

A Machine Learning Approach to Player Value and Decision Making in Professional Ultimate Frisbee

Other Sports

20251407

1. Introduction

In the past decade, Ultimate Frisbee – commonly known as ‘ultimate’ – has transformed from a largely amateur sport to a professional arena with dedicated athletes and multiple leagues including the Ultimate Frisbee Association, the Premier Ultimate League, and the Western Ultimate League. Unlike established professional sports with sophisticated analytical frameworks like baseball's sabermetrics or football's Next Gen Stats, ultimate has historically relied on basic counting statistics such as goals, assists, and blocks, with analysis often limited to post hoc volunteer-tracked metrics. The emergence of professional leagues has been pivotal in driving more thorough data collection, with new tracking systems now capturing unprecedented detail – recording aspects of every throw, including thrower and receiver location, throw outcome, and game time. Despite these advancements, analytics in ultimate are still underdeveloped, leaving room for more refined methods to assess player contributions and team strategy.

This paper presents the first empirical work to leverage this wealth of new data, specifically the four full seasons of spatial data provided by the Ultimate Frisbee Association, to introduce a machine learning framework for estimating player value and decision quality. We aim to capture more intricate aspects of gameplay – going beyond traditional metrics to provide a more accurate assessment of each throw's contribution to team success. We use two primary models: 1) a Completion Probability (CP) model, which estimates the probability of a throw being successfully completed at any given target location, and 2) a Field Value (FV) model, which assigns positional value on the field as measured by scoring probability. We derive metrics from the CP model, FV model, and their combination that effectively evaluate player contributions, throwing performance, and team strategies. These metrics offer a data-driven approach to analyzing ultimate, establishing a more comprehensive understanding of the game.

In this paper, we accomplish the following:

Section 2: Provide an overview of ultimate, including the sport's background, our dataset, and previous work.

Section 3: Introduce and develop two key models: a CP model to estimate the probability of a throw being completed, and a FV model to quantify the probability of scoring.

Section 4: Leverage these models to create novel metrics, focusing on player contribution, throwing performance, and team strategy.

Section 5: Evaluate the metrics to analyze their effectiveness, interpret what they measure, and assess their overall utility.

Section 6: Demonstrate the practical application of these metrics in real-game scenarios, highlighting their ability to identify multiple MVP players, quantify thrower ability in the context of decision-making, and uncover overlooked players who provide significant value to their team.

Section 7: Discuss the limitations of our study, avenues for future work and conclude.

1.1 Dataset

Our analysis leverages a dataset provided by the Ultimate Frisbee Association, covering four full seasons of professional play from 2021 to 2024. The dataset includes 327,179 recorded throws across 604 games. This comprehensive collection is the largest of its kind in ultimate, offering a novel opportunity to explore throw-level metrics at an unprecedented scale. Key aspects of the dataset include its size and scope – covering all teams in the league and representing professional-level play across four years – and its rich spatial and contextual features. Each entry provides precise field locations for both throwers and receivers, enabling detailed spatial analysis, while game-state variables such as the score differential and remaining time add important context for evaluating decision quality. A typical scoring possession involves a series of passes that advance the disc down the field, culminating in a catch in the end zone to complete the goal (Figure 1).

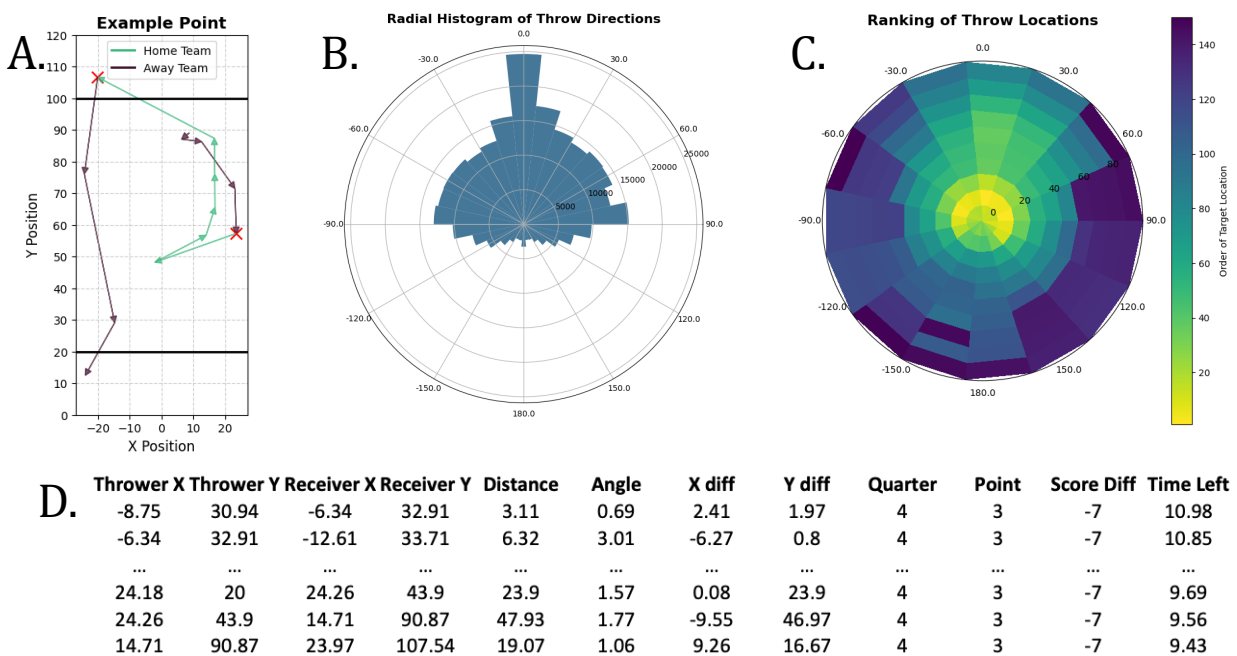


Figure 1: Dataset Description (A) Example point illustrating 2 turnovers and a goal (B) Radial histogram displaying the most frequently used angles by throwers (C) Radial plot showing the most targeted locations relative to the thrower's position (D) Example data

2. Background

We provide context for our research by outlining the basic rules of the sport and reviewing previous work in ultimate analytics.

2.1. Ultimate Frisbee Rules

Our data is based on the gameplay format of the Ultimate Frisbee Association, the premier men's professional league. In a typical match, two teams of seven players compete on a field that measures 80 yards in length and 53 $\frac{1}{3}$ yards in width, with 20-yard end zones on either side. Players can pass the disc in any direction and must establish a pivot foot after receiving a pass. The player with the disc must release it within a 7-second stall count kept track of by game officials.

Each game is divided into four 12-minute quarters, with teams alternating who receives the disc first at the beginning of each quarter. The game follows a "continuous play" format, where, apart from goals, fouls, or timeouts, the game rarely stops, maintaining a fast-paced flow. Possession changes, known as "turnovers", occur through either an incomplete pass (the disc is dropped, lands out of bounds, or is intercepted) or a stall (if the thrower does not release the disc before the 7-second count). When a turnover occurs, the opposing team takes immediate possession at the spot of the turnover and attempts to move the disc in the opposite direction to score. A goal occurs when a team completes a pass into the opposing team's end zone. This earns one point, after which the scoring team initiates a "pull" (similar to a kickoff in football) to begin the next point. If a game ends in a tie at the end of regulation, a 5-minute overtime period is played, with the second overtime being a sudden death scenario where the next team to score wins.

2.2 Previous Work

Published research on ultimate analytics remains limited, with one of the most notable contributions being Weiss and Childers' "Maps for Reasoning in Ultimate" [1] which laid important groundwork for understanding player performance through advanced metrics. Their study introduced innovative methodologies, such as completion and scoring maps, and provided initial insights into the value of throws in relation to expected scoring outcomes. Weiss et al. manually collected data on 3,195 throws from 10 exhibition club ultimate games, utilizing a smoothed K Nearest Neighbors model based on thrower and receiver coordinates. This work led to the creation of metrics like Expected Point Outcome and Effective Contribution to help quantify the expected value of a throw and the impact of a player's actions [2]. While their study was an important step forward, it was limited by the relatively small dataset, data collected on amateur rather than professional play, and a focus on thrower and receiver positions, without incorporating broader game state variables or modeling entire point outcomes.

Our study both provides several novel approaches and improves on this work. We do this by leveraging a significantly larger and more comprehensive dataset. Additionally, we apply more robust modeling techniques, resulting in clearer, more actionable metrics that are designed for practical use in real-game scenarios. Our focus on player value, decision-making, and strategic outcomes offers a more complete understanding of ultimate.

In addition to this Weiss et al., analyses on blogs and web applications have also occasionally explored ultimate statistics with varying degrees of rigor and depth [3,4,5,6]. This gap in formal research highlights the need for further exploration and development of advanced analytics in ultimate. In other sports, analytics have revolutionized strategy, player development, and team performance, providing insights that have significantly advanced the understanding of the game such as Carter et al.'s work which laid the foundation for the widely impactful concept of Expected Points Added in football – a cornerstone of modern football analytics [7]. Prior to the Ultimate Frisbee Association's in-depth data collection starting in 2021, similar progress in ultimate analytics was nearly impossible. With the current dataset, we have the opportunity to create a more refined framework that defines and quantifies offensive contribution and player performance in a way that surpasses traditional counting statistics.

3. Models

We introduce two key models: CP and FV. These models leverage play-by-play data and game state to predict the result of a given throw and the result of a point, respectively. The following sections detail the structure, implementation, and validation of these models.

3.1 Features

The features used in the models are derived from the thrower, receiver, throw, and game contexts.

- **Thrower Context:**
 - Thrower X and Thrower Y: The cartesian coordinates of the thrower's position.
- **Receiver Context:**
 - Receiver X and Receiver Y: The cartesian coordinates of the receiver's position.
- **Throw Context:**
 - Throw distance: The distance between the thrower and the receiver.
 - Throw angle: The throw angle, measured in degrees, with 0° forward, 180° backward, positive to the right, and negative to the left of the thrower.
 - Y differential: The difference in the y-coordinates of the thrower and receiver.
 - X differential: The difference in the x-coordinates of the thrower and receiver.
- **Game Context:**
 - Game quarter: The current game quarter.
 - Quarter point: The point number in the current quarter.
 - Score differential: The point difference between the teams.
 - Time: The time remaining in the current quarter.

3.2 Model Definitions

To evaluate strategic decision-making in the context of gameplay, we define two complementary probabilistic models: FV and CP.

3.2.1 Field Value Model

The FV model predicts the probability that a point culminates in a goal based on the thrower and game context. Mathematically, we model this as:

$$Goal_p(X^t) \sim \text{Bernoulli}(\pi(X_{i(p)}^t, X_{i(p)}^g))$$

Where:

- $\pi(X_{i(p)}^t, X_{i(p)}^g) = P(Goal_p = 1 \mid X_{i(p)}^t, X_{i(p)}^g)$
- $X_{i(p)}^t$ represents the thrower context and $X_{i(p)}^g$ represents the game context
- $i(p)$ denotes the i th pass of point p , with $i = 1, \dots, N_p$ and $p = 1, \dots, N_g$
- N_p is the total number of passes in point p
- N_g is the total number of points in game g

The model allows us to assess the strategic value of disc location continuously over the field. We estimate $\pi(X_{i(p)}^t, X_{i(p)}^g)$ via an ensemble of decision trees, which we fit using XGBoost.

3.2.2 Completion Probability Model

The CP model estimates the probability of a pass being successfully completed. It incorporates information about the thrower context, receiver context, throw context, and game context. Formally, this is expressed as:

$$Complete_i \sim \text{Bernoulli}(\pi(X_i^t, X_i^r, X_i^{th}, X_i^g))$$

Where:

- $\pi(X_i^t, X_i^r, X_i^{th}, X_i^g) = P(Complete_i = 1 \mid X_i^t, X_i^r, X_i^{th}, X_i^g)$
- X_i^t represents the thrower context, X_i^r represents the receiver context, X_i^{th} represents throw context, and X_i^g represents the game context
- i indexes all throws

As with the FV model, the CP model is estimate using an ensemble of decision trees and fitted using XGBoost.

3.3 Performance

We fit our models using 5-fold cross-validation and validate them against three hold-out sets: a player-based set of 50 throwers with 200+ throws each, a temporal set from the latest 10 games to simulate future predictions, and a random 20% subset. This approach ensures robust performance across both temporal and player-specific variations.

Both CP and FV models demonstrate strong performance with high values in Area Under the Curve (AUC), accuracy, positive predictive value (PPV) and negative predictive value (NPV) scores (Table 1). The FV and CP models significantly outperform their respective baselines, which were developed using Weiss and Childers' methodology, across all datasets and measures.

Table 1: CP and FV Model performance compared to baseline.

	Metric	FV Baseline	FV Model	CP Baseline	CP Model
Random	Accuracy	0.708	0.769	0.932	0.946
	AUC	0.622	0.791	0.726	0.906
	PPV	0.709	0.769	0.935	0.947
	NPV	0.512	0.765	0.454	0.905
Temporal	Accuracy	0.694	0.760	0.934	0.948
	AUC	0.629	0.787	0.713	0.900
	PPV	0.696	0.759	0.939	0.949
	NPV	0.446	0.762	0.423	0.903
Player	Accuracy	0.730	0.789	0.936	0.950
	AUC	0.607	0.789	0.755	0.913
	PPV	0.730	0.792	0.941	0.951
	NPV	0.482	0.773	0.461	0.933

4. Metrics

We now introduce the metrics developed from our models to improve the evaluation of player contributions and decision-making. We first introduce metrics derived from the CP model, then metrics derived from the FV model, and finally a metric that integrates both models.

4.1 CP Metrics

The Expected Completion Probability (xCP) is the throw's prediction based on the CP model, defined as:

$$xCP = CP(X^t, X^r, X^{th}, X^g)$$

Where the notation remains consistent with previous definitions. Building on this, Completion Percentage Over Expected (CPOE) quantifies the difference between actual completion rate and xCP .

Let $C = \{0,1\}$ be an indicator for whether the throw was completed (1) or not (0), then CPOE is defined for a specific throw as:

$$CPOE = C - xCP$$

This metric is directly comparable to the CPOE statistic commonly used to evaluate quarterbacks in football [8] and is well-suited for assessing the performance of throwers in ultimate.

4.2 FV Metrics

FV metrics measure the contribution of a throw, considering not just raw yardage gained, but also its impact on the offense's positioning and potential scoring opportunities.

Expected Contribution (EC), originally introduced by Weiss and Childers [2], is redefined here with several improvements including the use of a more comprehensive dataset, the integration of additional game and throw context, and an enhancement of FV to model entire points rather than individual possessions.

Let FV_s be the field value at the starting location (or the location of the thrower), FV_e be the field value at the end of the throw, and FV_o be the opponent's field value at the end location of the throw given a turnover. Other notation is consistent as previously used.

EC is defined for a single throw as:

$$EC = FV_e - FV_s \quad \text{if } C = 1$$

$$EC = -FV_o \quad \text{if } C = 0$$

Further, we introduce an enhancement to the EC metric, which we call Adjusted Expected Contribution (aEC). This adjustment scales credit for scoring events so that the total contribution of all throws in a goal-scoring possession sum to exactly one. In contrast, the EC of goal-scoring possession throws sum to one minus the field value at the start of the possession. By eliminating the dependency on the starting FV, aEC provides a more equitable distribution of contributions across possessions. This also makes the metric more intuitive: for example, a single throw possession that results in a goal will have a contribution of one, directly reflecting the impact of that throw.

aEC is defined for an individual throw as:

$$aEC = \frac{FV_e - FV_s}{1 - FV_p} \quad \text{if } C = 1$$

$$aEC = -FV_o \quad \text{if } C = 0$$

Where notation is consistent as above and FV_p being the FV of the starting throw of that possession.

Because both throwers and receivers contribute to the success of a throw, EC and aEC metrics can be calculate separately for both players. In this paper, we use the metrics Receiver EC (R-EC), Thrower EC (T-EC), and their sum, Total EC (Tot EC), to quantify the individual contributions of each player. The same approach is applied to aEC, where we calculate Receiver aEC (R-aEC), Thrower aEC (T-aEC), and Total aEC (Tot aEC).

4.3 CP and FV Combined Metric

We combine CP and FV to create a metric that captures the strategic impact of a throw in ultimate. Expected Throw Value (ETV) quantifies the expected value of a throw by integrating the probability of success with the potential field value at the throw's endpoint.

ETV for a single throw is defined as:

$$ETV = xCP \times FV_e - (1 - xCP) \times FV_o$$

The positive term ($xCP \times FV_e$) accounts for the potential impact of the throw (FV_e) factoring in its risk (xCP). The negative term ($(1 - xCP) \times FV_o$) reflects the scoring opportunity of the opponent at the same location (FV_o) considering the probability of turnover ($1 - xCP$). This structure enables ETV to provide a comprehensive measure of a throw's strategic value by considering both the potential gain from a successful throw and the risk of conceding a point while accounting for the probability of each outcome occurring.

The ETV ranges from -1 to 1:

- A value of 1 represents a throw with 100% completion into the end zone, resulting in a certain goal.
- A value of -1 indicates the opposite: a certain interception in the defending end zone.

5. Metric Analysis

To effectively evaluate our novel metrics, we assess them based on their ability to distinguish between players (discrimination), their consistency over time (stability), their independence from other performance measures, and their relationship with established metrics [9,10]. Discrimination reflects a metric's capacity to differentiate players based on true performance differences, while stability indicates how reliably a metric predicts future performance. Independence evaluates whether a metric provides unique insights without significant overlap with others, and understanding its relation to established metrics helps validate its relevance and contextual significance.

5.1 Discrimination and Stability

We assess how well our novel metrics capture player value and decision-making reliability by evaluating their discrimination and stability against traditional baseline metrics. Discrimination reflects a metric's ability to distinguish between players by measuring how much of the observed variance in performance is due to true differences rather than random noise. Stability evaluates the consistency of a metric in representing a player's performance over time, indicating its reliability for predicting future contributions. Stable metrics reduce the risk of overvaluing context-dependent or short-term fluctuations. We compare our novel metrics with conventional counting statistics (Figure 2). Our analysis shows that the novel metrics perform comparably to traditional metrics in both discrimination and stability, offering a reliable alternative for evaluating player performance.

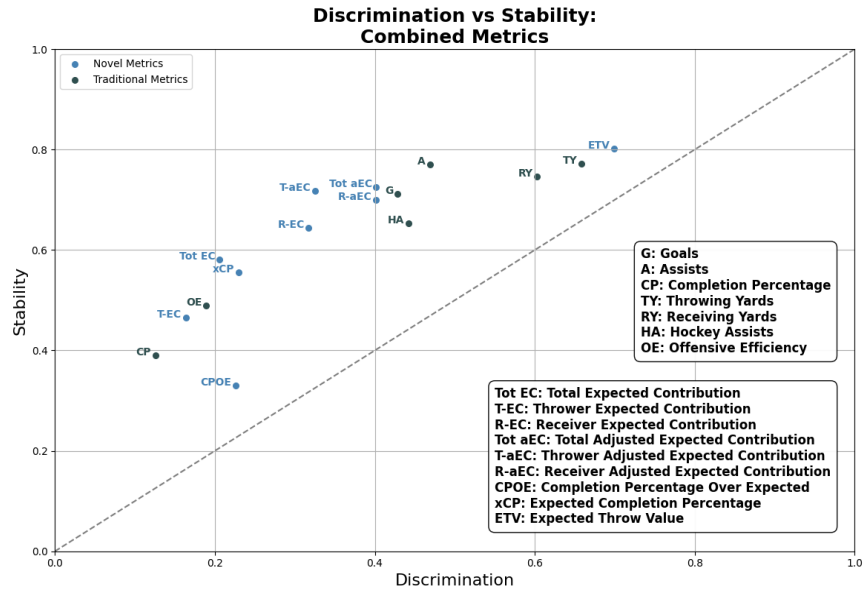


Figure 2: Discrimination vs. stability of ultimate metrics. Discrimination reflects a metric's ability to differentiate player performance, while stability measures consistency over time. Higher values indicate greater reliability and informativeness. Novel metrics (blue), like Expected Throw Value (ETV) and Completion Percentage Over Expected (CPOE), match traditional metrics (black) like goals (G) and assists (A) in both discrimination and stability.

5.2 Independence

Using a Gaussian copula model to measure relationships between metrics quantifies their dependencies, providing a numerical representation of how much each metric offers unique information. This comparison identifies metrics that contribute distinct insights and avoid redundancy. Each metric's independence score is calculated based on its regression on established metrics.

Our analysis reveals that most novel metrics exhibit strong independence, except for ETV (Figure 3). This suggests that ETV primarily reflects overall player involvement and throw frequency, rather than unique insights into throw quality or decision-making. As a result, ETV is more effective for throw selection and optimization of team strategy independent of individual performance. In contrast, adjusted metrics like aEC provide distinct insights with higher independence than EC, while CPOE and xCP isolate specific performance aspects, such as execution quality and shot selection.

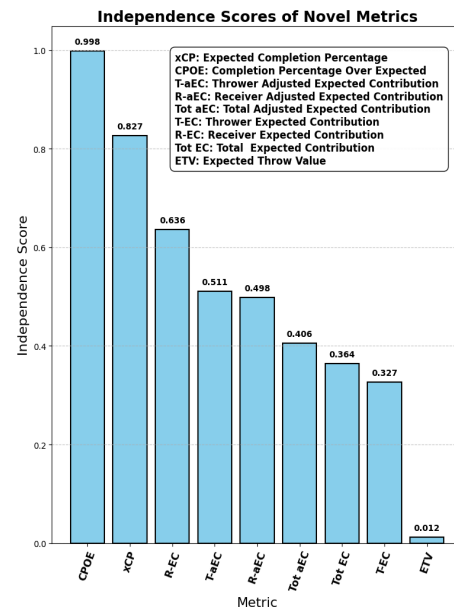


Figure 3: Independence scores of novel against traditional metrics. Higher scores reflect greater independence, indicating that a metric provides unique and non-redundant information for performance analysis.

5.3 Relation to Established Metrics

To assess the relevance of our metrics, we evaluate how closely they align with established metrics. By analyzing the correlation between the novel and traditional metrics, we can determine the specific aspects of player performance captured by the new metrics. We employed hierarchical clustering and used a dendrogram to visualize the relationships between the novel and traditional metrics (Figure 4).

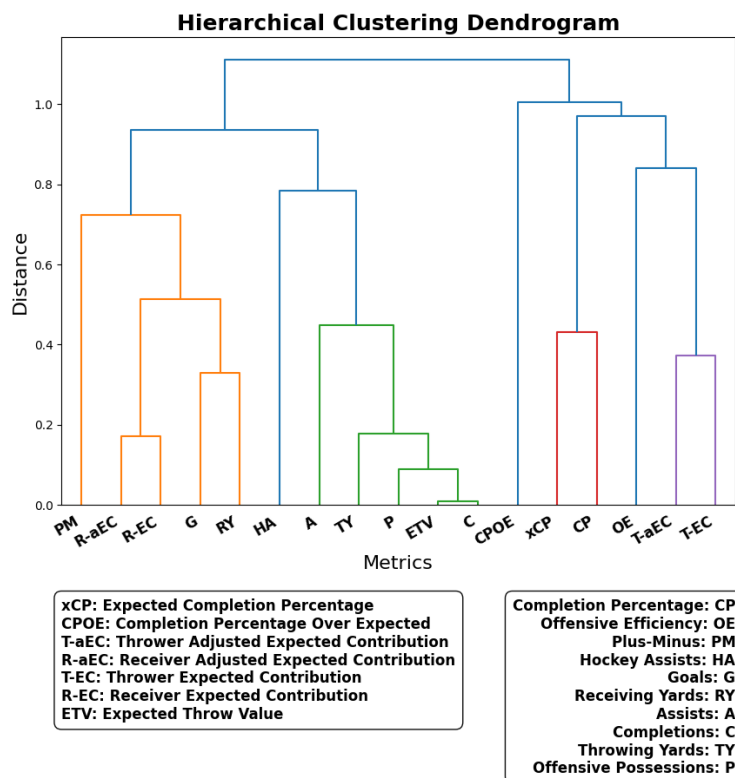


Figure 4: Dendrogram illustrating the relationships among performance metrics in ultimate. By grouping metrics based on similarity, the dendrogram highlights how different metrics capture overlapping or distinct aspects of player and team performance.

The dendrogram reveals several expected relationships, such as R-EC and R-aEC closely aligning with other receiver-based metrics, like goals. In contrast, T-EC and T-aEC are more strongly associated with efficiency metrics, suggesting that a thrower's effectiveness is linked to overall team performance. Meanwhile, ETV clusters near traditional volume metrics, such as the number of throws, highlighting its focus on distribution rather than the quality of individual performance. Lastly, CPOE and xCP show the strongest, yet still modest, correlations with completion percentage and offensive efficiency metrics, indicating that while they capture some related aspects of performance, they provide distinct insights into throw execution and shot selection.

6. Applications

The insights gained from our analyses translate into practical applications that enhance decision-making, strategy, and player evaluation in ultimate. We offer data-driven methods that can optimize team performance and player assessment.

6.1 Decision-Making

By plotting outcomes for every possible throw, we can visualize the most effective strategies in real-time. This allows teams to make informed decisions based on game state and field positioning. Utilizing Shapley Additive Explanations (SHAP) to identify feature importance, we identify the key factors that influence throw value, offering actionable insights into how specific decisions impact scoring opportunities.

6.1.1 Field Plots

We identify optimal target locations by plotting the smoothed ETV for every possible throw given a disc location and game state (Figure 5). In the middle of the field, ETV suggests advancing the disc by throwing closer to the sidelines. This is consistent with known strategies, as many teams use a 'vertical stack', positioning players in the middle of the field while isolating the edges to create space.

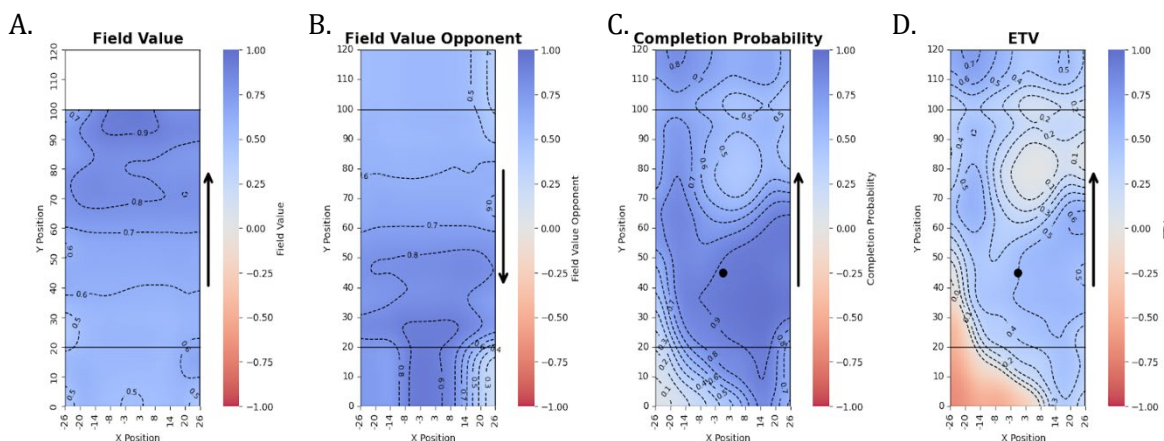


Figure 5: Field plots illustrating different aspects of Expected Throw Value (ETV) with the disc at position (0, 50) in the first quarter, a score differential of 0, and 11 minutes remaining. **(A)** Field value for each receiver location **(B)** Field value of the opponent for each turnover location **(C)** Completion probability every possible throw **(D)** ETV derived from on prior field plots

When the disc is near the sidelines, it's often more effective to 'reset' the disc into the center of the field rather than attempting to advance it toward the end zone. This also aligns with recognized principles in ultimate, where coaches emphasize resetting the disc as a strategic option rather than forcing a low percentage throw down the sideline (Figure 6).

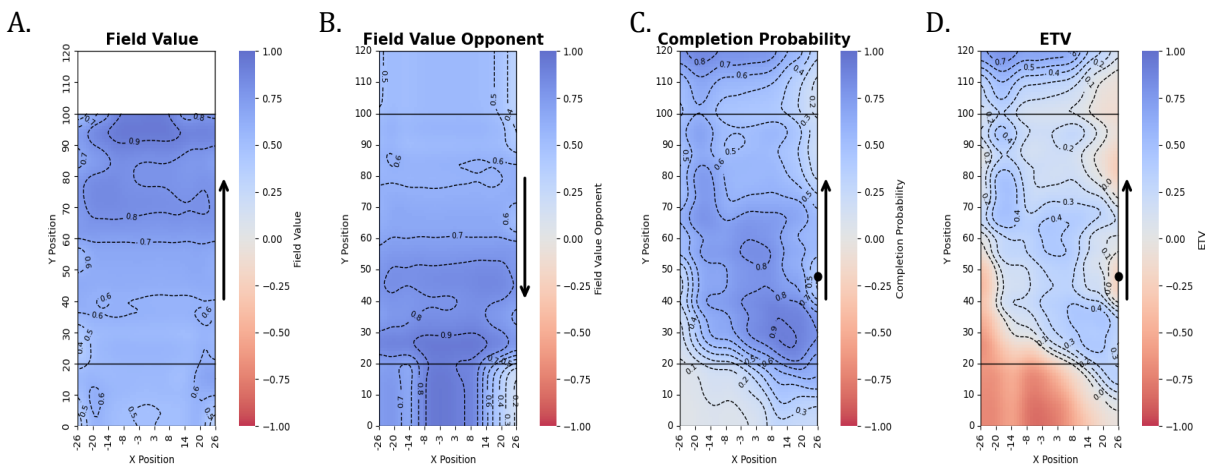


Figure 6: Field plots illustrating different aspects of Expected Throw Value (ETV) with the disc at position (27, 50) in the first quarter, a score differential of 0, and 11 minutes remaining. **(A)** Field value for each receiver location **(B)** Field value of the opponent for each turnover location **(C)** Completion probability every possible throw **(D)** ETV derived from on prior field plots

These findings enable data-driven strategies to improve player positioning, real-time decision-making, and scoring efficiency. They can aid training by emphasizing situational awareness and optimal positions for yardage gaining vs disc resets, encouraging a more efficient offense.

6.1.2 Feature Trends

SHAP provides a method for interpreting the contributions of individual features to a model's predictions, quantifying how each feature impacts the final output. In the context of the ETV metric, SHAP allows for a detailed analysis of how different factors, such as player position and throw difficulty, influence throw value. This interpretability reveals which variables most affect ETV and how they interact. In these plots, each point corresponds to a single throw in our dataset, with the horizontal axis representing the feature value and the vertical axis indicating the SHAP value, which quantifies the contribution of that feature to the model's output.

We identify trends in features such as throw distance and y differential which show a positive correlation with ETV up to a threshold, after which further increases in distance yield diminishing returns, reflecting the potential decline in throw accuracy or higher turnover risk. Additionally, plots examining time remaining in the quarter reveal the effect of player fatigue, with a marked decrease in ETV as time elapses. These plots provide a precise understanding of how different features influence scoring probability and throw completion (Figure 7).

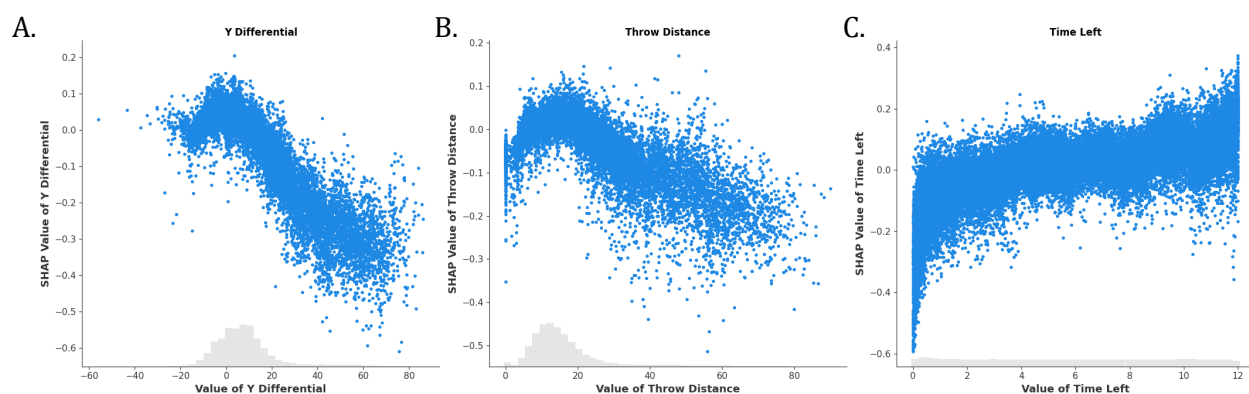


Figure 7: SHAP interaction plots visualize the marginal contribution of features to Expected Throw Value (ETV). Each plot shows how the SHAP value (impact on final prediction) of a feature varies as the value of that feature. Positive SHAP values indicate a positive impact on the prediction, while negative values indicate a negative impact. **(A)** Y Differential **(B)** Throw Distance **(C)** Time Left

6.1.3 Feature Interactions

Coloring SHAP value scatter plots provides an insightful way to visualize feature interactions and their collective impact on model predictions. By incorporating color based on other features, we can better understand how different factors interact to influence ETV.

When examining the SHAP scatter plot of the thrower's x-position, the color gradient reveals how various features such as throw distance and field position interact. These scatter plots highlight a clear advantage to central positioning on the field (Figure 8). Specifically, when plotting the thrower's x-position and coloring by other features, we observe that both long-yardage throws and throws toward the end zone result in higher ETV when made from the middle of the field, compared to low-yardage throws or throws not directed towards the end zone. In contrast, near the sidelines, lower yardage throws and those not targeting the endzone tend to have higher ETV. This illustrates and quantifies the strategic benefit of maintaining central positioning, where throwers have more options and greater control over play development, allowing for aggressive, yardage-gaining throws. Additionally, the model suggests that throwing to the opposite side of the field often provides increased value, as ETV rises when both the thrower and receiver are positioned at opposite extremes of the field. This reinforces the idea that central positioning enhances flexibility and supports forward-moving, high-value plays, while sideline positioning restricts options, making lateral throws more favorable.

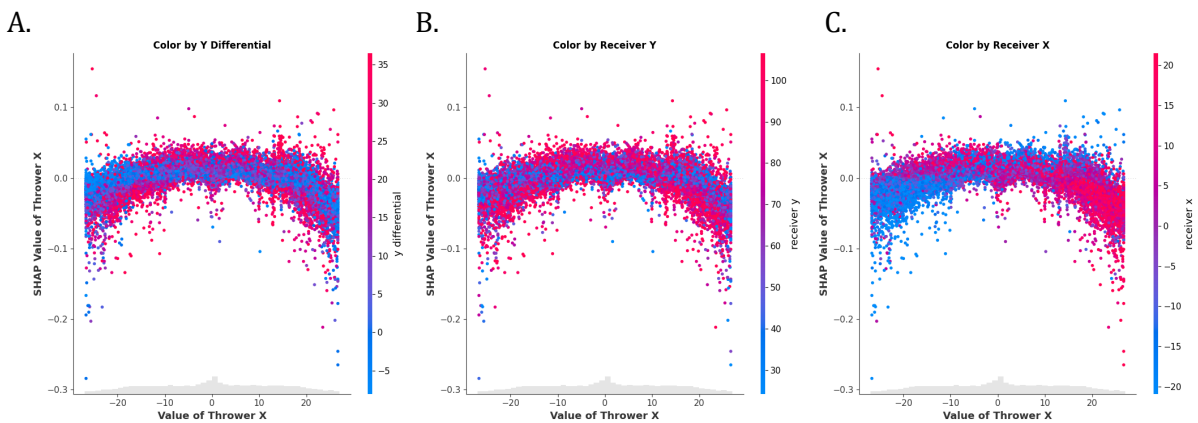


Figure 8: SHAP interaction plots visualize the marginal contribution of features to Expected Throw Value (ETV). The color gradient represents the value of an interacting feature, with warmer colors indicating higher values and cooler colors indicating lower values. Each plot shows how the SHAP value (impact on final prediction) of a feature varies as the value of that feature and an interacting feature change. Positive SHAP values indicate a positive impact on the prediction, while negative values indicate a negative impact. **(A)** Y Differential **(B)** Receiver Y **(C)** Receiver X

6.2 Player Assessment

6.2.1 Play Contribution

The EC and aEC metrics substantially improve upon traditional yardage-based statistics, such as throwing yards and receiving yards, by incorporating the positional value differences in throws. EC provides a more comprehensive assessment of the value generated by factoring in contextual elements like field position and situational impact. aEC goes a step further by scaling contributions so that each goal-scoring possession has the same total value, resulting in a more interpretable and context-independent measure.

Empirical validation underscores the effectiveness of these metrics, with the top five players by total EC (Table 2) include 2021 MVP Ben Jagt and four other All-UFA team selections from their respective seasons. Notably, when examined by season, the MVP ranked #1 in EC in 2021 and 2022, #7 in 2023 and #4 in 2024, reflecting a strong alignment between EC rankings and recognized elite performance. Similarly, aEC also proves valuable in identifying top players (Table 3), including those with unique play styles such as high-risk, high-reward play of 2017 MVP Jonathan Nethercutt and 5 time All-UFA selection Pawel Janas.

Table 2: Top 5 player-seasons by Expected Contribution (EC).

Player	Year	Total EC
Ryan Osgar	2022	23.26
Quinn Finer	2022	18.91
Jordan Kerr	2023	17.48
Ben Jagt	2021	16.35
Kyle Henke	2022	14.63

Table 3: Top 5 player-seasons by Adjusted Expected Contribution (aEC).

Player	Year	Total aEC
Ryan Osgar	2022	146.11
Jonathan Nethercutt	2022	116.13
Ben Jagt	2021	114.91
Pawel Janas	2023	112.69
Ross Barker	2022	107.32

Table 4: Expected Completion Probability (xCP) and Completion Percentage Over Expected (CPOE) for example players in the 2024 season.

Player	xCP	CPOE
Gus Norrbom	96.8%	+1.6%
Matt Jackson	96.1%	+1.9%
Jonathan Nethercutt	93.7%	+1.1%
AJ Merriman	89.7%	+1.3%
James Pollard	86.3%	-0.6%

In addition to identifying top players, coaches and players can evaluate how player roles and contributions interact by directly comparing T-EC to receiving R-EC. For instance, players with high R-EC but low T-EC emphasize their strength as a receiver (known as ‘cutters’), while players with higher T-EC demonstrate their effectiveness as a distributor (known as ‘handlers’). Another possible application is to analyze contributions on a per-possession basis. Coaches can gain a clearer picture of which players make the most significant impact when on the field, enabling data-driven decisions regarding playing time and roster composition. For example, players with a high number of possessions but low contributions per possession should be played less than those with higher contribution per possession with fewer opportunities.

6.2.2 Completion Evaluation

CPOE and xCP are complementary metrics that provide a nuanced view of throwing performance (Table 4). xCP highlights decision-making tendencies, such as Gus Norrbom’s high xCP reflecting his focus on safe, high-percentage throws, or James Pollard’s and AJ Merriman’s lower xCPs indicating a preference for riskier plays. By examining xCP, we can identify players’ strategic tendencies and roles within their teams, whether as steady distributors or aggressive playmakers.

CPOE, in contrast, evaluates a player’s ability to exceed or fall short of these baseline expectations, emphasizing their execution relative to the difficulty of their choices. Players like Matt Jackson, with both a high xCP and a strong CPOE, excel at reliably completing high-percentage throws. Meanwhile, Jonathan Nethercutt, with a lower xCP maintains a high CPOE demonstrating skill in completing challenging throws. These metrics improve upon conventional completion percentage by accounting for the context and complexity of throws, moving beyond raw outcomes to evaluate decision-making and execution. Together, CPOE and xCP paint a holistic picture of player performance. Coaches could use xCP to tailor offensive strategies, assigning high xCP players to handle key reset throws in high-pressure situations, ensuring possession retention. Similarly, players with low xCP and strong CPOE could be prioritized for initiating deep plays or critical end-zone throws where precision in challenging situations is important.

7. Discussion

7.1 Limitations

We address several limitations of our study and data including the inability to accurately attribute credit between throwers and receivers, as both contribute to the success of a play, potentially leading to discrepancies in player evaluations. Additionally, differences in rules, timing, and field dimensions between the Ultimate Frisbee Association, the Premier Ultimate League, and the Western Ultimate League could affect performance metrics, requiring adjustments and league-specific datasets. There is also a potential bias introduced by the location of completed and incomplete throws, as completed throws often land in different areas than incompletions, especially when they go out of bounds or are intercepted.

7.2 Future Work

Future research in the field of ultimate analytics presents several opportunities to refine and extend the application of our novel metrics. A key area for development is the inclusion of additional contextual data, such as stall count, teammate and defender locations, and other situational factors that are not currently captured systematically. These elements likely influence player decisions and outcomes and incorporating them would enhance the robustness of these metrics. Expanding the dataset to include these overlooked factors would also improve the predictive power and relevance of these metrics in various in-game contexts. Additionally, a significant avenue for future work lies in the development of defensive metrics, which remain underexplored. Integrating data on defensive positioning, pressure, and forced turnovers could offer a more balanced evaluation of player contributions, highlighting not only offensive strengths but also defensive impact.

7.3 Conclusion

In this paper, we introduce novel models and metrics learning that significantly enhances the analysis of player value and decision-making in ultimate. By leveraging a comprehensive dataset and modeling techniques, we develop two models to predict completion and score probability from thrower, receiver, throw, and game contexts. We use these models to create metrics that evaluate player contribution (EC and aEC), completion rates (CPOE and xCP), and guide overall strategy (ETV). These metrics offer a more nuanced and informative understanding of ultimate, surpassing the limitations of traditional statistics. Our findings demonstrate the reliability, stability, and unique insights provided by these metrics, making them valuable tools for coaches, analysts, and fans alike. The applications of these metrics are extensive, shaping strategy, player roles, playing time, quantifying impact, and more. By incorporating these metrics, teams can refine their strategies to improve both individual and team performance.

References

- [1] Weiss, J. C., & Childers, S. (2013). "Maps for reasoning in ultimate." *CEUR Workshop Proceedings*, 1969, 35-46. <https://ceur-ws.org/Vol-1969/paper-04.pdf>
- [2] Weiss, J. C., & Childers, S. (2014). Spatial statistics to evaluate player contribution in ultimate. Poster presentation at the MIT Sloan Sports Analytics Conference. <https://pages.cs.wisc.edu/~jcweiss/ssac2014/MIT%20SSAC%20Poster.pdf>
- [3] Wurtztack, P. (2021). "Better box score metrics: The return & introducing EDGE [Pt. 1]." *Ultiworld*. <https://ultiworld.com/2021/04/01/better-box-score-metrics-the-return-introducing-edge-pt-1/>
- [4] Scott, C. (n.d.). Retrieved from <https://medium.com/@colinscott4>
- [5] Geertsen, I. (2021). Player Efficiency Rating in the AUDL: Developing an Impact Metric for Ultimate Frisbee. *Bruin Sports Analytics*. https://www.bruinsportsanalytics.com/post/ultimate_per
- [6] Schmidt, H. (n.d.). UltiMaps: An ultimate frisbee analytics tool. Retrieved from <https://hirosme.shinyapps.io/UltiMaps/>
- [7] Carter, V., & Machol, R. E. (1971). "Operations research on football." *Operations Research*, 19(2), 541-544. <https://www.jstor.org/stable/169286>
- [8] Yurko, R., Ventura, S., & Horowitz, M. (2019). nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, 15(3), 163-183. <https://doi.org/10.1515/jqas-2018-0010>
- [9] Franks, A., D'Amour, A., Cervone, D., & Bornn, L. (2016). Meta-analytics: tools for understanding the statistical properties of sports metrics. *Journal of Quantitative Analysis in Sports*, 12(4), 151-165. <https://doi.org/10.1515/jqas-2016-0098>
- [10] ElHabr, T. (2023). "Meta-analytics for soccer." <https://tonyelhabr.rbind.io/posts/soccer-meta-analytics/#fnref2>