Adjusting Double Poisson Models to Predict the NCAA Division I Softball Championship

Authors' Names Redacted

September 2024

Abstract

While its viewership has surged in recent years, college softball remains an under-researched sport in the domain of sport analytics, partially due to a lack of a longstanding major professional league. However, the postseason format of major college softball - a four-stage layout with two four-team double elimination phases and two best-of-three series - presents an intriguing challenge for predictive models. Primarily focusing on the first of these four stages, we evaluate the effectiveness of a Double Poisson model in predicting the outcome of this competition. This model, previously developed and advanced for use in predicting the outcome of soccer matches, posits that a team's number of runs scored takes a Poisson distribution, with a mean based on its own offensive strength and its opponent's defensive strength, with strengths expressed in terms of scoring averages. Using game-by-game results for runs scored and runs against, we construct two additional pairs of factors used in constructing the means of these variables for each team that account for a team's strength of schedule and conference membership. From these additional pairs of factors, we construct three additional models: one that accounts for opponent quality, one that accounts for conference membership, and one that considers both components. We assess each of these models in their ability to predict recent editions of the NCAA Division I Softball Tournament, with the aim of better understanding factors associated with success in this compelling tournament format.

1 Introduction

The eight team Women's College World Series (WCWS), held in Oklahoma City, is the culmination of each edition of the NCAA Division I Softball Tournament. In recent years, both in-person attendance and television viewership for the event has surged. In the best-of-three finals of the 2024 WCWS, in which Oklahoma defeated rival Texas to win its fourth consecutive national title, television viewership on the ESPN family of networks reached the highest levels on record, with two million viewers tuning in (Callahan, 2024). The 2022 edition of the finals, also between Oklahoma and Texas, eclipsed the corresponding series of the Men's (Baseball) College World Series (MCWS), with an average of 1.7 million viewers tuning into the softball event, whereas the most-viewed game of the MCWS finals between Ole Miss and Oklahoma averaged 1.63 million viewers (Staff, 2022). Meanwhile, a recent expansion of the host stadium for the WCWS has enabled continued growth in attendance. Whereas the event hosted an average of 72,747 fans in the 2010s, it averaged 112,123 between 2021 and 2023, the first three seasons of increased seating capacity in Oklahoma City, an increase of over 54 percent (National Collegiate Athletic Association, 2024b).

Despite this live and televised success, college softball - with its compelling, multi-level postseason format - remains an under-researched sport in analytics literature. In this paper, we adjust a model, the Double Poisson, that was previously used to predict the outcomes of soccer matches by constructing two sets of adjustment variables and building three models off of these that are used in producing simulations of the tournament. The models will be evaluated against a base implementation of the Double Poisson in predicting the outcomes of the first round of major college softball's four-stage postseason, the regional round. The following section provides background on the Double Poisson model, while Section 3 describes the data used for the project and decisions made on which to include. Section 4 implements a base Double Poisson model on the three most recent college softball seasons - and describes the issues that arise from implementing this model in this setting. Section 5 details two pairs of adjustment factors to extend this model, while Section 6 applies these factors to construct models for the same three seasons, evaluating differences in performance. Lastly, we conclude with potential lines of future inquiry and results of this study in Section 7.

2 Background

2.1 NCAA Division I Softball Tournament Structure

The first NCAA Division I Softball Tournament was held in 1982 with a 16-team structure and culminated in a championship held in Omaha, Nebraska. Over time, the number of participants has evolved to the current 64-team setup, with 32 teams earning automatic bids by winning their respective conferences, and 32 teams being selected by a committee to receive at-large bids. Of these 64 teams, 16 are chosen to host four-team double elimination tournaments, termed *regionals*, on the first round of competition; the Tuscaloosa Regional from the 2024 edition of the tournament is shown below as an example. In this example, the "if necessary" game was not required because the undefeated team (Alabama) defeated the team with one loss (Southeastern Louisiana), thereby eliminating Southeastern Louisiana with the Lions' second loss.

The winners of these sixteen regionals move on to a best-of-three series against another regional winner, held at the higher seed's ballpark on the second weekend of the tournament and termed the *super regionals*. The eight super regional winners advance to the Women's College World Series (WCWS) in Oklahoma City. In WCWS play, the eight teams are split into two four-team double elimination brackets and essentially repeat the first two rounds over the course of nine days, culminating in the best-of-three WCWS Finals. The only difference is that the loser of the winner's bracket championship in each four-team bracket moves to the other four-team bracket instead of remaining in its own bracket (National Collegiate Athletic Association, 2024b). The complexity of this multistage tournament format, as well as the variation in team quality due to all conferences receiving at least one bid, make this a challenging format to build predictive models for.

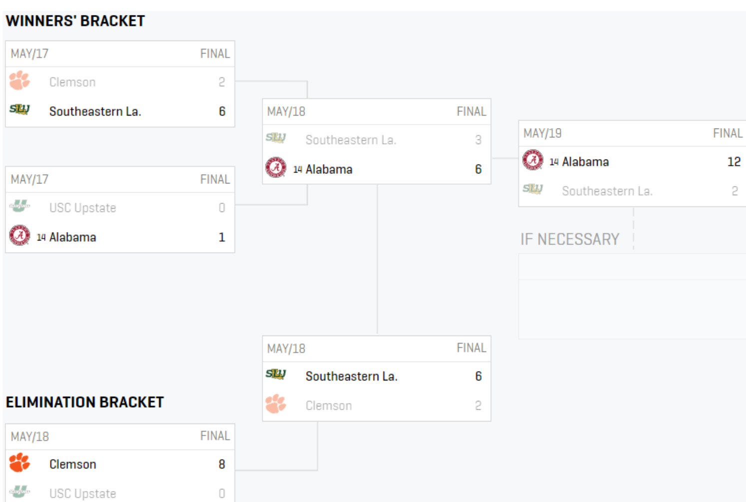

Figure 1: 2024 Tuscaloosa Regional Bracket

2.2 Double Poisson Models

Using Poisson random variables to model outcomes in English soccer was first proposed by M.J. Maher in 1982, under the justification that a team has many individual chances to score over the course of a match, but that the probability of any individual opportunity being successful was relatively low. In such a model, the number of goals scored by team i against team j, X_{ij}, is assumed to be Poisson distributed with a mean determined based on the offensive strength of team i and defensive vulnerability of team j, where the strengths are measured in expressions related to the number of goals scored and allowed (Maher, 1982)(Wackerly *et al.*, 2008). Under this model, the probability that team i defeats team j with X_{ij} as defined above is $P(X_{ij}) > P(X_{ji})$.

More recently, Penn and Donnelly have advanced this work, using a Double Poisson model to predict the results of the 2020 Euro soccer team tournament between European national teams. In their analysis, attacking (offensive) strengths and defensive vulnerability values are calculated, and matches against teams with exceedingly large defensive vulnerability values were removed prior to recalculating such values (Penn and Donnelly, 2022). In their work, issues with including teams with poor defensive quality were noted often; whereas their work removed such teams from the dataset, we include such teams (as even a team with a weak defensive strength may still win its conference and advance to NCAA tournament play) and attempt to construct adjustment variables to properly situate these teams.

3 Data

The dataset for these models includes all regular-season and postseason games from the 2022-2024 NCAA Division I college softball seasons. Game-by-game data was scraped from stats.ncaa.org using a purpose-built scraper in Python for each team. For the purposes of this work, only games between *full* Division I member schools were considered; games against member schools transitioning to Division I or those in Division II, III, or the NAIA were not considered as such schools may operate under different association-imposed scholarship restrictions and thus be inherently competitively imbalanced.

For each game, the date, score, and competing teams were recorded, and each team's conference affiliation was added using a file constructed by the scraper. In addition, the teams participating in each regional of each tournament were manually gathered from the NCAA website (National Collegiate Athletic Association, 2024a), along with the team's performance in the regional and tournament as a whole.

4 Base Model

4.1 Methodology

For all Double Poisson models in this work, we consider the offensive strength (O_{x_1}) and defensive vulnerability (V_{x_1}, where $V_{x_1 u}$ is the unscaled version) of team x_1 (out of a set of m full member institutions $x_1, x_2, ..., x_m$) in the following way. Given a team's set of n games prior to the NCAA tournament indexed by g, (g, f, a), where f represents the number of runs for (scored by) team x_1 in game g, and a represents the number of goals against team x_1 in game g, we have:

$$O_{x_1} = \frac{\sum_{g=1}^{n} f}{n}, \qquad V_{x_1 u} = \frac{\sum_{g=1}^{n} a}{n}, \qquad V_{x_1} = \frac{V_{x_{1u}}}{\frac{1}{m}\sum_{i=1}^{m} V_{x_{iu}}}.$$

Thus, O_{x_1} is the average number of runs scored in a particular year by team x_1 across all games prior to the tournament, and V_{x_u} (unscaled defensive vulnerability) is the average number of runs allowed in a particular year by team x_1 across all games prior to the tournament. The V_x value is then taken by dividing the unscaled defensive vulnerability by the average of all unscaled defensive vulnerabilities.

This definition of O_x and V_x differs significantly, however, from the derivation of those values in the previous work of Penn and Donnelly. In their work, offensive strength and defensive vulnerability parameters were estimated, using the maximum likelihood estimator, as the solutions to a nonlinear system of equations. The objective function for such a system involved, for m teams, two square matrices of dimension m. In the matrix used for games played, here denoted G, the entry g_{ij} denoted the number of international games played between countries i and j in their dataset. The number of goals scored by country i against country j was recorded similarly in a score matrix, here denoted S. Because the purpose of their work was to predict the Euro 2020 competition, only 54 teams were included in the model, so a unique solution was found to the system of equations. (Penn and Donnelly, 2022).

However, when attempting to apply the same process to estimate the parameters O_x and V_x in the context of college softball, the sparsity of the G and S matrices quickly caused the system to not converge at a unique solution. On average, a full Division I member institution competed against just under 25 distinct opponents over the course of the entirety of the regular season, which on average spanned just under 49 games, due to the series-based nature of softball. Because of this, the matrices G and S are highly sparse; with 296 full Division I teams competing in the 2024 season, an average of 271 zero entries would exist in each row and column. Therefore, we propose this base model for estimating the parameters concerning a team's offensive and defensive strength and aim to use adjustment factors to tune such estimates, with the aim of producing values that remain true to the concept of the Double Poisson model in considering offensive and defensive strengths in predicting scoring. In addition, this model aims to solve a problem found by Penn and Donnelly introduced by teams with large defensive vulnerabilities; whereas their work fixed the largest V value at 1, we scale our V values relative to the mean so that a single weak defensive team has less impact on the remaining teams' V values.

From these offensive and defensive vulnerabilities, we calculate the mean of the Poisson random variables representing the predicted number of runs scored by team a against team b and vice versa. Letting $\mu_{a,b}$ represent the mean of $R_{a,b}$, the predicted number of runs scored by a, and $\mu_{b,a}$ represent the mean of $R_{b,a}$, the predicted number of runs scored by b, these are expressed as follows in the base Double Poisson model:

$$\mu_{a,b} = (O_a)(V_b), \qquad \mu_{b,a} = (O_b)(V_a), \qquad R_{a,b} \sim Poisson(\mu_{a,b}), \qquad R_{b,a} \sim Poisson(\mu_{b,a}).$$

Therefore, the probability of team a defeating team b, $p_{a,b}$ is given by the following for the case where ties are possible:

$$p_{a,b} = P(R_{a,b} > R_{b,a}).$$

However, a significant difference between predictive modeling in soccer matches versus college softball is the issue of ties. In soccer, a game may end after regulation is played in a tie. However, ties are extraordinarily rare - and usually the result of weather-related interruptions, as was the case between Auburn and Virginia Tech in 2024. In that situation, the game concluded in a 5-5 tie due to unplayable field conditions following an extended rain delay near the end of the game (Virginia Tech, 2024). Because the Poisson distribution is discrete, $P(R_{a,b} = R_{b,a}) \neq 0$, so we must consider the situation of ties.

Ties in softball at the end of regulation play are, except in these rare cases, settled with extra innings. We assume that the strengths that a team has during the seven innings of regulation play carry over to extra innings, so the relative probabilities of victory should remain constant. While this assumption may not be true in select cases, such as a player becoming injured or otherwise unable to continue, it is likely not a major source of error, as softball teams routinely play more than seven innings per day, especially during early-season non-conference tournaments - and even the NCAA tournament. Under this assumption, we eliminate the possibility of ties by maintaining this proportionality and define $p_{a,b}$ in the following manner for the remainder of this work:

$$p_{a,b} = \frac{P(R_{a,b} > R_{b,a})}{P(R_{a,b} > R_{b,a}) + P(R_{b,a} > R_{a,b})}$$

4.2 Process

For each year under study, we calculated the O and V values as defined above for each tournament participant using the data described in section 3. For each of the sixteen regionals, a probability matrix was constructed such that the entry in the ith row and jth column is the calculated probability that the team in the ith row would defeat the team in the jth column, with ties accounted for as above. Two examples of these probability matrices are below:

Then, using the initial setup of teams in the regional and the probability distribution calculated in the prior step, each regional was simulated 10,000 times, with the winner, runner-up, third-place (1-2 record) team, and fourth place (0-2 record) team noted for each and analyzed.

2024 Tuscaloosa Regional (won by Alabama)	Alabama	Clemson	Southeastern Louisiana	USC Upstate
Alabama		0.298	0.260	0.494
Clemson	0.702		0.437	0.712
Southeastern Louisiana	0.740	0.563		0.753
USC Upstate	0.506	0.288	0.247	

2024 Durham Regional (won by Duke)	Duke	South Carolina	Utah	Morgan State
Duke		0.899	0.949	0.915
South Carolina	0.101		0.573	0.531
Utah	0.051	0.427		0.461
Morgan State	0.085	0.469	0.539	

Table I: Probability Tables for 2024 Tuscaloosa and Durham Regionals

4.3 Results

When applied to the 2024 season, the base Double Poisson model correctly predicted the regional winner an average of 4767.875 times (as indicated by the red line in the below plot) over the course of 10,000 simulations of the double-elimination bracket. Over the three-year dataset, this year's performance was slightly below average, as it was correct 5068.6875 times across all regional brackets in that span, on average.

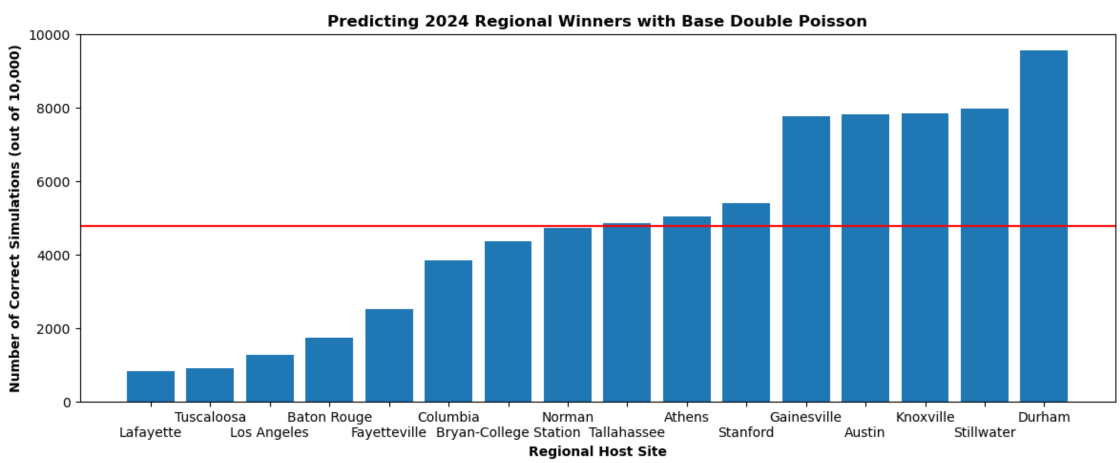

Figure 2: Predicting Regional Winners with Base Double Poisson (line indicates mean)

4.4 Analysis

Several interesting takeaways emerge from applying the base model to the 2024 postseason, including that the model predicted the three-time defending champion - and 2nd overall seed of all teams - Oklahoma Sooners to win their regional in less than half of trials. Under the base model, the Sooners are given a 0.472 probability to defeat Boston University, which was able to accumulate strong scoring and defensive averages against weaker competition over the course of the season. Meanwhile, the Tuscaloosa Regional was predicted correctly in just under nine percent of trials, as the host Crimson Tide were not favored by the model against any of their regional opponents; in reality, Alabama went undefeated in regional play and won the regional final in run-rule fashion. This illustrates a key issue with the model that adjustment factors aim to solve: that scoring averages are considered irrespective of the quality of opponent.

5 Adjustment Variables

To account for the vast difference in team quality between college softball teams and correct some of the errors associated with the above "base" model, we propose two pairs of variables to account for this: the conference adjustment variables and schedule adjustment variables. Each variable is constructed such that a "theoretically average" team would have a value equal to 1, and each pair of variables contains one variable adjusting the team's offensive strength and one variable adjusting the team's defensive vulnerability with respect to the attribute under study.

5.1 Conference-Adjustment Variables

The underlying assumption of the **conference-adjustment variables** is that teams compete in conferences that contain teams of relatively similar quality, so a team's conference affiliation provides some information about its expected performance. Because of this, games between teams of differing conferences may be used to construct an approximate judgment of the conference's quality. For instance, Marist, a member of the MAAC, was within the top 20 in offensive strength in 2024. However, the MAAC's schedule-adjusted defensive vulnerability, as defined below, was roughly 1.35, meaning that teams in the MAAC allowed non-conference opponents to score 35 percent more than their season-long averages in the typical game. This variable therefore adds context to Marist's high run-scoring totals.

Let the set (g, f, a, x_i) be the set of G non-conference games played by teams in conference C_1, indexed by $g \in \{1, 2, 3, ..., G\}$, in which teams in conference g scored f runs and allowed a runs against team x_g. Then the schedule-adjusted offensive strength, O_{C_1}, and schedule-adjusted defensive vulnerability, V_{C_1} are defined as follows:

$$O_{C_1} = \frac{\sum_{g=1}^{G}(\frac{f}{V_{x_{gu}}})}{G} \qquad V_{C_1} = \frac{\sum_{g=1}^{G}(\frac{a}{O_{x_g}})}{G}$$

Essentially, a conference's offensive strengths is the average, over all non-conference games, of the ratio between the number of runs scored by the team from that conference to the season-long average number of runs given up by the team's opponent. Conversely, the conference's defensive vulnerability is the average, over all non-conference games, of the ratio between the number of runs given up by the team from that conference to the season-long average number of runs scored by the team's opponent. We would reasonably expect a "theoretically-average" conference to have teams that give up as many runs as their opponents typically score, so this "theoretically average" conference would have $O_C = 1$ and $D_C = 1$. The top five and bottom five conferences in schedule-adjusted offensive strength and schedule-adjusted defensive vulnerability are shown below for the 2024 season. As may be expected, this set of variables primarily benefits schools in "Power" conferences:

Conference	2024 O_C Value
SEC	1.655
Big 12	1.634
ACC	1.482
Pac-12	1.365
Big Ten	1.284
...	...
MAAC	0.7412
NEC	0.7161
MEAC	0.6406
Horizon	0.6159
SWAC	0.5409

Conference	2024 V_C Value
SEC	0.440
Pac-12	0.594
Big 12	0.679
ACC	0.705
Sun Belt	0.797
...	...
America East	1.333
MAAC	1.346
Horizon	1.464
MEAC	1.584
SEAC	1.679

Table II: Top and Bottom Five Conference Offensive Strength, Defensive Vulnerability Values

5.2 Schedule-Adjustment Variables

Meanwhile, the underlying assumption of the **schedule-adjusted variables** is that strong teams will score more than their opponents usually allow - and conversely allow fewer runs than their opponents typically score, and that these quantities can shed light on the quality of a team's offense and defense. This can serve as a testament to the success of even the most successful offenses: Oklahoma's 8.1 runs per game in the 2024 season ranked near the top of college softball, but their offensive dominance becomes clear when their schedule-adjusted offensive strength value is examined as defined below: the Sooners scored 96 percent more runs, on average, than their opponents' runs allowed average.

Let the set (g, f, a, x_g) be the set of games played by team x_i over the course of a season, indexed by $g \in \{1, 2, 3, ...G\}$, where team x_i scores f runs and allows a runs against team x_g. Then the schedule-adjusted offensive strength of team x_i, denoted O_{S_i}, and schedule-adjusted defensive vulnerability of team x_i, denoted V_{S_i}, are given by:

$$O_{S_i} = \frac{\sum_{g=1}^{G} \frac{f}{V_{x_{g_u}}}}{G} \qquad V_{S_i} = \frac{\sum_{g=1}^{G} \frac{a}{O_{x_g}}}{G}$$

Essentially, a team's schedule-adjusted offensive strength is the average, over all games played, of the ratio between its number of runs scored in the particular game and the average number of runs allowed by its opponent. Conversely, a team's schedule-adjusted defensive vulnerability is the average, over all games played, of the ratio between its number of runs allowed in the game to the average number of runs scored by its opponent over the course of the season. We propose that a "theoretically average" team would score as much as each of its opponents usually allows - and give up as many runs as its opponents usually score - thereby having $O_S = V_S = 1$. The following gives the top five and bottom five schedule-adjusted offensive strength and defensive vulnerability values for the 2024 season:

Team	2024 O_S Value	Team	2024 V_S Value
Florida	2.194	Boston	0.296
Miami-Ohio	2.083	Duke	0.322
Florida State	1.999	Oklahoma	0.333
Oklahoma	1.962	Tennessee	0.340
Texas	1.919	Stanford	0.413
...	...	...	...
Arkansas Pine Bluff	0.483	UMES	1.882
Alcorn State	0.476	Detroit-Mercy	1.907
FDU	0.453	Alabama A-M	1.964
Detroit-Mercy	0.401	St. Bonaventure	2.100
Lafayette	0.399	Mississippi Valley State	2.131

Table III: Top and Bottom Five Schedule-Adjusted Offensive Strength, Defensive Vulnerability Values

6 Adjusted Models

6.1 Conference-Adjusted Double Poisson

As discussed in section 5, the conference adjustment variables, O_C and V_C, are defined such that a "theoretically average" team has $O_C = 1$ and $V_C = 1$. Therefore, we allow these to act as "multipliers" on the existing averages in the conference-adjusted Double Poisson model. Just as before, we define $\mu_{a,b}$ to be the mean of the Poisson distribution predicting the number of runs scored by team a against team b, and $R_{a,b}$ to be the Poisson random variable governed by $\mu_{a,b}$. For team a that is a member of conference i, and team b that is a member of conference j, we express the conference-adjusted Double Poisson as follows:

$$\mu_{a,b} = (O_a)(V_b)(O_{C_i})(V_{C_j}), \qquad \mu_{b,a} = (O_b)(V_a)(O_{C_j})(V_{C_i}),$$
$$R_{a,b} \sim Poisson(\mu_{a,b}), \qquad R_{b,a} \sim Poisson(\mu_{b,a}).$$

During the 2024 postseason, the conference-adjusted Double Poisson model correctly predicted each regional correctly an average of 6909.75 times out of 10,000 trials, just above its three-year average of 6505.33 correct predictions per regional. In the most recent season, this is an increase of over 2000 trials compared to the base Double Poisson model, which supports the idea that teams from stronger conferences tend to perform better in the postseason. One regional in particular that performed much better under this model was the aforementioned 2024 Tuscaloosa Regional, which was correctly predicted 8.8 percent of the time in the base model, but 68.8 percent of the time under the conference-adjusted model. Under the conference-adjusted Double Poisson model, Alabama (a member of the top-rated conference by adjustment variables), was favored in every matchup, which better reflects the reality of the bracket shown on page two.

Figure 3 shows this improvement, as six of the sixteen regionals in 2024 were predicted with over 97 percent accuracy by the conference-adjusted Double Poisson model. Conversely, a key issue with using conference adjusting variables alone is highlighted, it performed worse in predicting the Tallahassee Regional, with 48.33 percent of trials correctly predicting Florida State as the winner under the base model and 33.21 percent of trials in the conference-adjusted model reaching correct predictions. Auburn, a team in the SEC, competed in the Tallahassee Regional and received the benefit under this model of membership in the top offensive and defensive conference, despite the fact that they finished in the bottom third of the conference (Southeastern Conference, 2024).

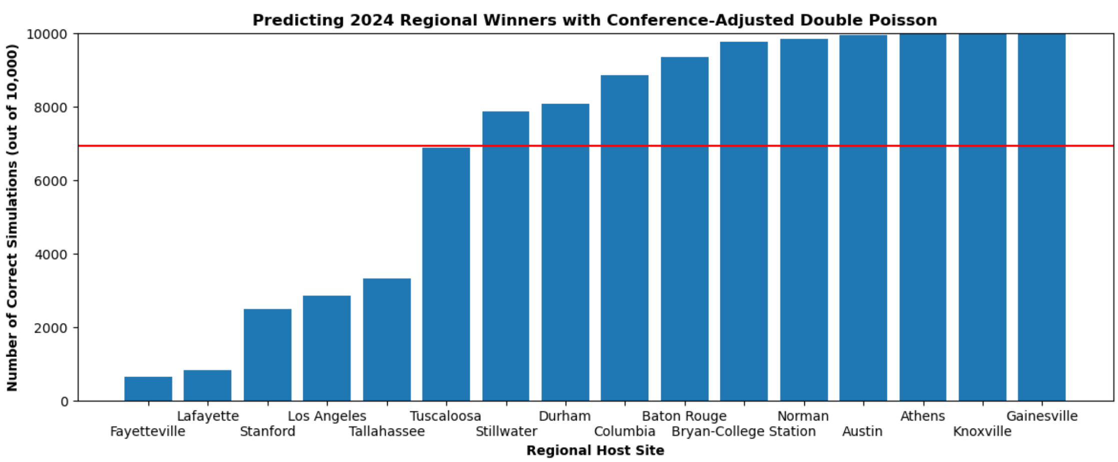

Figure 3: Predicting Regional Winners with Conference-Adjusted Double Poisson (line indicates mean)

6.2 Schedule-Adjusted Double Poisson

Similar to the conference-adjustment variables, the schedule-adjustment variables are defined such that the "theoretically average" team has $O_S = 1$ and $V_S = 1$, so they can effectively function as "multipliers" under this model. The schedule-adjusted Double Poisson attempts to control for a team's strength of schedule and incorporate their opponents' averages into the equation. This model, therefore, favors teams that do well with respect to the opponents on their schedule and attempts to control for the fact that teams may choose to play vastly different non-conference schedules in college softball. With $\mu a, b$ representing the mean of the Poisson random variable $R_{a,b}$ representing the number of runs scored by team a (with schedule adjustment variables O_{S_a} and V_{S_b}) against team b (with schedule adjustment variables O_{S_b} and V_{S_b}), we define the schedule-adjusted Double Poisson model as follows:

$$\mu_{a,b} = (O_a)(V_b)(O_{S_a})(V_{S_b}), \qquad \mu_{b,a} = (O_b)(V_a)(O_{S_b})(V_{S_a}),$$

$$R_{a,b} \sim Poisson(\mu_{a,b}), \qquad R_{b,a} \sim Poisson(\mu_{b,a}).$$

During the 2024 postseason, the schedule-adjusted Double Poisson model performed better than the base model, but not as well as the conference-adjusted model, as it correctly predicted each regional an average of 6184.31 times out of 10,000 trials. This is consistent with its three-year success rates, as the average number of correct predictions under this model was 6310.46. This gain in performance, while not as pronounced as that of the conference-adjusted model, is insightful, as it does not require any additional information beyond the scores of games (the same data used to build the base model), whereas the conference-adjusted model also requires each team's conference affiliation.

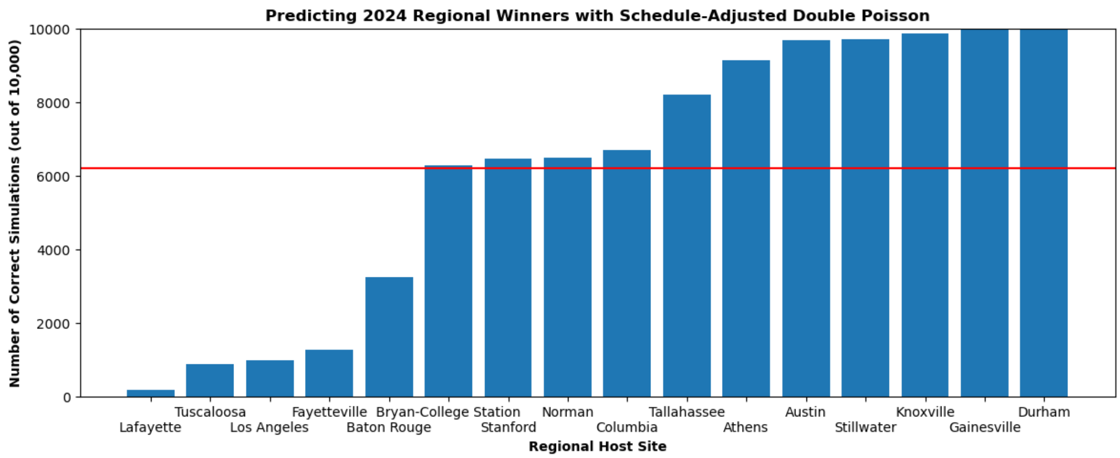

Figure 4: Predicting Regional Winners with Schedule-Adjusted Double Poisson (line indicates mean)

The aforementioned 2024 Tallahassee Regional demonstrates the improvements from conference-adjusted Double Poisson model, as it predicted that Florida State would defeat Auburn with probability roughly 0.86, whereas the conference-adjusted model assigned probability roughly 0.38 to that event. The events on the field (Florida State defeating Auburn 10-4 to cap an undefeated run through the regional) suggest that the schedule-adjusted model better accounts for games between a team at the top of a slightly weaker conference (Florida State) and a team near the bottom of a stronger one (Auburn) (National Collegiate Athletic Association, 2024a). This higher probability of a win for Florida State against Auburn translated to the Seminoles being correctly predicted to win the regional in almost 82 percent of trials under the schedule-adjusted model, compared to 48.57 percent of trials in the base model and 33.21 percent of trials in the conference-adjusted model.

6.3 Conference-and-Schedule-Adjusted Double Poisson

The final model, the conference-and-schedule-adjusted Double Poisson, aims to incorporate both factors associated with stronger conference membership and strength of schedule into its predictions. Using the same notation for $\mu_{a,b}$ and $R_{a,b}$ as before with team a in conference i and team b in conference j, we define:

$$\mu_{a,b} = (O_a)(V_b)(O_{C_i})(V_{C_j})(O_{S_a})(V_{S_b}), \qquad \mu_{b,a} = (O_b)(V_a)(O_{C_j})(V_{C_i})(O_{S_b})(V_{S_a}),$$

$$R_{a,b} \sim Poisson(\mu_{a,b}), \qquad R_{b,a} \sim Poisson(\mu_{b,a}).$$

Of the four models considered, the Double Poisson that adjusts for conference affiliation and strength of schedule produced the most regionals correctly predicted in greater than 90 percent of trials, as ten regionals fit that criteria, with five succeeding greater than 99 percent of the time, as shown in Figure 5. This model performed better than its three-year average of 69.75 percent of trials coming back with a correct regional winner, with 71.45 percent of 2024 regional champions correctly predicted. The nature of this model as a compromise between the two prior adjusted models is apparent, as this model performed considerably better on the Tallahassee Regional than the conference-adjusted model but fell short of the 82 percent threshold that the schedule-adjusted model reached. Meanwhile, in some cases, such as for the Baton Rouge Regional, this broader model outperformed both of the prior adjusted models, as this model predicted 97.74 percent of the time that LSU would win, whereas the conference-adjusted model, which greatly outperformed the schedule-adjusted model, made the same prediction in 93.27 percent of trials.

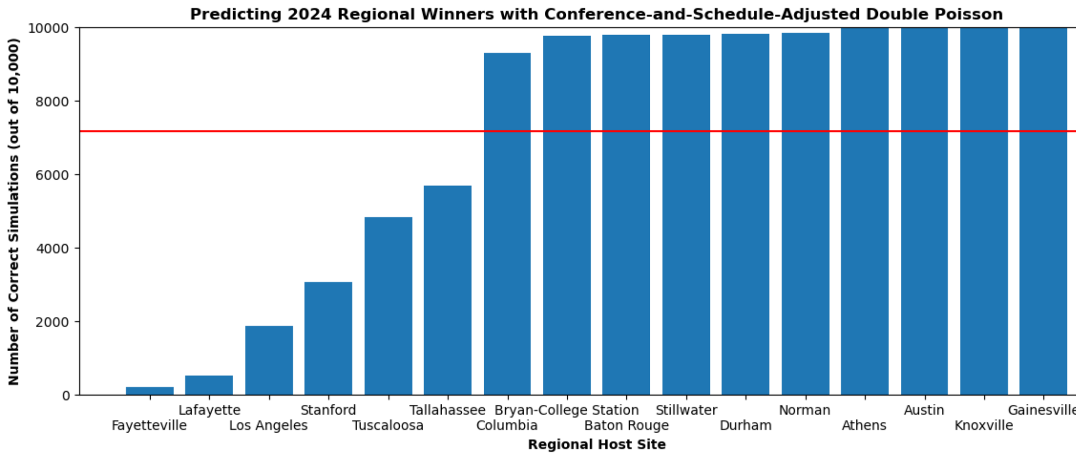

Figure 5: Predicting Regional Winners with Conference-and-Schedule-Adjusted Double Poisson

6.4 Analysis

As shown below, adjusted models substantially outperformed the base Double Poisson model in accuracy while simultaneously increasing the standard deviation of those prediction accuracy values across the three-year dataset. This both demonstrates their ability to incorporate some of the inherent inequities in the landscape of college athletics that scores and team affiliation alone cannot, but also the difficulty of those factors to fully address.

Mean Trials Correct (of 10,000)					Standard Deviation of Trials Correct				
Year	Base	SADP	CADP	CSADP	Year	Base	SADP	CADP	CSADP
2022	4940	5900	6021	6362	2022	3005	4028	3194	3970
2023	5498	6847	6585	7418	2023	2829	3420	3521	3506
2024	4768	6184	6910	7145	2024	2823	3691	3573	3803

Table IV: Mean and Standard Deviation of Correct Trials Between Models
SADP - Schedule-Adjusted Double Poisson, *CADP* - Conference-Adjusted Double Poisson,
CSADP - Conference-and-Schedule-Adjusted Double Poisson

7 Conclusion and Future Work

In this paper, we have proposed a base model that estimates the parameters of the Double Poisson for a setting in which team-versus-team matrices are sparse and evaluated the performance of the model against alternatives that take into account attributes of strength of schedule and conference affiliation. We have found that these alternative models bring with them greater accuracy, on average, but also greater variance in the distribution of those accuracy values. Future work in this area could involve assessing the relative performance of these models in the second (Super Regional) round of the NCAA Division I Softball Tournament, as well as the Women's College World Series, and constructing an algorithm to predict which of the four models would work best given a set of input teams. These investigations would serve to advance quantitative approaches to this under-researched, yet quickly-growing, sport.

References

Callahan, Katie (2024). "ESPN Viewership Clears the Bases During 2024 Division I NCAA Softball Season", **available at**: `https://espnpressroom.com/us/press-releases/2024/06/espn-viewership-clears-the-bases-during-2024-division-i-ncaa-softball-season/` (accessed 31 Aug. 2024).

Maher, M.J. (1982). "Modelling association football scores", *Statistica Neerlandica*, Vol. 36, (3), pp. 109–118. DOI: `https://doi.org/10.1111/j.1467-9574.1982.tb00782.x`.

National Collegiate Athletic Association (2024a). "2024 NCAA Division I Softball Official Bracket", **available at**: `https://www.ncaa.com/brackets/softball/d1/2024` (accessed 1 Sept. 2024).

— (2024b). "2024 Women's College World Series® Records Book", **available at**: `http://fs.ncaa.org.s3.amazonaws.com/Docs/stats/softball_wcws_rb/FullBook.pdf` (accessed 31 Aug. 2024).

Penn, Matthew J. and Donnelly, Christl A. (2022). "Analysis of a double Poisson model for predicting football results in Euro 2020", *PLOS One*, DOI: `https://doi.org/10.1371/journal.pone.0268511`.

Southeastern Conference (2024). "Standings - Softball - Southeastern Conference", **available at**: `https://www.secsports.com/standings/softball` (accessed 4 Sept. 2024).

Staff, The Athletic (2022). "Women's College World Series final outgains men's final, averages 1.7 million viewers", **available at**: `https://www.nytimes.com/athletic/4169889/2022/06/29/womens-college-world-series-final-outgains-mens-final-averages-1-7-million-viewers/` (accessed 31 Aug. 2024).

Virginia Tech (2024). "Peck rips grand slam to give Hokies 9-5 lead, before play was suspended in top of the seventh", **available at**: `https://hokiesports.com/news/2024/02/10/peck-rips-grand-slam-to-give-hokies-9-5-lead-before-play-was-suspended-in-top-of-the-seventh`.

Wackerly, Dennis D., Mendenhall III, William, and Scheaffer, Richard (2008). *Mathematical Statistics with Applications*, Thomson, pp. 131–135.