Expected Value Modeling of Batted-Ball Outcomes in LSU Softball

Omar Badawi, Charles Bloss, Thomas Emerick, Sebastian Escobar-Mesa, Connor Klebrowski

Department of Mathematics, Louisiana State University,
Baton Rouge Advisors: Dr. Nadejda Drenska & Dr. Peter Wolenski

Introduction

In softball, expected statistics quantify the quality of contact, independently of the batted-ball's observed outcomes, and form the foundation of modern hitting analytics. Two key measures are expected batting average (xBA) and expected weighted on-base average (xwOBA), which estimate the probability and run value of hits as functions of batted-ball characteristics such as how hard the ball is hit, exit velocity (EV), and the angle at which the ball comes off the bat, launch angle (LA).

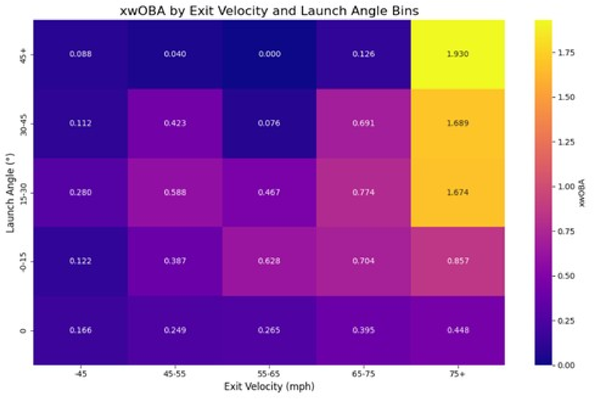
While expected statistics are now standard tools in professional baseball, these models remain underdeveloped at the collegiate softball level due to differences in data structure, sample size, and competitive environments. LSU Softball aims to leverage batted-ball tracking data to improve player development and game preparation.

In this project, we model the relationship between exit velocity, launch angle, and hit outcomes using a statistical learning pipeline originally trained on MLB Statcast data and later adapted to LSU Softball. Outcome probabilities are first estimated through k-nearest neighbors classification in exit velocity and launch angle space. These discrete probability fields are then smoothed using generalized additive models to obtain continuous probability surfaces, which are used to construct xBA and xwOBA directly from LSU Softball Trackman data.

Objectives

- Analyze the relationship between EV and LA on batted-ball outcomes using probabilistic classification methods.
- Construct LSU-specific xBA and xwOBA as functions of EV and LA.
- Estimate local outcome probabilities in EV-LA space using kNN classification.
- Regularize and smooth discrete probability estimates using Generalized Additive Models (GAM) to obtain continuous outcome surfaces.
- Develop interpretable visualizations of expected outcomes across EV-LA space for applied use in hitter evaluation and player development.

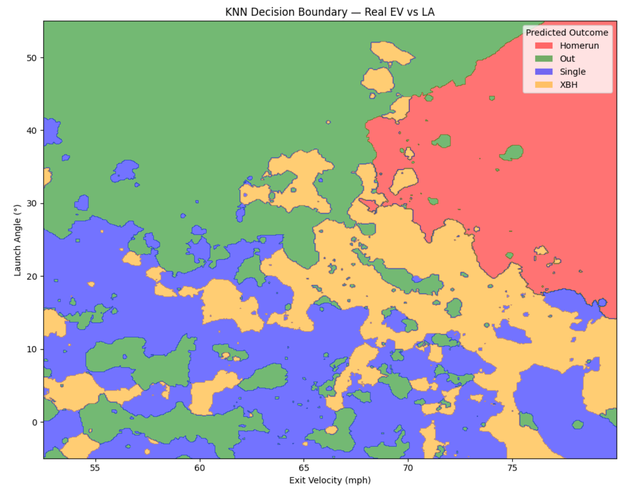
Binned xwOBA Surface



LSU xwOBA by EV and LA (Binned)

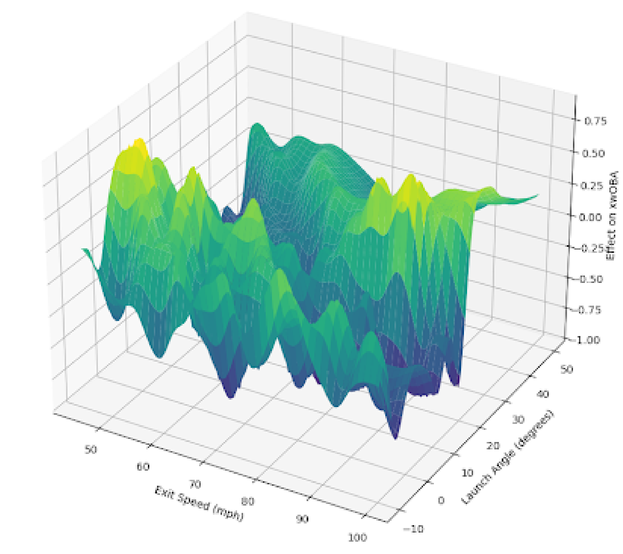
LSU-specific expected run value computed from smoothed outcome probabilities across EV–LA bins.

Decision Boundary and xwOBA Surface



kNN Decision Boundary in EV-LA Space

Local outcome classification from k-nearest neighbors, showing dominant batted-ball outcomes (out, single, extra-base hit (XBH), home run (HR)) by contact profile.



xwOBA by Exit Velocity and Launch Angle Bins

Smoothed EV–LA interaction surface from the GAM showing the combined effect of EV and LA on xwOBA.

Mathematical Formulation

Let (EV, LA) denote exit velocity and launch angle, with outcome

$$Y \in \{\text{Out}, \text{Single}, \text{XBH}, \text{HR}\}.$$

Local Outcome Probabilities (kNN):

$$\hat{p}_c(EV, LA) = \frac{1}{k} \sum_{j \in \mathcal{N}_k(EV,LA)} \mathbf{1}_{\{Y_j = c\}}.$$

Expected Batting Average (xBA):

$$\text{xBA}(EV, LA) = \hat{p}_{\text{Single}} + \hat{p}_{\text{XBH}} + \hat{p}_{\text{HR}}.$$

Expected Weighted On-Base Average (xwOBA):

$$\text{xwOBA}(EV, LA) = \sum_c w_c \, \hat{p}_c(EV, LA).$$

Continuous Surface Estimation (GAM):

$$\text{xwOBA}(EV, LA) = \beta_0 + f_1(EV) + f_2(LA) + f_3(EV, LA).$$

In these equations, w_c denote fixed linear outcome weights, and β_0 is the global intercept term. This framework combines local probabilistic classification with smooth nonparametric estimation to model expected offensive value across contact space.

Model Performance

The GAM used to estimate LSU-specific xwOBA achieves strong predictive accuracy:

$$\text{RMSE} = 0.163, \quad \text{MAE} = 0.099, \quad R^2 = 0.944.$$

The kNN outcome classifier attains a mean cross-validated F1 score of

$$0.831 \pm 0.020,$$

with highest accuracy for outs and home runs.

These results validate the kNN → GAM framework for local classification and continuous expected-value estimation.

Moving Forward

- Integrate model outputs into hitter evaluation and player development workflows.
- Validate model stability using additional seasons of LSU Softball Trackman data.
- Extend the framework to situational expected-value modeling.
- Explore player-specific calibration and opponent-adjusted expected statistics.
- Incorporate weather-adjusted corrections to account for environmental distortions.

Acknowledgments

- We thank Dr. Zach Jermain and the LSU Softball staff for the project opportunity and insights.
- We thank Dr. Nadejda Drenska and Dr. Peter Wolenski for their guidance on this project, and Fernando Heidercheidt, Maganizo Kapita, and Matthew Lemoine for their assistance throughout this work.

References

[1] T. Petersen, "Augmenting statcast expected batting average (xBA) with sprint speed," *MLB Technology Blog*, 2019. Published September 21, 2019.

[2] T. Nestico, "Modelling xwOBA (with KNN)," *Medium: Thomas Nestico*, 2024. Published February 14, 2024.

[3] W. Zhao, V. S. Akella, S. Yang, and X. Luo, "Machine learning in baseball analytics: Sabermetrics and beyond," *Information (MDPI)*, vol. 16, no. 5, 2025.