Modeling Expected Possession Value in the American Ultimate Disc Association (AUDL)

Nate August

Introduction

The American Ultimate Disc Association (AUDL) is a developing semi-professional Ultimate Frisbee (Ultimate) league with teams across the United States and Canada [1]. Ultimate is played on a field comparable to an American Football field. 7 players from each team are present on the field for each point. The team in possession of the disc scores when a teammate receives the disc in the attacking endzone. Players are not allowed to move when in possession of the disc, making throwing the disc between players ("passing") the only method for advancing field position towards the endzone [15].

Recent partnerships with DraftKings [11], a sports-betting platform and LSports [12], a provider of real-time sports analysis indicate that the AUDL is moving towards other highly-digitized and analyzed sports such as soccer, football and basketball. While these mainstream sports have received considerable attention from the data science and sports science communities, published literature on Ultimate has been largely limited to injury risk [8,9,10] and sociology-related disciplines [4,5]. Within the popular media, however, an increasing availability of play-by-play data has spurred considerable discussion around improving both team and player metrics [13, 14] primarily through combining simple aggregates like yards-gained with common statistics like goals, assists and blocks. Analysis of spatiotemporal throw-by-throw data is largely ignored due to the volume (~250 per game) [7] and complexity of analyzing such data. This suggests a ripe opportunity for the data science community to contribute by providing useful representations for and analysis of throw-by-throw data in Ultimate.

This study aims to build on previous work in both Ultimate and other sports to develop characterizations of offensive possessions using spatiotemporal data from the 2021 AUDL season. I propose developing a Expected Possession Value (EPV) model to explore how spatio-temporal features impact offensive success. EPV is analogous to a “stock-ticker” for a given possession where the value summarizes available information into a single value. In Ultimate, where a possession can result in at most one point, EPV represents the probability of scoring. In this study, conditional probability and state-transition modeling approaches are explored as methods for generating EPV estimates in Ultimate.

Related work

In other sports, spatiotemporal data have been explored extensively with notable results. In basketball, player and ball tracking data have been applied extensively to interrogate player and team behavior [19,3]. Cervone et al., propose a framework for estimating expected possession value (EPV) throughout an offensive possession, shedding new light on spatial strategy [19]. EPV has also been explored in rugby, where work was done to investigate optimal schemes for discretizing spatial data into zones [22]. In soccer, pass quality has been evaluated to identify the best passers and receivers as well as locations from which certain teams are most dangerous [16,17]. Other work has focused on characterizing team strategy as a whole using pass data to identify differences in play strategy when home vs. away [18].

The sole available work utilizing spatiotemporal data in Ultimate is authored by Lam et al. [7]. To analyze the spatial attributes, the authors represent offensive possessions with a state transition model. In this formulation, there are four transient states representing zones on the field where a pass originates and absorbing states that note the end a possession, either via scoring or a turnover. The state transition model enabled examination of rates of transitions between states for winning and losing teams. While a useful first step, this study led to very limited practical insights in the spatial patterns of the game, likely due to the coarse spatial resolution used for analysis.

Data and data processing

About the Data

Play-by-play data is publicly-available through audlstats.com, a site maintained by the AUDL. A python package AUDL-Advanced-Stats was used to download and interact with the data [2]. Data from 134 games over the 2021 season were used in the analysis. The dataset contains the origin and destination for each through as well as the thrower and receiver. Figure 1 diagrams the entry and exit states of a possession. To reduce noise, possessions that were ended by a time cap were removed because teams often take riskier throws in order to score before the end of a quarter.

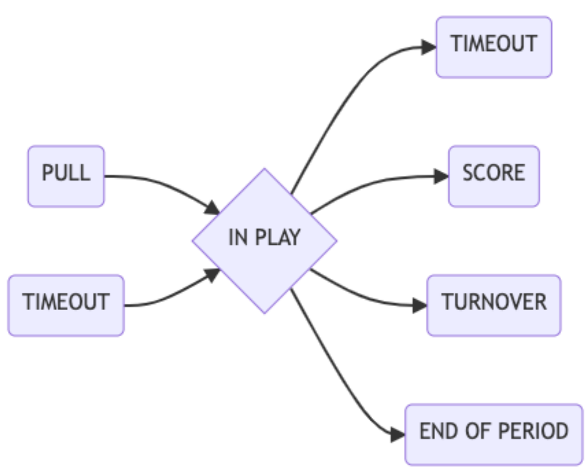

Figure 1. Possession States. A possession is defined as beginning from a pull (analogous to "kick-off" in American Football) or a timeout. During timeouts, teams are allowed to make personnel changes. When in play, the disc is passed between members of the same team. A possession ends when the offensive team calls a timeout, scores, turns the disc over or when the play clock for the period falls to 0.

The final dataset consisted of 71,341 passes from 10,507 possessions. The average completion rate for passes was 92.9% and the percent of possessions resulting in a score was 51.5%. An example of a possession is shown in Figure 2.

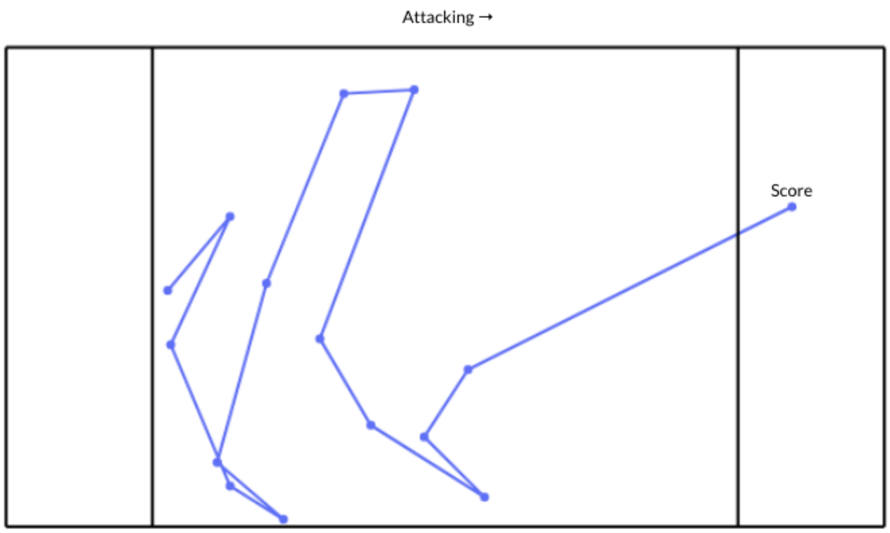

Figure 2. Example possession from 2021 AUDL Division Playoffs. Raleigh Flyers vs. DC Breeze. This is the 5th possession for the Flyers and ends in a score.

Zone Discretization

The origin and destination locations for each throw were discretized into 51 zones. 50 10m X 10m zones represent the primary field of play, including the defending endzone. The attacking endzone (25m x 50m) is represented by a special absorbing zone. The successful reception of a pass within the attacking endzone signals a score and the end of possession. Relative to existing work in Ultimate [7], this zone scheme is higher in resolution by roughly an order of magnitude.

Throw Types

Throw type labels are assigned based on the trajectory classifications developed by AUDL-Advanced-Stats [2]. Classifications are shown in Figure 3.

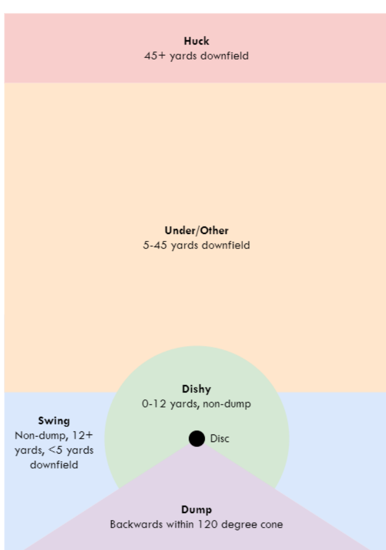

Figure 3. Throw Type Classifications. Throws are assigned one of five labels based on the scheme provided in AUDL-Advanced-Stats [2]. The scheme considers the horizontal and vertical travel of the pass relative to the throwers location. For example, if a throw travels 3 yards vertically (downfield) and 15 yards horizontally that would be labeled as a swing.

Player Characteristics

Anyone that has played or watched sports will share the intuition that different players have different characteristics which may impart valuable information that can be used to improve the accuracy of modeled outcomes. Given that the AUDL dataset contains only disc movements and does not directly track player movements, our consideration of player characteristics will be limited to throwing behavior. Even with this limited scope, there is not enough data on the behavior of each player in each zone of the field to derive useful characteristics when each player is considered individually. This is a challenge that other authors have acknowledged even in domains like basketball where researchers have orders of magnitude more observed possessions than the present case [19].

To address this, players are clustered into thrower archetypes using KMeans++ [20]. Since throwing is the sole behavior of interest, the clustering is performed on features that describe what a player does once they have the disc. Features are normalized to the number of times a player is in possession of the disc to limit the influence of playing time on cluster assignment. Some examples of cluster features include vertical yards per completion, percent huck attempts and completion percent. Principal Component Analysis is used to reduce the dimensionality of the feature space from 19 to 10 principal components. The optimal number of clusters identified by using silhouette coefficients [21] was found to be 2.

Inspection of the clusters shows that the two clusters fit with commonly-held notions of the two primary positions in Ultimate Frisbee: “Cutters” and “Handlers”. Handlers (approximated by cluster `1.0`) are typically stronger throwers and have a bigger role in distributing the disc while cutters (approximated by cluster `0.0`) are primarily receivers and often avoid longer throws. Figure 4. compares the clusters on a selection of features.

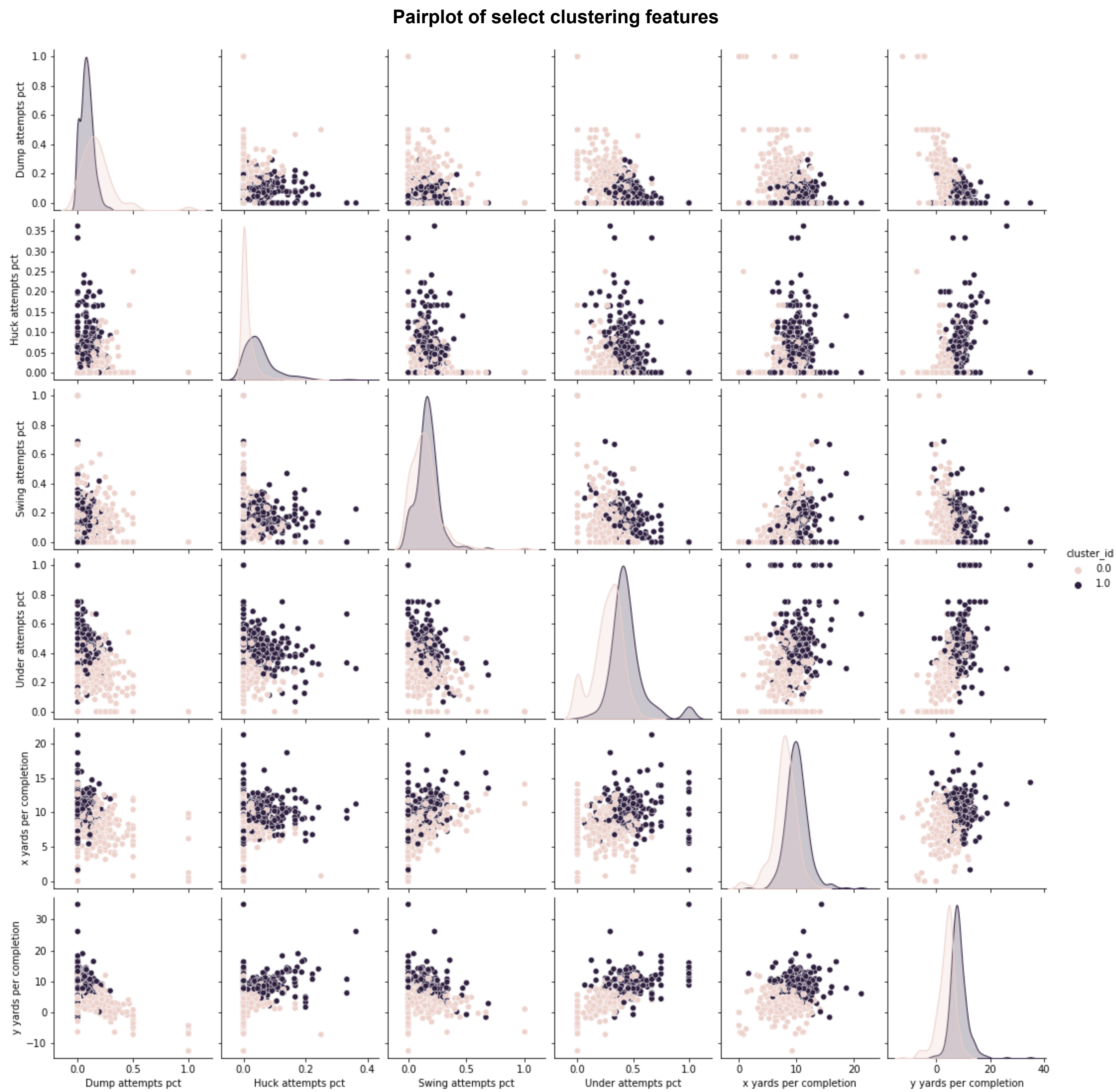

Figure 4. Pairplot of selected features used in player clustering. Players in cluster `0.0` tend to have fewer yards per completion (both horizontal and vertical), make more dump attempts and fewer huck and under attempts. This is consistent with the common notion of the “Cutter” position in Ultimate. Players in cluster `1.0` are the opposite, representing the notion of the “Handler” position.

Spatio-temporal Feature Space

The space of possession histories $\mathcal{F}_t$ for our spatio-temporal models is simplified into four attributes for every timestamp t:

$$\mathcal{F}_t = M \times Z_t \times C_t \times R_{t-1}$$

Where M represents team, Z_t represents the zone at time t, C_t represents the cluster label of the player in possession of the disc at time t and R_{t-1} represents the previous throw type. Previous throw type is included as a potential proxy for the flow of positioning of players on the field, since player location is not directly available in this dataset. The idea is that the opportunities available to a thrower will be dependent on the previous zone because players will be positioned according to the previous zone. While this is an obvious simplification of all possible play histories, the constrained feature space ensures parameter estimation is tractable.

Methods

Expected possession value (EPV) ν_t at time t can be understood as the probability of the possession ending in a score $X = 1$ given the history of the possession up to time $\mathcal{F}_t$:

$$\nu_t = P[X = 1|\mathcal{F}_t]$$

Two approaches to modeling EPV are investigated in this work: conditional probability (CP) and state transition (ST).

Conditional Probability Model (CP)

Conditional probability is the simplest formulation of EPV estimation. Intuitively, it is a count of how many times a possession with history $\mathcal{F}_t$ resulted in a score divided by the total number of times $\mathcal{F}_t$ is observed n.

$$P[X = 1|\mathcal{F}_t] = \frac{\sum_{i=1}^{n} \mathbb{I}(X = 1)}{n}$$

This method presents challenges in constraining the space of possible histories such that a reasonable number of examples for every history exists. It also has the potential to be more reactive to noise than a state-transition model because it is so directly dependent on data. To ensure tractability, four formulations of this conditional probability are considered that leverage different combinations of our feature space. For example, CP_{MZ} is the probability of scoring given team m is in possession at zone z. All four formulations are shown below:

$$CP_M = P[X = 1|M = m]$$
$$CP_{MZ} = P[X = 1|M = m, Z_t = z]$$
$$CP_{MZR} = P[X = 1|M = m, Z_t = z, R_{t-1} = r]$$

$$CP_{MZC} = P[X = 1 | M = m, Z_t = z, C_t = c]$$

Notably, a model using all combinations of the feature space is not calculated because the data is too sparse to provide reasonable probabilities.

State Transition Model (ST)

In the process model, possessions are modeled as a state-transition process where transition probabilities π_{ij} for all discrete states representing $i, j \in Z$ where Z is the space of all combinations of possession history values. In the simplest case, there is a state for every zone on the field, with special absorbing states representing scores or turnovers. A transition π_{ij} represents the probability of the disc moving from state i into state j. A simulation can then be executed to generate sample paths for "playing out" the possession given any starting state by performing a random walk between states until an absorbing state is reached. Our EPV is found by calculating the rate at which sample paths beginning from i at time t eventually terminates in a score $X_\epsilon = 1$:

$$P[X = 1 | \mathcal{F}_t] = P[X_\epsilon = 1 | Z_t = i] = \sum_{k=1}^{n} \mathbb{I}[x_\epsilon = i]$$

The transition probabilities are estimated as counts of all transitions from i into j, divided by all transitions from i.

Two formulations of state spaces were explored in this work. In the first, ST_{MZ} states represent only zones in the field for each team. In the second ST_{MZC} each state represents a combination of team, zone and player cluster.

Software

All analyses and modeling was performed in Python using open source packages NumPy [24], Scipy [26], and Scikit-Learn [25].

Evaluation

Evaluation of EPV presents a challenge because there are no direct measurements of the value; only the end result of a possession can be observed. Previous works evaluating EPV focus on parameter reproducibility in evaluation [19, 22]. The idea is that if the estimated parameters are similar across different datasets for the same team, the model is likely to represent reality in a useful way. Here, we use Kullback-Liebler Divergence D_{KL} to evaluate how similar two distributions of parameters are for our models [23]. In the discrete case the D_{KL} of two distributions P, Q is given by:

$$D_{KL}(P, Q) = \sum_{xinX} P(x) log(\frac{P(x)}{Q(x)})$$

To enable comparison between models with different numbers of parameters, the divergence metric is normalized by dividing by the number of parameters in the model.

In addition to consistency, the EPV estimates are evaluated by how well they predict the outcome of a possession using mean absolute error (MAE). While prediction is not a principal outcome of this work, MAE should provide some indication of how well the models describe the data. MAE is used instead of mean-squared error because the values are probabilities. More common classification metrics are not used to avoid the need for identifying a threshold upon which the possession is assigned to either “Score” or “Turnover”.

During evaluation, the models are fit to some number of games and then tested on other games using cross-validation. Cluster labels are assigned during each iteration using only the training set.

Results

Model Performance

Summarized Model Performance

	Mean Absolute Error		K-L Divergence	
	mean	std	mean	std
CP-M	0.47131	0.02870	**0.00312**	0.00423
CP-MZ	0.44864*	0.02690	0.10393	0.06590
CP-MZC	0.44839*	0.02730	0.20046	0.10914
CP-MZR	0.45573*	0.02615	0.33253	0.06180
ST-MZ	0.45270*	0.02180	0.00694	0.00528
ST-MZC	**0.44168***	0.07555	0.00991	0.00568

Figure 5. Summarized model performance when evaluated with 5 training games and 5 testing games. Mean values noted with * are not statistically different at $p < 0.05$. Statistical significance calculated using Welch’s T. Minimum values shown in bold.

For Normalized K-L Divergence, the naive conditional probability model CP-M scored the best with the state transition models (ST-MZ, ST-MZC) close behind. The other conditional probability models increased in divergence according to their complexity. CP-MZR, with a number of parameters equal to {zones} x {throw types}, showed the greatest divergence between distributions of parameters between train and test sets.

On the prediction metric, all models other than CP-M performed similarly. This is not shocking considering prediction isn’t the core objective. It does however provide a strong signal that the state transition models are not providing completely erroneous results as they show agreement with more simplistic models.

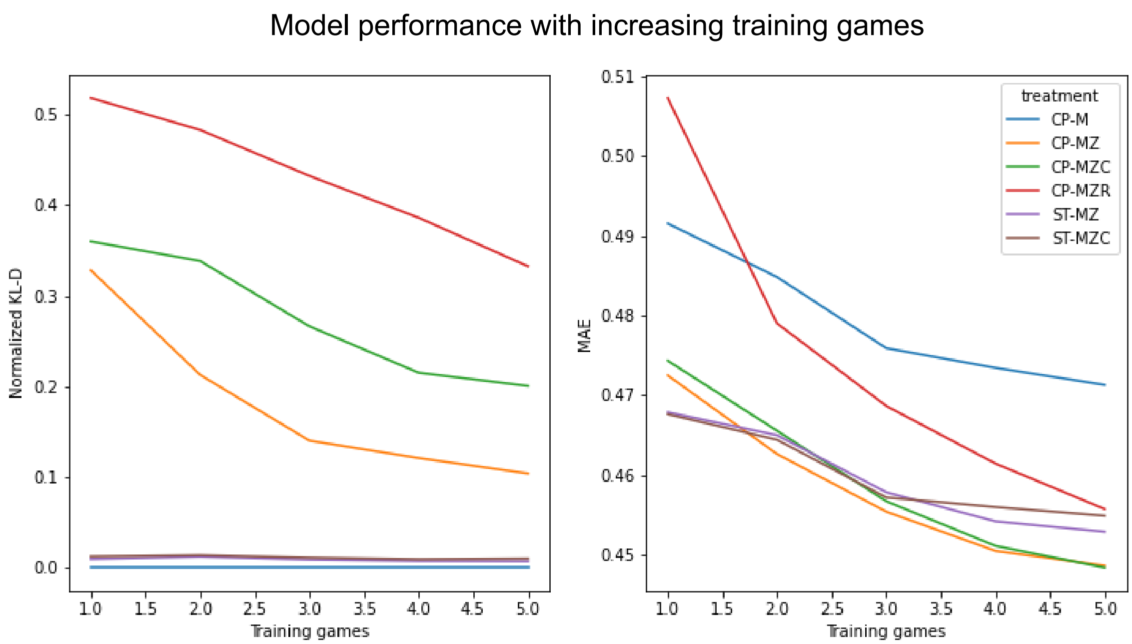

Figure 6. Model Performance with increasing training games. Average cross-validation results from models trained on a number of games and then evaluated on the same number of games.

All models demonstrated increasing performance when exposed to additional training data. CP-M, the most naive model, showed the lowest divergence, but also the worst predictive performance. State models ST-MZ AND ST-MZC performed similarly across both metrics. CP-MZ and CP-MZC had similar trends in divergence and showed the greatest predictive power in the iteration with the most training data. ST-MZR had the worst divergence and poor predictive power when provided with few training examples, but the predictive power improved to be close to the best as training data increased.

EPV Charts

When examining the EPV charts, a few clear trends emerge. EPV increases as the disc is moved closer to the endzone, as exemplified in Figure 7. When the disc is close to the sideline and far from the endzone, EPV is typically lower than when the disc is in the center of the field, however this effect diminishes as the disc moves closer to the endzone. Comparing between modeling approaches, the state-transition model estimates appear more moderate, where the conditional probabilities can appear noisey. For throw number 6 in figure 7, for example, CP-MZR has the scoring probability at 100%. This is obviously erroneous and likely due to very few direct observations of that combination of features. The state-transition models help buffer some of those limitations in the training data.

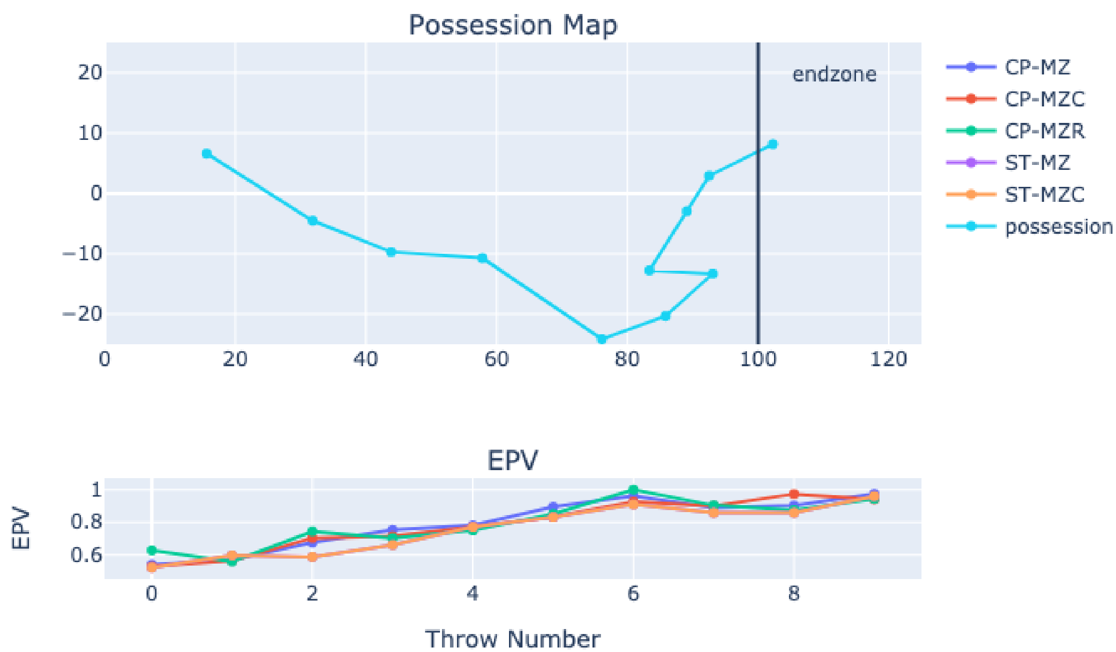

Figure 7. Example EPV. 2021 AUDL Division Playoffs. Raleigh Flyers vs. DC Breeze. Raleigh possession number 8.

State-Transition Model Path Generation

The state-transition models have an added benefit in being able to generate simulated possession paths based on a starting state. The closer these simulation paths are to reality, the better our EPV estimates will be. Figure 8. displays examples of real and simulated possession paths produced from ST-MZ. While the true possessions appear a bit more consistent in their behavior, this small sample of generated paths appears to be plausible.

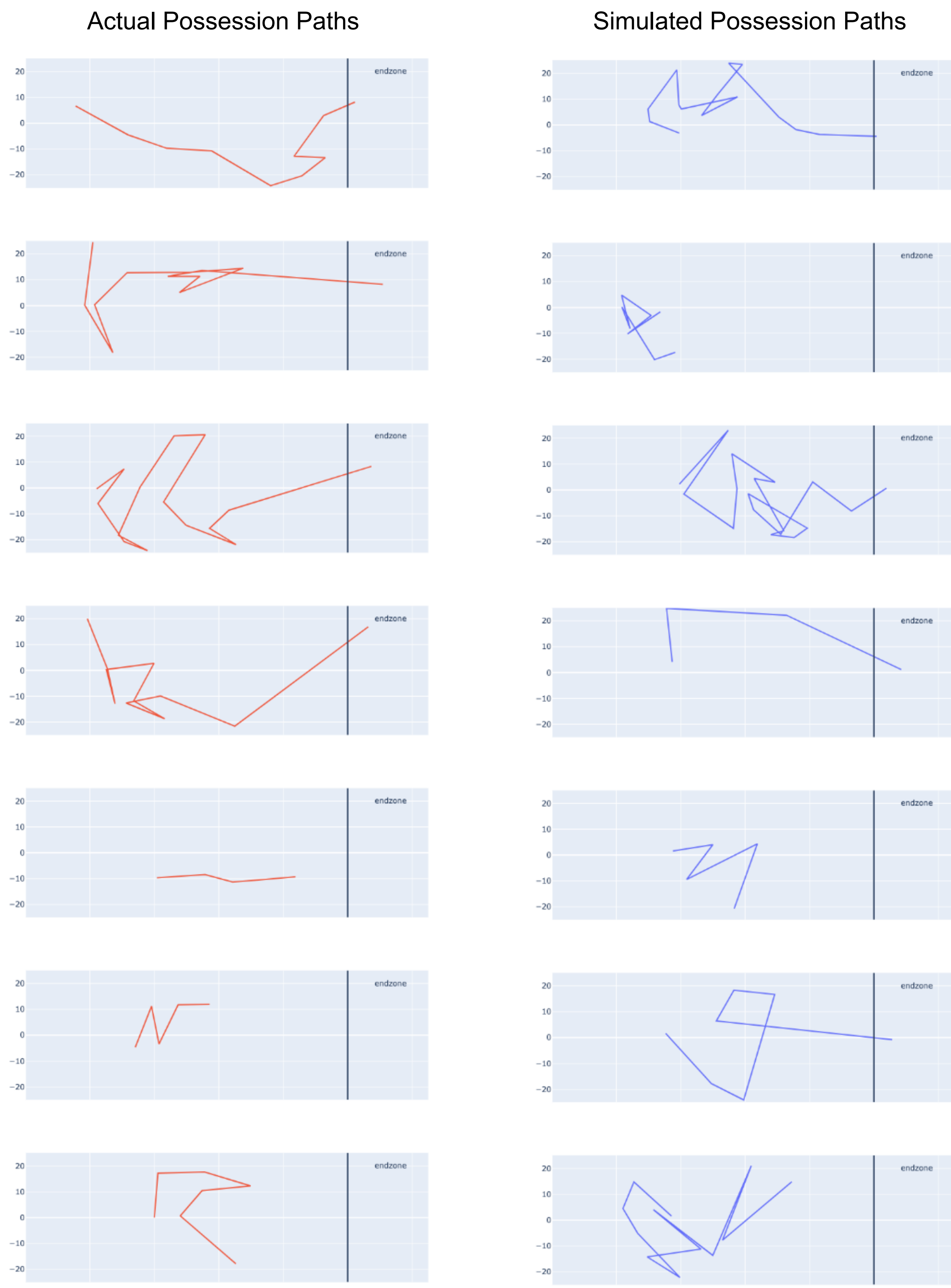

Figure 8. Left: The first 7 possessions from the Raleigh Flyers at the AUDL Divisional Championships. Right: 7 simulated possession paths generated from a ST-MZ model fit to all data on the Raleigh Flyers from 2021.

Conclusion

This work demonstrates that spatio-temporal modeling can generate meaningful EPV estimates in Ultimate Frisbee using a state-transition modeling. We have shown that state-transition models provide greater consistency and are likely better estimators of EPV than simple conditional probability approaches.

Further work on model improvements are open for investigation. Integrating the previous throw feature into the state transition model could be a viable path for improving the descriptive power of these models. Additionally, careful study of the paths generated from the state-transition models could inform further evolution.

Analysis comparing EPVs from different teams could yield useful insights for team development or developing opponent-specific strategies. For this to become viable, better approaches to characterizing and displaying EPVs would be important steps.

References

[1] American Ultimate Disc Association, "About the AUDL," Available: https://theaudl.com/. [Accessed: Oct, 26, 2021]

[2] J. Lithio., "AUDL-Advanced-Stats" Available: https://github.com/JohnLithio/AUDL-Advanced-Stats [Accessed: Oct, 26, 2021]

[3] P. Lucey, A. Bialkowski, P. Carr, Y. Yue and I. Matthews, "'How to Get an Open Shot': Analyzing Team Movement in Basketball using Tracking Data", In *8th Annual MIT Sloan Sports Analytics Conference*, 2014

[4] J. Neville, "Dressed to play: An analysis of gender relations in college women's ultimate Frisbee," *International Review for the Sociology of Sport* vol. 54, no. 1, p.38-62, Feb. 2019

[5] J. Amoroso, J. Coakley, R. Robelo-Goncalves, R. Antunes, J. Valente-dos-Santos, and G. Furtado, "Teamwork, Spirit of the Game and Communication: A Review of Implications from Sociological Constructs for Research and Practice in Ultimate Frisbee Games," *Soc. Sci.* vol. 10, no. 8, p. 300, Aug. 2021

[6] F. R. Goes, M. S. Brink, M. T. Elferink-Gemser, M. Kempe, and K. A.P.M Lemmink, "The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data," *J. Sports Sciences*, vol. 39, no. 5, p. 523-532, 2021

[7] H. Lam, O. Kolbinger, M. Lames and T. G. Russomanno, "State Transition Modeling in Ultimate Frisbee: Adaptation of a Promising Method for Performance Analysis in Invasion Sports," *Front. in Psychology*, vol 21, April 2021

[8] M. C. Hess, D. I. Swedler, C. S. Collins, B. A. Ponce and E. W. Brabston, "Descriptive Epidemiology of Injuries in Professional Ultimate Frisbee Athletes," *J. Athl. Train* vol. 55, no. 2. p. 195, 2020

[9] K. H. Reynolds and S. E. Halsmer, "Injuries from Ultimate Frisbee," *Wisconsin Medical Journal* vol. 105, no. 6, p. 46-49, 2006

[10] D. F. Pulido and R. P. Lystad, "Epidemiology of Injuries in Ultimate (Frisbee): A Systematic Review," *Sports* vol. 8, no. 12, p. 168, Dec. 2020

[11] C. Eisenhood, "DraftKings Now Taking AUDL Bets in Some States," Ultiworld.com, July 22, 2021. [Accessed: Oct, 26, 2021]

[12] American Ultimate Disc League, "American Ultimate Disc League and LSports Data Enter Into $3 Million Data Distribution and Co-Development Strategic Partnership," CISION PR Newswire, April 22, 2021. [Accessed: Oct, 26, 2021]

[13] P. Wurtztack, "Better Box Score Metrics: Fixing the Plus/Minus Stat [Pt. 7]," Ultiworld.com, August 30, 2021. [Accessed: Oct, 26, 2021]

[14] I. Geersten "Player Efficiency Rating in the AUDL: Developing an Impact Metric for Ultimate Frisbee," Bruins Sports Analytics, June 9, 2020. [Accessed: Oct, 26, 2021]

[15] American Ultimate Disc League, "Offical Rules: 2021 Season" Available: https://theaudl.com/sites/default/files/AUDL%20Rule%20Book%20v2021.pdf [Accessed: Oct, 26, 2021]

[16] P. Power, H. Ruiz, X. Wei, and P. Lucey, "'Not All Passes Are Created Equal:' Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data", In KDD ’17, 2017, pp. 1605-1613

[17] F. R. Goes, M. Kempe, L. A. Meerhoff, and K. A.P.M. Lemmink "Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches," *Big Data*, vol. 6, no. 4, 2018

[18] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews, "Assessing Team Strategy using Spatiotemporal Data,", *KDD '13: The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013

[19] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, "A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes", *J. Amer. Stat. Assoc.* vol. 111, no. 514, p. 585-599, 2016

[20] D. Arthur, S. Vassilvitskii, “k-means++: The Advantages of Careful Seeding”. Stanford. 2006. Available: http://ilpubs.stanford.edu:8090/778/1/2006-13.pdf

[21] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”, *J. Computation and Applied Mathematics,* vol. 20, p. 53-65, 1987

[22] T. Sawczuk, A. Palczewska, and B. Jones, “Development of an expected possession value model to analyse team attacking performances in rugby league.” *PLoS ONE* vol. 16 no. 11. 2021. Available https://arxiv.org/pdf/2105.05303.pdf

[23] S. Kullback and R. A. Leibler, “On Information and Sufficiency”, *Ann. Math. Statist.* vol. 22, no. 1, pp. 79-86, 1951

[24] C. R. Harris, K. J. Millman, S. J. van der Walt, “Array programming with NumPy” *Nature* vol. 585, pp. 357–362

[25] F. Pedregos et. al., “Scikit-learn: Machine Learning in Python” *JMLR* vol. 12, pp. 2825-2830, 2011

[26] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. *Nature Methods*, vol. 17, pp. 261–272, 2020