

College Football Chaos

Anissa Williams
Data Science
College of Charleston
Charleston, SC

anissawilliamschs@gmail.com

Braeden Mefford
Data Science
College of Charleston
Charleston, SC

meffordbm@g.cofc.com

ABSTRACT

College football's appeal lies in its unpredictability, where momentum swings and explosive plays often defy expectations. We introduce the Chaos Factor, a novel metric designed to quantify volatility in college football games. Derived from win probability fluctuations, lead changes, and explosive plays, chaos captures dimensions of unpredictability that traditional Elo ratings overlook. Using play-by-play data from the 2018–2023 seasons, we evaluate chaos across Power Five matchups and demonstrate its predictive utility. Our findings show that chaos highlights games where volatility spikes, offering insights into outcomes that Elo alone fails to anticipate. This work contributes to sports analytics by formalizing game volatility as a measurable construct and by showing its potential to complement existing ranking and prediction systems.

KEYWORDS

Win probability volatility, chaos factor, explosive play, conference, college football analytics, ELO, rankings

1 Introduction

College Football is a national pastime, and it is especially enjoyed by hundreds of millions of fans. The venues can sometimes exceed 80,000 spectators, all there to cheer on their team, and enjoy the spectacle. The ultimate prize at stake is a playoff and national championship. Each week, these games are reviewed, and the outcomes directly influence the CFB rankings, which lead to the eligibility of playing in the postseason games leading to the championship. Ranking is one variable - which is largely subjective and includes the team's past performance (partly Elo), and the level of competition [1]. This leads one to question the effectiveness and fairness of the metric [2], when so much is on the line, especially later in the season. We propose introducing another predictor - the chaos factor.

We seek to introduce another aspect of college football worth quantifying and utilizing when predicting outcomes historically or through similar parameters - volatility, explosive plays, and lead changes. These factors can help

us determine how a team may perform against another, and whether a team is particularly susceptible to high chaos (more volatility, instability, and unpredictability) that should be factored into these critical outcomes that make or break a team's season.

2 Literature Review

Quantifying the excitement fans experience while watching football has always captured the interest of statisticians who follow the game. One such example comes from a 2012 MIT paper where the authors aim to determine whether 'psychological momentum' from a big play during a NFL game has a spill over effect onto the ensuing plays (Johnson et. al., 2012) [3]. While the Johnson et al. paper set distinct parameters for 'good' and 'bad' momentum, their overall conclusion was that plays defined under their momentum criteria had no significant effect on the following plays.

A paper that had success in its win prediction analysis is the 2014 Leung & Joseph paper on Sports data mining. [4] They contend that due to the limited sample size in NCAA football such as limited head to head match ups year over year, and the limited time that individual players have their respective teams, make it difficult to accurately predict game outcomes. The way Leung & Joseph overcome this shortfall is by using data mining tools to compare a team's historical performance without directly comparing it to the head to head opponent. They also cite influences in advanced statistics from the Society for American Baseball Research (SABR). SABRmetrics are known for their ability to reach a deeper understanding of the game they study by using more advanced statistical techniques to analyze a player or game. Leung & Joseph follow in the footsteps of SABRmetrics by using stats such as RPI, Pythagorean wins, offensive strategy, and turnover differential. Their results speak for themselves as their win prediction has at least a 30 percentage point advantage over more direct analysis across 2 seasons. We intend to learn from the related literature by using advanced stats such as 'explosive plays', 'lead change', and 'win probability volatility' to gain a

deeper understanding of how games develop under 'chaotic' circumstances.

3 Data

3.1 Data Source

Our data comes from the CFBFastR database which tracks a variety of play by play and aggregated data by team, season, and player from the years 2018-2023. We selected the raw play by play data from the 2018-2023 seasons so that we would be able to craft our advanced statistics with as much control as possible, aggregating them with our own parameters.

3.2 Data Cleaning

The play by play data had a number of issues that needed to be corrected before we could perform analysis. The original data only identified teams under who had possession of the ball, not by their team name. Additionally, there were a number of inconsistencies among how plays were identified. We then came up with a simplified approach of creating a scoreboard which normalized the attributes in the data so we could have an easier path towards feature implementation in the later stages of the project. The reconstructed scoreboard left us with a play by play database that was easier to read and produced the score for each team after the play.

3.3 Data Decisions

Upon viewing the cleaned data, we opted to include only the Power 5 conferences, because these are the conferences considered for the College Football Playoff. This was a tradeoff that eliminates some of the historical non-Power 5 v. Power 5 upsets. However, considering the parity of most conferences, and the aforementioned importance of the Power 5 Conference inclusion to the playoffs formula, this was justified.

4 Feature Development

To quantify our chaos feature, we wanted to design a statistic that factored in lead changes, win probability volatility, and explosive plays to maintain the effective strategies implemented in the previous literature, while still wanting to test if explosive plays or 'momentum' has any influence on a game. In addition, as we moved farther along in our analysis, we realized we wanted to incorporate other variables to examine. This led to a new data pull and merge to include the pregame elo scores for the home and away teams.

5 Methodology

We define the Chaos Factor as a weighted combination of win probability volatility, lead changes, and explosive plays:

$$\text{Chaos} = w_1 \cdot \text{WinProbVolatility} + w_2 \cdot \text{LeadChanges} + w_3 \cdot \text{ExplosivePlays}$$

Chaos is computed at the game level and aggregated to season averages for both matchups and the league overall.

Rivalry games are compared against league averages to contextualize volatility. Elo ratings provide a baseline measure of team strength, with expected win probability defined as:

$$\text{HomeWin} = \frac{1}{1 + 10^{\frac{(E_H + HFA) - E_A}{400}}}$$

We compare Elo differentials (Δ Elo) with chaos scores to evaluate predictive utility. Models using Elo alone are contrasted with models incorporating both Elo and chaos, with accuracy assessed via correlation, upset classification, and predictive error metrics.

6 Results

Pregame Elo differentials across 2018–2023 exhibit a wide distribution, with most contests clustered around moderate strength gaps but a distinct tail of heavily lopsided matchups (Figure 1). Elo provides a useful baseline measure of team strength, yet it does not capture volatility spikes that occur even in games with large rating disparities.

Distribution of Elo Differentials (2018–2023)

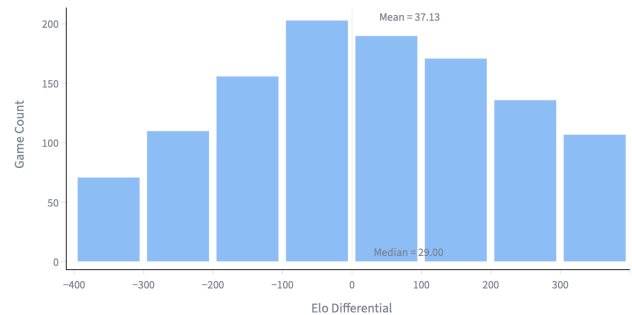


Figure 1. Distribution of pregame Elo rating differentials across Power Five matchups, 2018–2023. Most contests fall within moderate ranges, but a long tail of lopsided games underscores the need for volatility measures such as the Chaos Factor.

College Football Chaos Charleston, 2025

Chaos trends highlight these volatility spikes, particularly in late-season contests with playoff implications. Although actual outcomes were not included in this analysis, predicted upsets were disproportionately associated with high chaos values. Models that combined Elo and chaos improved upset identification compared to Elo alone, demonstrating that chaos offers complementary predictive insight and enhances forecasting accuracy.

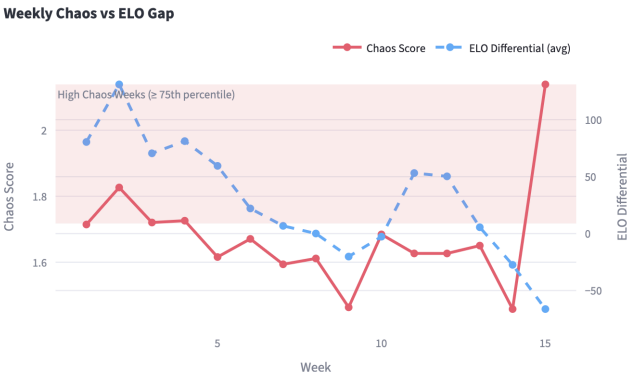


Figure 2. Comparison of Elo rating differentials and Chaos Factor trends across the season. Chaos spikes in late-season games despite stable Elo predictions, revealing volatility not captured by traditional strength metrics.

A case study of Alabama's 2023 season illustrates the divergence between Elo and chaos. Despite maintaining an Elo rating above 2200 throughout the season, Alabama experienced multiple high-chaos weeks, including a dramatic spike in Week 15 (Figure 3). This suggests that even elite teams are susceptible to volatility, and that Chaos Factor captures dimensions of unpredictability not reflected in strength-based metrics.

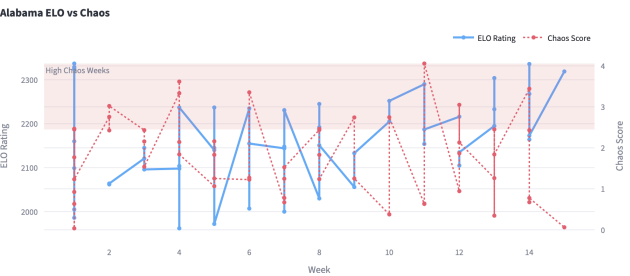


Figure 3. Alabama's Elo rating vs. Chaos Factor across the seasons. Despite consistently high Elo, multiple high-chaos weeks reveal volatility not captured by strength metrics.

Rivalry matchups often defy league-wide volatility norms. Figure 3 illustrates the Alabama vs Texas A&M series, where chaos scores fluctuate sharply across seasons. The 2021 contest approached the predicted chaos threshold of 3.66, underscoring the rivalry's tendency to produce unpredictable outcomes regardless of Elo-based expectations.

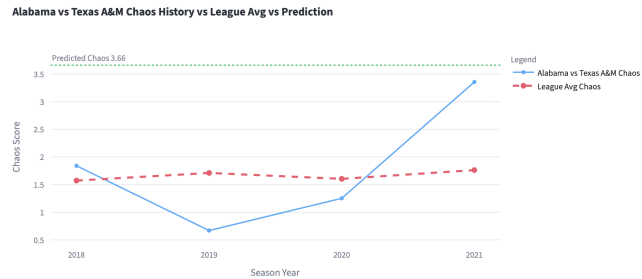


Figure 4. Chaos Factor scores for Alabama vs Texas A&M compared to league-wide averages and predicted chaos. Rivalry volatility exceeds baseline expectations, with future matchups projected to remain highly unpredictable.

7 Conclusion

This study introduces the Chaos Factor as a formal measure of volatility in college football. By quantifying win probability swings, lead changes, and explosive plays, chaos captures dimensions of unpredictability overlooked by Elo ratings. Analysis of Power Five games from 2018–2023 shows that chaos identifies high-volatility contests, contextualizes rivalry dynamics, and improves upset prediction when combined with Elo.

Chaos thus represents a valuable complement to traditional strength-based metrics, offering decision-makers and analysts a richer understanding of game outcomes. Future work should explore integrating chaos with sentiment analysis, excitement indices, and other advanced predictors to further enhance forecasting and deepen the narrative of college football analytics.

REFERENCES

[1] Benjamin Lucas 2024. *An Analysis of NCAA College Football Playoff Team Selections Using an Elo Ratings Model*. DOI: <https://doi.org/10.48550/arXiv.2403.03862>

[2] Bill Radjewski. 2021. Talking Tech: Calculating Elo Ratings for College Football. DOI: <https://blog.collegefootballdata.com/talking-tech-elo-ratings/>

[3] A.W. Johnson, A.J. Stimpson, T. Clark, Turning the Tide: Big Plays and Psychological Momentum in the NFL, 2012, MIT Sloan Sports Analytics Conference. DOI: <https://hsl.mit.edu/biblio/turning-tide-big-plays-and-psychological-momentum-nfl-poster/>

[4] Carson K. Leung, Kyle W. Joseph 2014. Sports Data Mining: Predicting Results for the College Football Games, *Procedia Computer Science*, Volume 35, 2014, Pages 710-719, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2014.08.153>.