expected Points Framework for Penalty Decisions in Rugby Union

Abstract

Following a penalty in rugby union, teams typically choose between attempting a shot at goal or kicking to touch to pursue a try. We develop an *Expected Points* (EP) framework that quantifies the value of each option as a function of both field location and game context. Using phase-level data from the 2018/19 Premiership Rugby season (35,199 phases across 132 matches) and an angle-distance model of penalty kick success estimated from international records, we construct two surfaces: (i) the expected points of a possession beginning with a lineout, and (ii) the expected points of a kick at goal, taking into account the in-game consequences of made and missed kicks. We then compare these surfaces to produce decision maps that indicate where kicking for goal or kicking to touch maximizes expected return, and we analyze how the boundary shifts with game context and the expected meters gained to touch. Our results provide a unified, data-driven method for evaluating penalty decisions and can be tailored to team-specific kickers and lineout units. This study offers, to our knowledge, the first comprehensive EP-based assessment of penalty strategy in rugby union and outlines extensions to win-probability analysis and richer tracking data.

1 Introduction

Rugby union presents a recurring strategic decision immediately after a penalty with four possible ways to restart play: going for an immediate three points via a kick at goal or pursuing a potential seven-point try by either a scrum, a tap-and-go (both at the point of the penalty) or a kick to touch where the attacking team restarts play with a lineout. Generally, except when right on the opposition goal-line, a lineout is perceived to be a superior choice than a tap-and-go or a scrum since the attacking team can safely advance the ball down the field, with the lineout occurring where the ball crosses the touch line following the kick. This means that, once within distance to make a penalty kick $\sim 60\text{m}$ from the opponent try line, the decision following a penalty is routinely between taking a lineout or kicking at goal.

While the choice is often straightforward near the opposition's corner, given a certain point difference, or late in a match, many game states are ambiguous and governed by convention, perceived kicker ability, and the expected quality of a team's lineout. In contrast, American football has long relied on *Expected Points* (EP), *Expected Points Added* (EPA), and *win-probability* (WP) frameworks to guide fourth-down decision-making, translating field position

and context into expected returns and actionable recommendations (Romer, 2006; Burke, 2009; Yurko et al., 2018). In that setting, WP and EP estimators are typically built using statistical machine learning models (e.g., Random Forests and gradient boosting), as is done in Lock and Nettleton (2014) and Baldwin (2021).

In rugby union, public research is only beginning to close this methodological gap, which can be largely attributed to a lack of publicly-available data. Martinez-Arastey et al. (2025) propose a transparent, phase-level modeling framework for calculating Expected Points on English Premiership data, highlighting the difficulty of reliably predicting minority outcomes (e.g., penalty kicks) and providing a reproducible benchmark rather than a deployable decision tool. Separately, public-facing analyses model goal-kicking success as a function of distance and lateral angle, offering practically-useful (albeit context-free) estimates (Monpezat, 2023). Building on these strands—and informed by lessons from the NFL literature on decision modeling—we directly evaluate the penalty decision.

Our contribution is a decision-ready EP framework for the penalty decision. We construct two context-aware surfaces: (i) the expected points of a possession beginning with a lineout, estimated from phase-level data by field location, and (ii) the expected points of a kick at goal, which integrates a continuous angle-distance model of kick success together with the *continuation value* of misses (e.g., 22 m drop-outs and in-play recoveries) rather than treating misses as zero. Comparing these surfaces yields decision maps (zero-difference contours) indicating where each option maximizes expected return, and sensitivity analyses showing how the frontier shifts with kicker accuracy and expected meters gained to touch. We also make explicit the translation from penalty location to resulting lineout location and document assumptions required when bridging club-phase data and international kicking records. To our knowledge, this is the first comprehensive EP-based assessment of penalty strategy in rugby union that produces decision-ready artifacts; complementing prior methodological advances and isolated kicking models while laying the groundwork for future win-probability analysis in rugby. We proceed with an overview of the data, then continue with an outline of our methodology and results.

2 Data

2.1 Phase-Level Data

Our primary dataset consists of phase-level event logs from the 2018/19 Premiership Rugby season compiled by Martinez-Arastey et al. (2025), covering 132 matches and 35,199 phases of play. Each record describes the on-field state at the start of a phase: the initiating event type, a zonal field location and side, the current score differential, and the next score in points for or against the team in possession. We harmonize timestamps to “seconds remaining in match” so that time is strictly decreasing within a game and directly comparable across halves.

The unit of observation in the dataset is a *phase*, defined as the period between subsequent rucks. The phase count resets following restarts (e.g., lineouts, scrums, kick-offs) or major infringements that halt play. A *possession sequence* is an ordered set of phases by the

same team without an opposition touch that changes possession. Unless otherwise stated, Expected Points estimates condition on the *first* phase of a possession sequence (e.g., a lineout), because that is the decision-relevant entry state. Figure 1 summarizes the zonal discretization of the pitch used in the phase-level data.

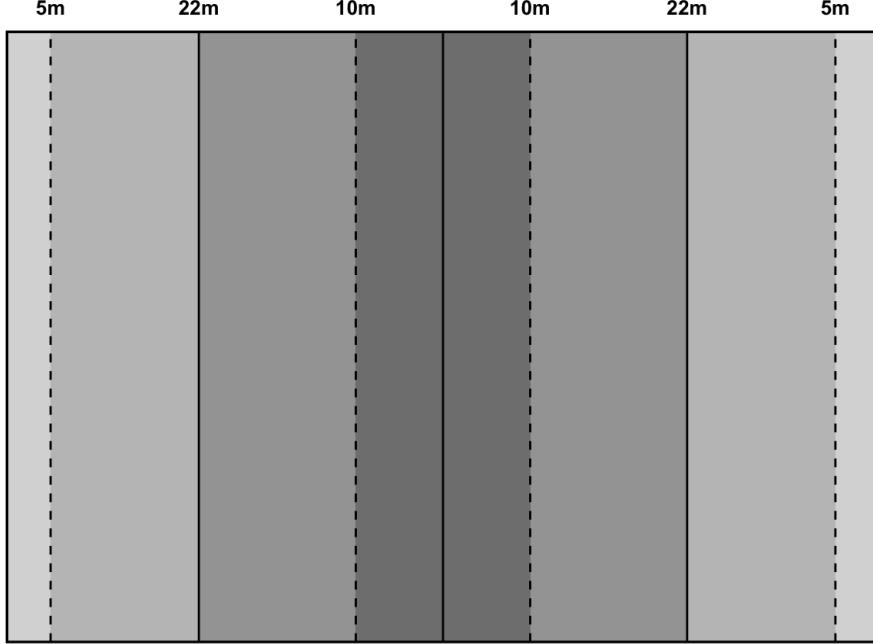


Figure 1: Field zones used in the phase-level dataset

2.2 Phase-Level Data Processing

For the EP model, the unit of analysis is a possession that begins with a lineout. We therefore restrict the dataset to opening phases of play (i.e., observations with Phase = 1) whose initiating event is a lineout. Each such record represents a distinct possession entry state. We then construct several derived variables that encode field position, time, and game context:

- **Field position.** The original zone labels are converted into a continuous coordinate, `meter_line`, which records the distance in meters from the team’s own goal line at which the lineout originates.
- **Time remaining.** Total seconds remaining in the match are transformed to `Sec_Remain_Half`, the seconds remaining in the current half, to better align with rugby’s two-period structure.
- **Score and card context.** `Points_Difference` records the score differential (for the team in possession), while `Card_Diff` tracks the net player advantage in terms of yellow or red cards.

- **Team strength.** `WinPct_Diff` is the running difference in win percentage between each team and its opponent at kickoff, computed from their prior matches.

To address potential double counting across consecutive possessions, we create a grouping identifier, `run_id`, which indexes sequences of possessions that (i) begin with lineouts for the same team, (ii) share the same point differential at the start of the possession, and (iii) have the same eventual scoring outcome. Within each `run_id` we also compute `n_same`, the number of such consecutive possessions. The role of `run_id` in the EP model is discussed in Section 3.2. The final structure of this reduced dataset is illustrated in Table 1.

Table 1: Processed Rugby Union Phase-Level Dataset

ID	Round	Home	Away	Phase	Team_In_Poss	Points_Difference	Sec_Remain_Half	Card_Diff	WinPct_Diff	meter_line	run_id	n_same	Points
8	1	Harlequins	Sale	1	Home	0	2302	0	0	31.0	1	1	3
10	1	Harlequins	Sale	1	Home	3	2166	0	0	45.0	2	1	-3
14	1	Harlequins	Sale	1	Away	-3	2103	0	0	13.5	3	2	3
17	1	Harlequins	Sale	1	Away	-3	2043	0	0	31.0	3	2	3
53	1	Harlequins	Sale	1	Home	0	1313	0	0	55.0	4	1	7
56	1	Harlequins	Sale	1	Away	0	1188	0	0	86.5	5	1	-7
67	1	Harlequins	Sale	1	Away	-7	926	0	0	31.0	6	1	7

2.3 Kicking Data

Our kicking data come from the analysis in [Quarrie and Hopkins \(2015\)](#), which studied 582 international rugby matches played between 2002 and 2011. The resulting dataset consists of 3,802 penalties, aggregated into a 5 m-by-5 m grid over the pitch, with the observed proportion of successful kicks recorded for each grid square. These cell-level success rates serve as the raw input for our continuous angle-distance model of kick success. The structure of this dataset is included in Table 2.

Table 2: Structure of the penalty-kicking dataset. Each entry gives the observed proportion of successful penalties in that grid cell.

Distance from Try Line (m)	Distance from Left-Hand Touchline (m)													
	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	55–60	60–65	65–70
0–5
5–10
10–15
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
60–65

3 Methodology

3.1 Expected Points Framework

We evaluate the penalty decision—kicking at goal versus kicking to touch—by comparing the expected points (EP) of the two options at a given field location. Let (x, y) denote the location of a penalty on the field, where x is the distance in meters from the opposition try line and y is the lateral distance from the center of the pitch. For a given assumed

gain in territory from a kick to touch, denoted d_{touch} , we define the decision quantity as the difference in expected points, with positive values favoring a lineout:

$$\Delta EP(x, y; d_{\text{touch}}) = EP_{\text{lineout}}(x_{\text{LO}}) - EP_{\text{kick}}(x, y),$$

where $EP_{\text{kick}}(x, y)$ is the expected points from attempting a kick at goal from (x, y) , and $EP_{\text{lineout}}(x_{\text{LO}})$ is the expected points of the next scoring event from the ensuing lineout at location x_{LO} on the sideline. Here x_{LO} denotes the resulting lineout x -coordinate after kicking to touch; we make this translation rule explicit in Section 5.1.

The remainder of this section describes how we estimate EP_{lineout} and EP_{kick} .

3.2 Expected Points of a Lineout

To value a lineout, we model the expected points of the *next* scoring event using a multiple linear regression, based on possession sequences that begin with a lineout in a given field zone. Only the first phase of each possession is included, so that downstream phases from the same decision state are not double-counted. Formally,

$$EP_{\text{lineout}} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{meter_line} + \hat{\beta}_2 \cdot \text{Card_Diff} + \hat{\beta}_3 \cdot \text{WinPct_Diff},$$

where the $\hat{\beta}$ coefficients are estimated from the regression. This specification captures downstream contingencies observed in matches: losing one's own lineout, turnovers in subsequent phases, or the opponent scoring following a turnover. The resulting quantity represents the *expected next-score value* of having a lineout at that field position, given competition-level play and the current game context. In the decision maps below, we evaluate this surface at the `meter_line` implied by each lineout location x_{LO} .

3.2.1 Dealing with Double Counting

Double counting can arise not only at the phase level (which we have already addressed by restricting to Phase = 1), but also across multiple consecutive possessions with effectively the same state and outcome. Table 3 illustrates three successive possessions that all begin with a lineout for the same team, under the same score differential, and with the same eventual points outcome.

Table 3: Multiple possessions with the same outcome following consecutive lineouts

ID	Round	Home	Away	Phase	Team_In_Poss	Points_Difference	Sec_Remain_Half	Card_Diff	WinPct_Diff	meter_line	run_id	n_same	Points
19	1	Harlequins	Sale	1	Home	21	454	0	0	69.0	11	3	7
20	1	Harlequins	Sale	1	Home	21	378	0	0	13.5	11	3	7
21	1	Harlequins	Sale	1	Home	21	146	0	0	13.5	11	3	7

Consider a group of n consecutive lineouts sharing the same `run_id`. Treating all n as independent observations would over-represent that specific game situation in the regression. To prevent this, we randomly retain a single representative observation from each group and discard the remaining $n - 1$. Let Y_1, \dots, Y_n denote the n lineouts in such a group. The retained observation \tilde{Y} is drawn uniformly:

$$\tilde{Y} \sim \text{Uniform}(\{Y_1, \dots, Y_n\}).$$

This procedure ensures that each `run_id` contributes at most one observation to the lineout EP model, mitigating bias from repeated, structurally identical possessions. Our approach is conceptually similar to the cluster-level subsampling used by [Brill et al. \(2025\)](#), who randomly retain a subset of plays per simulated game to reduce within-game dependence when estimating win probability. After sampling, we obtain a total of 2,046 lineout observations for estimation.

3.2.2 Regression Summary

The estimated regression coefficients for the lineout EP model are reported in Table 4.

Table 4: Regression summary for expected points of a lineout

Coefficient	Estimate	Std. Error	t value	Pr(> t)
Intercept	3.2545	0.2093	15.553	< 2e-16 ***
meter_line	-0.0586	0.0044	-13.283	< 2e-16 ***
Card_Diff	0.8802	0.3052	2.884	0.00397 **
WinPct_Diff	0.6503	0.3430	1.896	0.05809 .

The negative coefficient on `meter_line` reflects the intuitive decrease in expected next-score value as the lineout moves farther away from the opposition try line. Positive coefficients on `Card_Diff` and `WinPct_Diff` indicate that both a player advantage and stronger underlying team quality are associated with higher expected points from the same field position. Figure 2 visualizes the relationship between field position and expected points for a lineout, holding other covariates fixed.

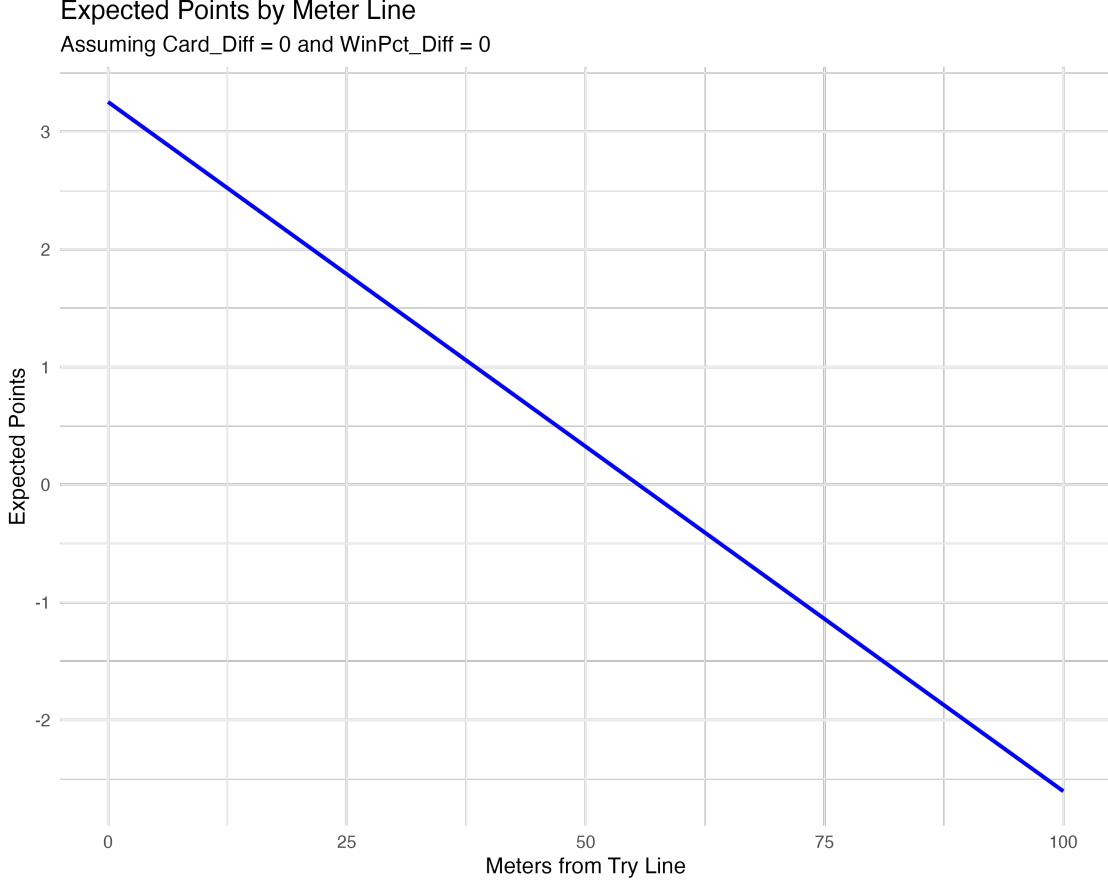


Figure 2: Average expected points of the next score given lineout position

3.3 Expected Points of a Penalty Kick

The expected points of a penalty kick at goal is the sum of the value of a make and the continuation value of a miss, weighted by their respective probabilities. Unlike lineouts, kick outcomes depend primarily on the kick’s distance and angle to the posts rather than just its raw field coordinates. Let d denote the distance from the penalty spot to the center of the posts, and θ the lateral angle. Then

$$EP_{\text{kick}}(d, \theta) = \mathbb{P}_{\text{make}}(d, \theta) \cdot 3 + (1 - \mathbb{P}_{\text{make}}(d, \theta)) \cdot EP_{\text{miss}}(d, \theta),$$

where 3 points are awarded for a successful kick, $\mathbb{P}_{\text{make}}(d, \theta)$ is the probability of a successful kick from (d, θ) , and $EP_{\text{miss}}(d, \theta)$ is the expected points of the continuation state following a miss (e.g., change in possession, change in field position).

The field coordinates (x, y) of the penalty uniquely determine (d, θ) through the geometry of the pitch, so $EP_{\text{kick}}(x, y)$ can be obtained by evaluating $EP_{\text{kick}}(d, \theta)$ at the corresponding (d, θ) . In the next section, we describe how we estimate $\mathbb{P}_{\text{make}}(d, \theta)$ using the international kicking data and how we approximate $EP_{\text{miss}}(d, \theta)$ from restart sequences in the phase-level dataset.

4 Modeling Kick Outcomes

4.1 Estimating Kick Success Probability

We model the probability that a penalty kick at goal is successful as a smooth function of the kick's *distance* (d) and *lateral angle* (θ) to the center of the posts. Our model takes the form

$$\mathbb{P}(\text{make} \mid d, \theta) = \text{logit}^{-1}\{\beta_0 + f(d, \theta)\},$$

where $f(d, \theta)$ is a two-dimensional spline estimated by a generalized additive model (GAM) with a logistic link and a quasi-binomial family to allow for overdispersion. Smoothing is selected by Restricted Maximum Likelihood to balance fit and smoothness. We validate the surface with standard calibration diagnostics and cross-validation. Estimated kick success probabilities, with dashed lines at probabilities of $(0.2, 0.4, 0.6, 0.8)$, are shown in Figure 3.

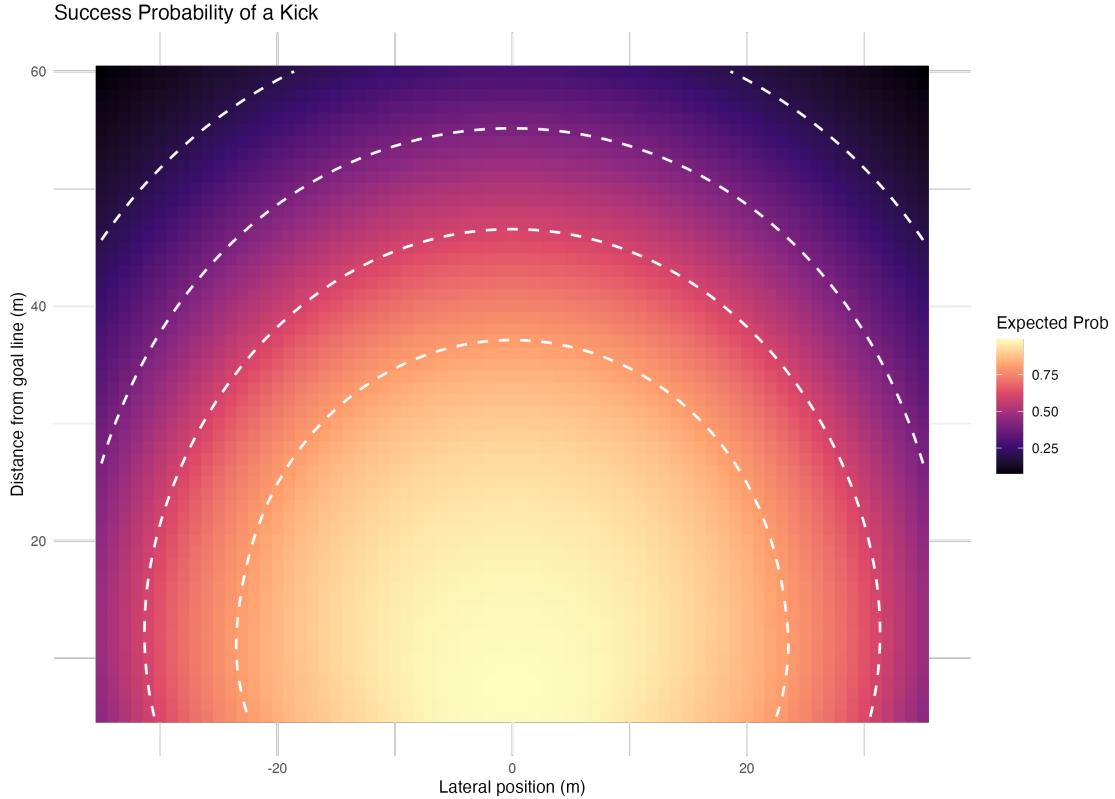


Figure 3: Estimated probability of kick success by position on the pitch. Note that the goal posts are located at lateral position 0. Kick probabilities begin at 5 m from the goal line

4.2 Continuation Value of a Missed Kick

In rugby, a missed penalty does not end the sequence immediately. After a miss, the ball can be grounded in-goal by the defending team or it can travel past the dead-ball line, in which

case play restarts with a 22 m drop-out and the penalty-kicking team regains possession. If the ball remains live, the defense may return it from inside the 22. In practice, kicks that are returned in play occur relatively infrequently, and the labeling in our phase data does not reliably distinguish returned penalties from other open-play kicks. We therefore adopt the simplifying assumption that a missed penalty leads to a 22 m drop-out restart.

Operationally, we identify restarts that (i) do not follow a change in score (excluding post-score kick-offs) and (ii) did not occur at the start of a half, and compute the expected next-score value of those entries by restart zone, using the same EP framework as for lineouts (Section 3.2). Table 5 reports average expected points following kick restarts by original kick location. These restart values provide a coarse but data-driven approximation to $EP_{\text{miss}}(d, \theta)$ as a function of the original kick location.

Table 5: Average expected points following kick restarts by zone

Location	n	Avg. Expected Points
10m–22m (opp)	4	2.75
Half–10m (opp)	17	1.24
10m–Half (own)	27	-0.63
22m–10m (own)	45	1.24
Overall Average	93	0.76

4.3 Penalty Kick Expected Points Surface

Combining the kick success surface and the continuation values yields an expected points surface for penalty kicks. For each location on the field, we map (x, y) to (d, θ) , evaluate $\mathbb{P}_{\text{make}}(d, \theta)$ from the GAM, and approximate $EP_{\text{miss}}(d, \theta)$ using the restart values in Table 5. These quantities are then substituted into

$$EP_{\text{kick}}(d, \theta) = \mathbb{P}_{\text{make}}(d, \theta) \cdot 3 + (1 - \mathbb{P}_{\text{make}}(d, \theta)) \cdot EP_{\text{miss}}(d, \theta)$$

to obtain the expected points of a kick at goal from that location. The resulting expected points surface for penalty kicks is shown in Figure 4.

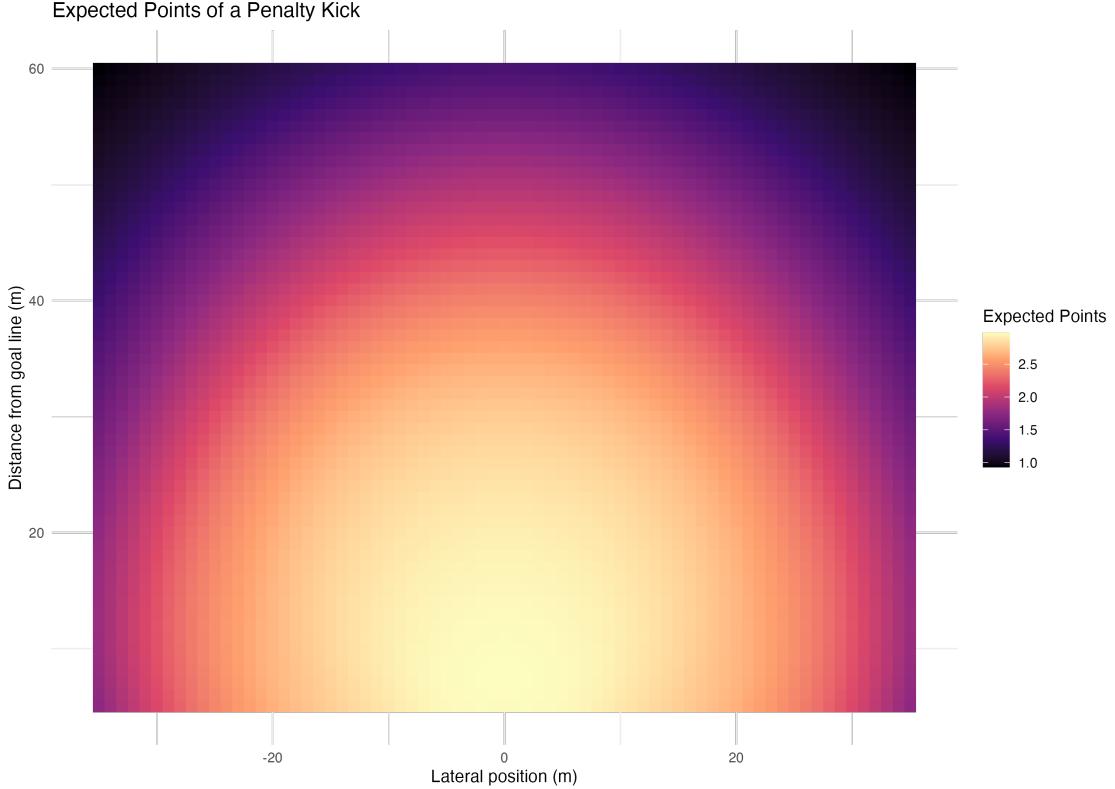


Figure 4: Estimated expected points of a penalty kick by position on the pitch

5 Constructing the Penalty Decision Surface

5.1 Adjusting for Meters Gained on Kick to Touch

At each penalty location on the field, we compare the value of attempting a kick at goal with the value of kicking to touch and taking the ensuing lineout. The kick-at-goal value is calculated at the penalty spot, using the distance and lateral angle to the posts implied by that location. By contrast, the lineout value is calculated at the location where the lineout would actually occur *after* the kick to touch.

Let (x, y) denote the penalty location, and let d_{touch} denote the expected meters gained from a kick to touch along the length of the field. Starting from x (distance from the opposition try line), we approximate the resulting lineout location as

$$x_{\text{LO}} = \max\{5, x - d_{\text{touch}}\},$$

shifting the ball d_{touch} meters closer to the opposition try line and truncating at the 5 m line if necessary, since the laws prohibit lineouts closer than 5 m from the try line. The lateral coordinate y is assumed unchanged by the kick to touch, so the lineout occurs at (x_{LO}, y) .

For a given assumed gain d_{touch} , the expected points of choosing the lineout option from penalty location (x, y) is then

$$EP_{\text{lineout}}(x_{\text{LO}}),$$

evaluated via the regression in Section 3.2. In our baseline decision maps we consider a grid of plausible translation distances

$$d_{\text{touch}} \in \{0, 5, 10, 15, 20, 25\} \text{ m},$$

reflecting different assumptions about how much territory can be gained by kicking to touch.

5.2 Decision Maps

Given EP_{lineout} and EP_{kick} , the relative value of a lineout versus a kick at goal at the same penalty location is

$$\Delta EP(x, y; d_{\text{touch}}) = EP_{\text{lineout}}(x_{\text{LO}}) - EP_{\text{kick}}(x, y),$$

where (x, y) is the penalty location, d_{touch} is the assumed meters gained by kicking to touch, and $x_{\text{LO}} = \max\{5, x - d_{\text{touch}}\}$ is the resulting lineout x -coordinate. Positive values of $\Delta EP(x, y; d_{\text{touch}})$ favor choosing the lineout, while negative values favor taking the kick at goal, with $\Delta EP(x, y; d_{\text{touch}}) = 0$ tracing out the indifference frontier between the two options.

By mapping out $\Delta EP(x, y; d_{\text{touch}})$ across the field, we obtain decision maps that show how location and assumed meters gained shape the optimal choice. Figure 5 displays these maps for a range of values of d_{touch} .

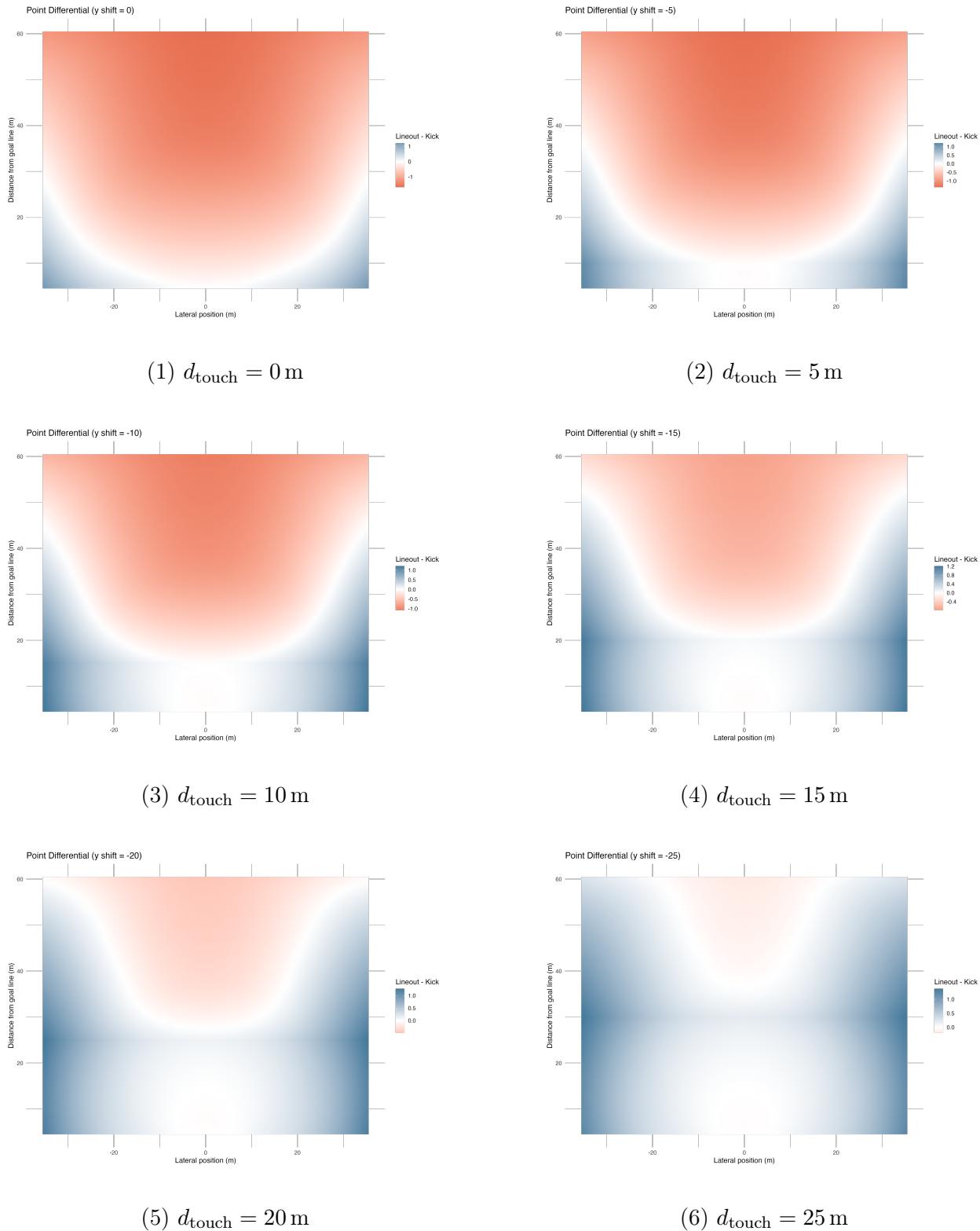


Figure 5: Decision maps comparing the expected points of a lineout versus a kick at goal for different assumed meters gained to touch d_{touch} .

6 Scenario Analysis

Changing the game context—that is, changing the covariates in our regression model—allows us to construct hypothetical decision surfaces under different circumstances.

6.1 Manpower Advantages and Disadvantages

In rugby union, manpower differentials occur when a player receives a yellow or red card and is sent off the field (for 10 minutes for a yellow and for the rest of the game for a red card). This directly affects the `Card_Diff` covariate in the lineout EP model.

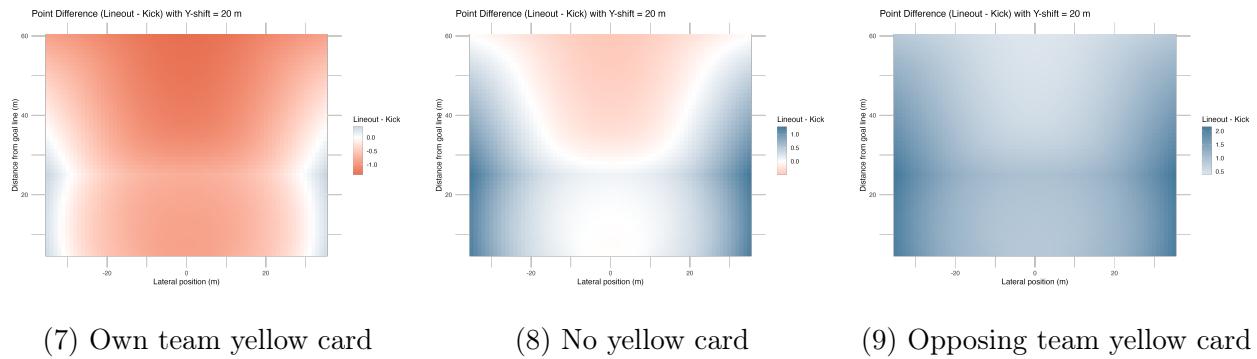


Figure 6: Decision maps under different manpower states, holding other covariates fixed.

As shown in Figure 6, the decision surface changes significantly depending on the presence and direction of a yellow card, with kicks being preferred more often when the attacking team is down a player and lineouts being preferred more often when the attacking team has a numerical advantage.

6.2 Strong vs. Weak Teams

We can also examine how the decision surface changes for teams with positive or negative win-percentage differentials by varying `WinPct_Diff`.

As illustrated in Figure 7, stronger teams (positive `WinPct_Diff`) realize larger expected points from lineouts across much of the pitch, making the aggressive option more attractive. Weaker teams (negative `WinPct_Diff`) are relatively better off taking points from penalty kicks, though the effect is less pronounced than the impact of yellow cards.

7 Case Study: New Zealand vs. South Africa

We now apply the model to a real match to illustrate both single-decision analysis and aggregate decision quality. The data were collected by the authors from a match between New Zealand and South Africa played on September 16th, 2025.

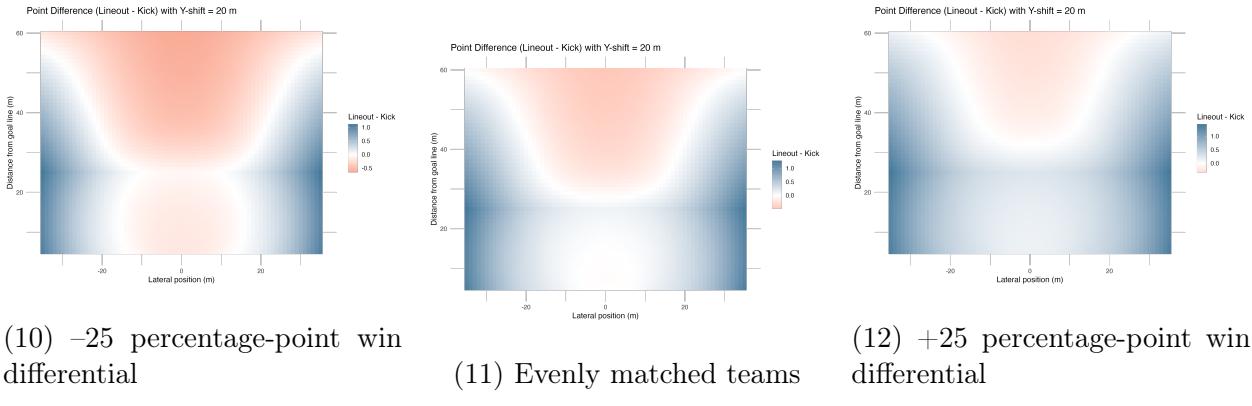


Figure 7: Decision maps for teams with different win-percentage differentials.

7.1 A Single Penalty Decision

Consider a penalty awarded to South Africa in the 22nd minute, 15 m from the left touch line and 30 m from the opposition try line. We can set up this decision graphically by varying the expected distance that a kick to touch would gain. There were no yellow or red cards in effect, and we assume the teams were evenly matched (difference in win percentage equal to 0).

Figure 8 displays $\Delta EP(x, y; d_{\text{touch}})$ as a function of d_{touch} for this specific penalty location. The lineout becomes preferable once the expected gain exceeds roughly 16 m.

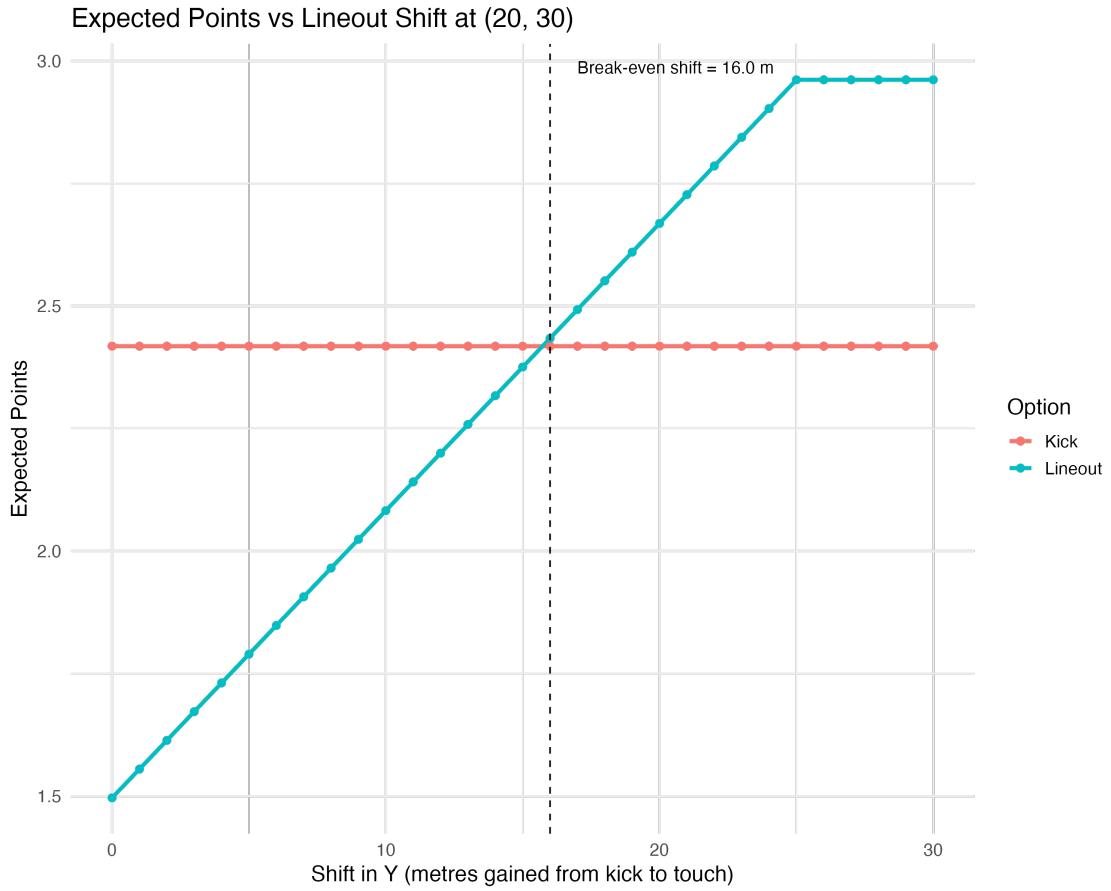


Figure 8: Expected points difference between lineout and kick at goal as a function of meters gained to touch

From this field position it would be reasonable to gain 20–25 m of territory by opting for the lineout. Conservatively assuming a 20 m translation, the model’s recommendation at the penalty location is shown in Figure 9.

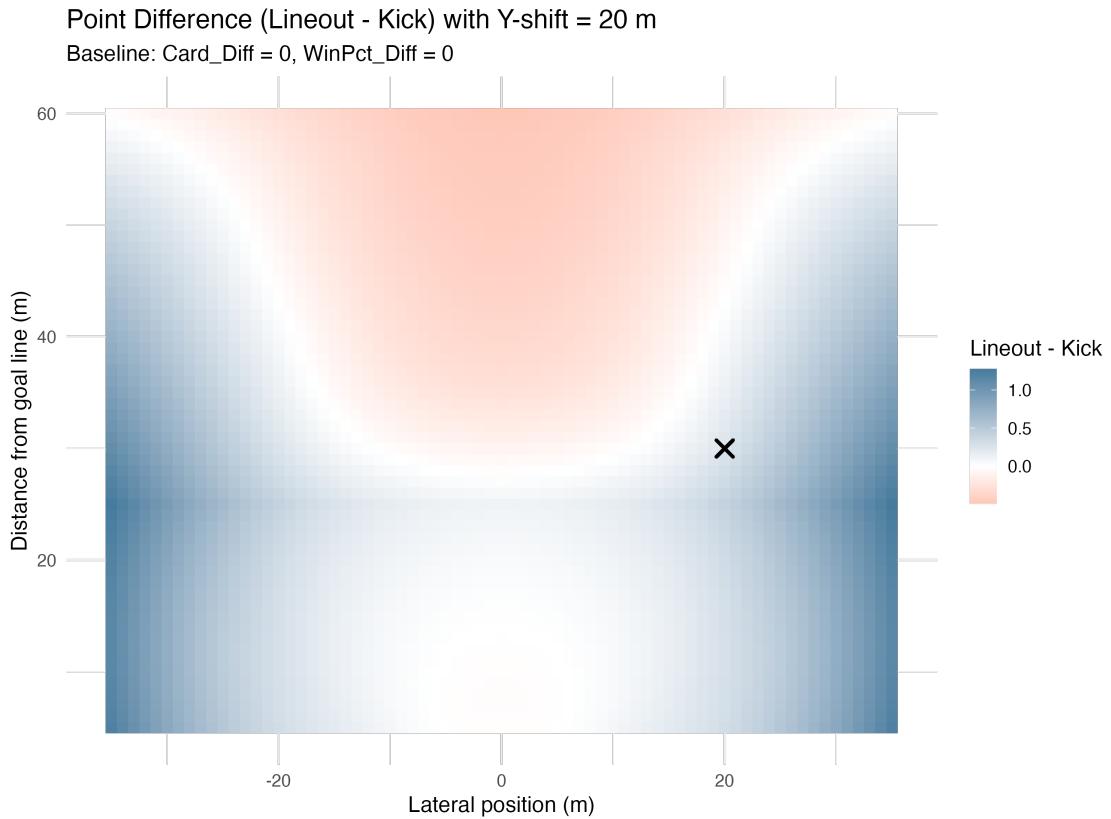


Figure 9: Penalty decision at the observed field position (marker denotes the penalty location)

At this point on the pitch, the optimal decision from the model is to go for the lineout. The expected points for each option at the marker location, together with the resulting $\Delta EP = EP_{\text{lineout}} - EP_{\text{kick}}$, are reported in Table 6.

Table 6: Expected points of lineout vs. kick at the marker location

	EP Lineout	EP Kick	ΔEP
Value	2.67	2.42	0.25

Here $\Delta EP = 0.25$ in favour of the lineout, implying that by electing to kick at goal in this situation, South Africa, in expectation, forfeited approximately 0.25 points relative to the lineout option.

7.2 All Penalty Decisions in the Match

We can also compare decision quality over all relevant penalties in the match. For each penalty, we define a *regret* quantity

$$R = EP_{\text{optimal}} - EP_{\text{actual}},$$

where EP_{optimal} is the expected points of the model-preferred option (lineout or kick) and EP_{actual} is the expected points of the option actually chosen. Positive values of R indicate points lost in expectation by deviating from the model’s recommendation. Table 7 presents all penalty decisions taken within 60 m of the opponents’ try line for both teams, together with the expected points of each option, the model-implied optimal decision, and the associated regret. We assume a flat 20 m gained for each lineout and again treat the teams as equally matched.

Table 7: Expected points for lineouts and kicks: actual vs. optimal decision

Team	Lineout EP	Kick EP	Decision	Optimal Decision	R
NZ	1.79	1.85	kick	kick	0.00
NZ	2.26	1.93	lineout	lineout	0.00
SA	1.56	1.87	lineout	kick	0.31
SA	2.67	2.42	kick	lineout	0.25
SA	1.91	1.66	kick	lineout	0.25
SA	1.50	1.60	lineout	kick	0.10
NZ	2.73	2.64	kick	lineout	0.09
SA	2.08	2.07	lineout	lineout	0.00
NZ	2.96	2.39	lineout	lineout	0.00
NZ	2.96	2.53	lineout	lineout	0.00
SA	2.61	2.39	lineout	kick	0.22
SA	1.14	1.31	lineout	kick	0.17
SA	2.90	2.84	lineout	lineout	0.00

Summary metrics of decision quality for this match are reported in Table 8.

Table 8: Summary metrics of decision quality for the case-study match

Metric	Value
Total regret R	1.39
Proportion of Optimal Decisions	0.46

As summarized in Table 8, the model’s optimal EP decision was made 46% of the time, leading to a total regret of 1.39 points in expectation between both teams. While the proportion of “incorrect” decisions seems relatively high, the expected points lost from sub-optimal decision making is modest. For context, the game ended 24–17 to New Zealand, so 1.39 points would not have swung the outcome. Given that these international teams are at the pinnacle of the sport, the observed level of inefficiency is perhaps unsurprising: most decisions are near the indifference frontier, where the cost of making the “wrong” choice is relatively small.

8 Discussion

8.1 Conclusions

This paper develops a decision-ready Expected Points framework for evaluating whether to kick for goal or kick to touch following a penalty in rugby union. By combining a lineout-based EP model from phase-level Premiership data with an angle-distance model of penalty kick success and a continuation value for missed kicks, we obtain two context-aware EP surfaces and a derived decision surface that identifies where each option maximizes expected return.

The resulting decision maps make the penalty decision explicit as a function of field location, expected meters gained to touch, player advantages, and team-strength differentials. Scenario analyses show how the indifference frontier shifts with yellow cards and team quality, while the case study of New Zealand vs. South Africa demonstrates how the framework can be used to evaluate both individual choices and aggregate decision quality through a regret measure. In that match, the model suggests that sub-optimal decisions were relatively frequent but typically near the decision boundary, so that total regret in expected points was modest relative to the final score margin.

Because the framework is modular, it can be adapted to different competitions or tailored to specific teams by refitting the underlying EP components with alternative data sources. The lineout and kicking modules can be updated independently as richer data become available, providing a transparent foundation for decision-support tools in both professional and amateur settings.

8.2 Limitations and Future Directions

Several assumptions and data limitations should be acknowledged, along with concrete directions for addressing them in future work.

First, the model combines phase-level data from club competitions with penalty-kicking data from international matches, introducing potential inconsistencies in player quality, decision-making, and context. Future work could develop fully competition-specific models by collecting phase-level and kicking data from the same league and year, or by fitting hierarchical models that allow competition-level differences in both lineout and kicking performance to be estimated explicitly.

Second, the analysis restricts team choices to kicking for goal or kicking to touch, even though teams occasionally opt for a scrum or a quick tap—particularly near the try line—albeit infrequently. Extending the framework to additional decision branches is natural: EP models can be trained for scrums and tap-and-go restarts using the same phase-level machinery, enabling multi-armed decision maps that compare all realistic options from a given penalty location.

Third, the analysis is limited to a single season, which constrains the sample size and reduces the generalizability of the findings across competitions or years. Future work could pool multiple seasons and competitions, possibly with random effects for season and league, to

increase sample size, stabilize estimates in rare field zones, and quantify between-competition variation in EP surfaces.

Fourth, selection bias may influence the results (Brill et al., 2025). Stronger teams are both more likely to generate attacking lineouts in the opposition half and more likely to earn penalties. As a result, lineouts in advanced field positions are disproportionately taken by stronger sides, potentially inflating expected points estimates for lineouts near the try line. A similar dynamic is present for the kicking data, since stronger sides will have generated more of the observed penalties, and coaches may only elect to kick when they deem the attempt makeable for their kicker. Future research could focus on further de-biasing EP estimates with respect to the underlying decision processes that generated the data.

Fifth, the data do not track individual kicker ability when aiming for touch to gain territory, making it necessary to assume a translation distance between the penalty location and the resulting lineout. Future analyses trained on proprietary tracking or event data could incorporate player- and team-specific models for both goal-kicking accuracy and meters gained to touch, producing personalized decision maps that reflect the strengths of particular kickers and lineout units rather than competition averages.

Sixth, the current datasets lack the granularity of precise (x, y) field coordinates for penalty locations and lineouts, relying instead on zonal encodings and an approximate mapping to continuous coordinates. Improvements in data collection—for example, through optical or GPS tracking—would allow EP surfaces to be estimated directly on a fine spatial grid, supporting smoother decision boundaries, more accurate modeling of angle-dependent effects, and better treatment of corner and touchline situations.

Lastly, this modeling framework optimizes expected points. A common critique of expected-points-based decision models is that they maximize expected points rather than expected win probability (the quantity that ultimately matters in a game). A natural extension is to develop win-probability models that incorporate field location, score differential, time remaining, and player imbalances resulting from yellow or red cards. Decisions could then be evaluated directly on the scale of win probability, with EP retained as an interpretable intermediate quantity.

8.3 Reproducibility

All code used to process the data, fit the models, and generate the figures and tables in this article is available on [GitHub](#).

References

- Baldwin, B. (2021). Nfl win probability from scratch using xgboost in r. Open Source Football Blog.
- Brill, R. S., Yurko, R., and Wyner, A. J. (2025). Analytics, have some humility: A statistical view of fourth-down decision making. *The American Statistician*, 79(3):393–409.

- Burke, B. (2009). The 4th down study – part 3. Advanced Football Analytics (formerly Advanced NFL Stats), published September 16, 2009.
- Lock, D. and Nettleton, D. (2014). Using random forests to estimate win probability before each play of an nfl game. *Journal of Quantitative Analysis in Sports*, 10(2):197–205.
- Martinez-Arastey, G., Datson, N., Smith, N., and Robins, M. (2025). Foundations of expected points in rugby union: A methodological approach. *Journal of Sports Analytics*, 11(0):1–14. Creative Commons Attribution 4.0 License.
- Monpezat, B. (2023). Predicting penalties and conversions success: Rugby world cup 2023 kicks probability insight. Data Ruck Blog.
- Quarrie, K. L. and Hopkins, W. G. (2015). Evaluation of goal kicking performance in international rugby union matches. *Journal of Science and Medicine in Sport*, 18(2):195–198.
- Romer, D. (2006). Do firms maximize? evidence from professional football. *Journal of Political Economy*, 114(2):340–365.
- Yurko, R., Ventura, S., and Horowitz, M. (2018). nflwar: A reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*. arXiv preprint arXiv:1802.00998v2.