Analysis of the Potential Implementation of an Automatic Strike Zone

Devon Goetz, Zack Kopstein, Simran Pabla, Claire Wichman

Sports Lab

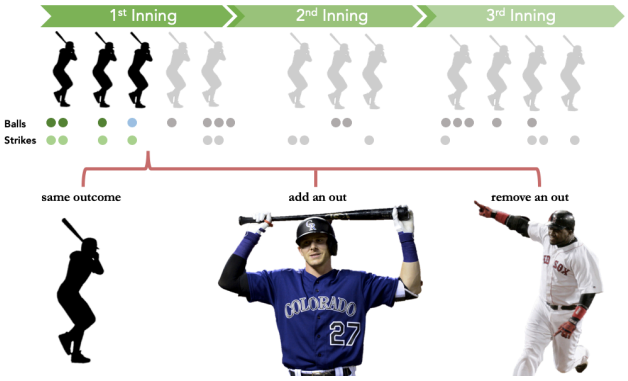
Project Overview

The MLB currently possesses the ability to implement an Automatic Strike Zone, eliminating the need for an umpire to be calling balls and strikes. Before using it, however, it is important to know how this addition might change the game.

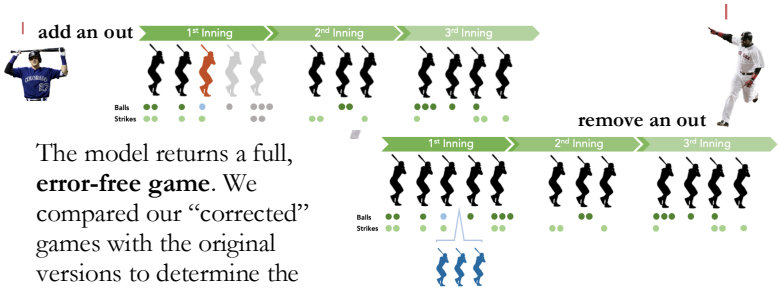
Methodology & Approach

I. Existing Discrepancies: After determining that our metrics of focus would be **strikeouts**, **walks**, **runs scored**, and **pitches thrown**, we grouped at bats by the number of offensive and defensive mistakes they contained. **Offensive mistakes** (true strikes that had been called balls) are advantageous for the batting team, or while **defensive mistakes** (true balls that have been called strikes) aid the fielding team. We defined a **true strike** as a ball that was inside, or touched any part of, the strike zone, which was 17 inches across the plate and customized on each pitch to the height of the batter, using top and bottom data provided. This breakout allowed us to compare the average number of each metric with the average number in a perfect, mistake-free at bats in order to quantify the impact

II. Predicted Changes: We created a **predictive model** that steps through each pitch. When it finds an umpire error, it corrects the state (ex. if the third pitch took the count to 2-1, but it was an incorrectly called ball, the new count would be 1-2). From there, it predicts the **next most likely state** for the game to enter, and so on, until we reach the end of the at bat. After that, there are three possibilities: the outcome is the same, the batter is out when he previously got on, or the batter is on when he previously got out.



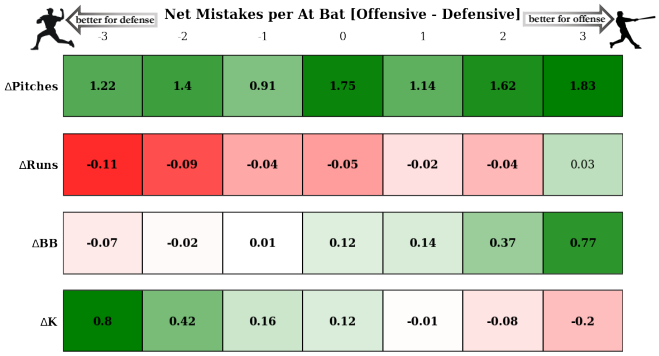
If nothing changes, we stitch reality back on immediately. If we add or remove an out, we have to adjust the runners and step through a probability function that allows us to preserve real at bat outcomes and move runners accordingly. If we remove an out, we have to add in fake batters, using the transition matrix, until we get back to the correct number of outs. Then we can return to reality, again moving our new set of runners as needed based on true at bats.



The model returns a full, **error-free game**. We compared our "corrected" games with the original versions to determine the difference the mistakes made in the metric outcomes.

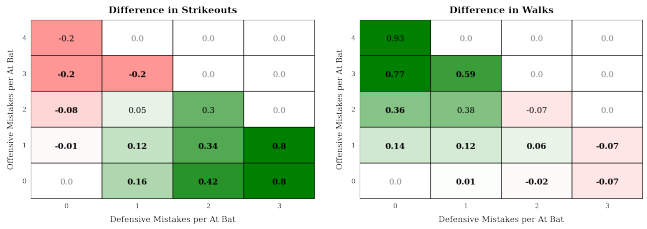
The main objective is to quantify the impact of the automatic strike zone. In order to accomplish this task, our team analyzed the difference between existing innings, grouped by their mistake characteristics, and created a predictive model to allow us to analyze how a game, if called correctly, may have ended differently.

Results: Existing Discrepancies



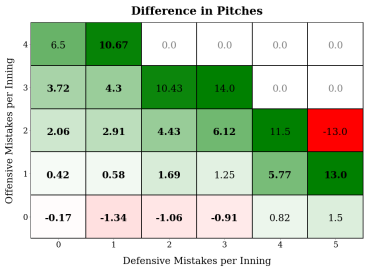
Note: Δ values calculated as [X Mistake Inning Average − 0 Mistake Inning Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.

For each at bat, we took the net mistakes (offensive minus defensive) in order to get a broad strokes picture of the impact. Strikeouts and walks had clear and opposite trends, so we classified at bats by the combination of offensive and defensive mistakes they contained, so we could get a clearer picture of how each type of mistake was changing the outcomes.



Results: Predicted Changes

The following metrics are calculated as the original innings **with mistakes** minus the new, **corrected** innings.



Pitches follow an anticipated trend; fewer pitches are thrown after correcting offensive mistakes but more are thrown after correcting defensive mistakes.

Conclusions & Moving Forward

Our analysis of existing innings shows that umpires have a tangible effect on the outcome of at bats, and our corrected inning pitch trend leads us to conclude that the model does eliminate some of the unfair advantages from mistaken calls.

This introduces a lot of interesting possibilities about the implementation of the automatic strike zone, and could enable future experiments that allow for the manipulation of the shape or consistency of the zone.