Uncertainty Analysis of Horizontal Break and Induced Vertical Break in Major League Baseball

Austin Gardiner

November 3, 2025

1 Introduction

In baseball, modern player performance analysis relies heavily on advanced measurement systems that track the trajectory and spin characteristics of pitches. Systems such as TrackMan, Rapsodo, and HawkEye use combinations of high-speed cameras, radar, and image processing techniques to track the ball and its spin throughout each pitch. Pitchers have increasingly sought to maximize pitch movement arising from Magnus and wake effects to make the ball more difficult to hit. This movement, known as break, can be separated into *horizontal break* and *induced vertical break*.

Although pitch-tracking systems have improved in accuracy over time, all measurements remain subject to uncertainty due to limited temporal and spatial resolution, calibration, and data processing algorithms. Furthermore, computing horizontal and vertical break, as seen in Eqs. 1–3, requires identifying the precise moments when the ball leaves the pitcher's hand and when it crosses the plate. Understanding and quantifying these uncertainties is essential for accurately interpreting player performance and ensuring that data-driven decisions, such as player development, scouting, and in-game strategy, are based on reliable information.

This project aims to analyze available MLB pitch-tracking data to estimate the uncertainty associated with reported horizontal and induced vertical break. The analysis will apply the principles of error propagation through the Monte Carlo Method (MCM) and evaluate the uncertainty of polynomial fit coefficients to quantify total measurement uncertainty.

2 Objectives

The primary objectives of this project are:

1. To determine the uncertainty of ball position at any point during the pitch based on measurement noise and instrument precision limits.
2. To develop an algorithm to determine the release point and its associated uncertainty.
3. To calculate the initial horizontal and vertical velocities and quantify their uncertainties.
4. To perform a total uncertainty analysis using Monte Carlo uncertainty propagation.

3 Plan to Meet Objectives

The project will address each objective in the order presented. Analysis will focus on data collected by the HawkEye tracking system, which uses an array of high-speed cameras to record the baseball's three-dimensional position throughout each pitch. This camera-based data will serve as the basis for all uncertainty calculations and modeling steps.

3.1 Important Equations

The following equations describe horizontal and induced vertical break, as well as the time of flight of the baseball:

$$B_h = x_f - (x_0 + v_{0,x}\Delta t) \tag{1}$$

$$B_v = y_f - (y_0 + v_{0,y}\Delta t + \frac{1}{2}g\Delta t^2) \tag{2}$$

$$\Delta t = t_f - t_0 \tag{3}$$

3.2 Ball Position

The uncertainty in the ball's position will be estimated by analyzing the resolution, field of view, frame rate, and distance from the cameras to the pitching area. These parameters determine the precision limits of the measured trajectory data. The analysis will quantify how these instrumental factors propagate into spatial uncertainty for each position measurement.

3.3 Point of Release

To determine the point of release, the project will follow a method similar to that used by the HawkEye team. In this method, two curve fits are constructed: one describing the path of the pitcher's arm and another describing the ball's flight trajectory after release. The release point is determined by locating the intersection of these two fitted curves, and the release time is determined by interpolation. This process is shown in Fig. 1. The end point of the pitch will be defined as the intersection of the ball's flight path with a vertical plane located at the front edge of home plate. From these two points, the total time of flight and its associated uncertainty will be calculated.

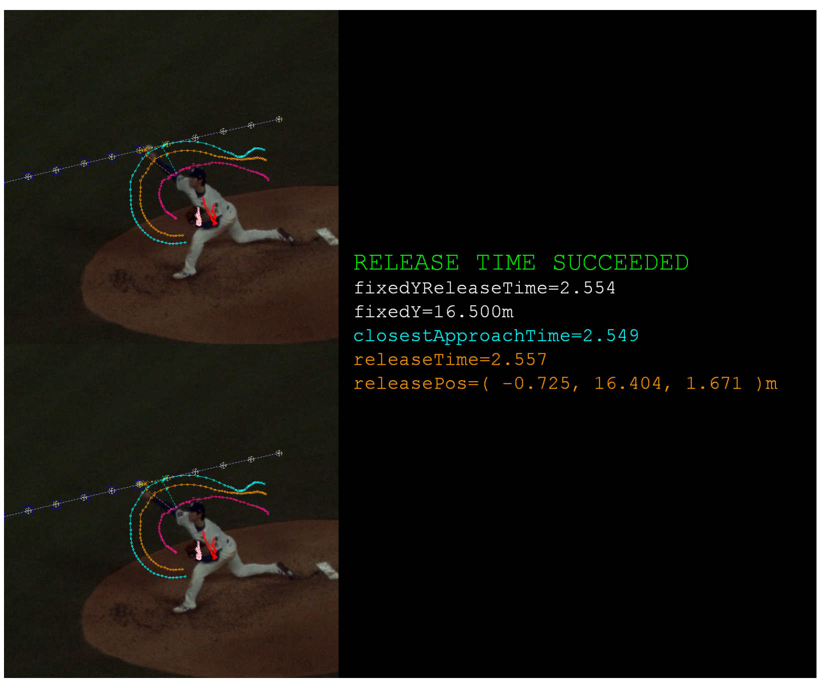

Figure 1: HawkEye release time calculation method

3.4 Initial Trajectory

The initial trajectory will be obtained by fitting polynomial curves to position-versus-time data in the horizontal (x) and vertical (y) directions. The corresponding initial velocities, $v_{0,x}$ and $v_{0,y}$, will then be used in Eqs. 1 and 2, respectively. Horizontal break can be modeled as the difference in current position from the initial linear trajectory as seen in Fig. 2. Induced vertical break must also account for gravitational acceleration. Thus, the vertical trajectory will be compared to the path of an object in free fall with the same initial vertical velocity. Induced vertical break along the flight path of a curveball is shown in Fig. 3.

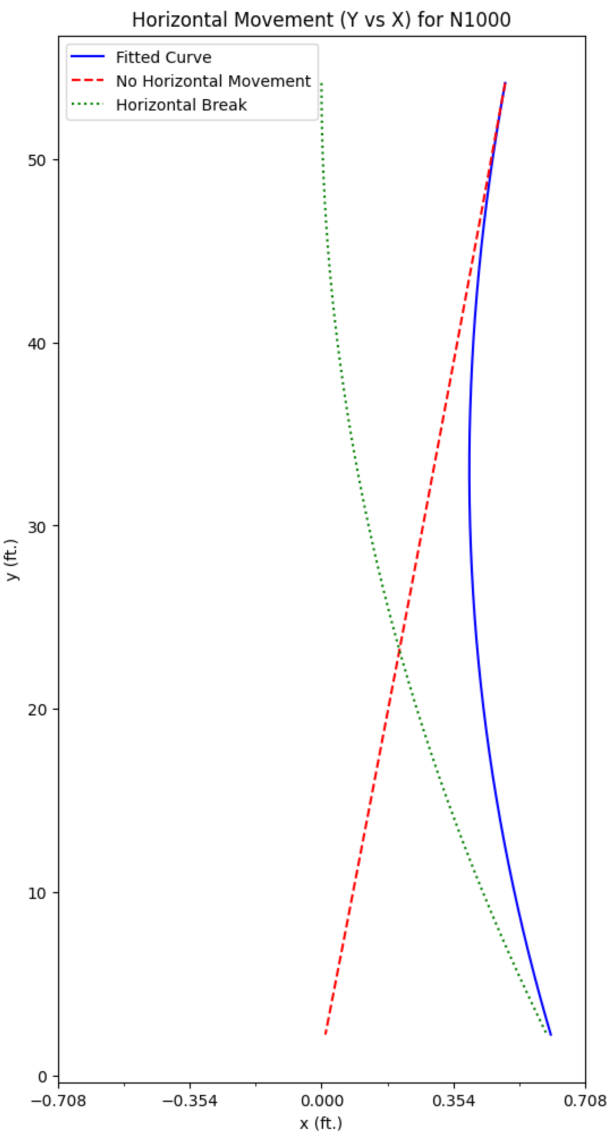

Figure 2: Horizontal break visualization

Uncertainties in the polynomial fit parameters will be determined by constructing a 3×3 α matrix:

$$\alpha_{l,k} = \sum_{i=1}^{N} \left[\frac{1}{u_{y_i}^2} x_i^{l-1} x_i^{k-1} \right] \tag{4}$$

The covariance matrix ϵ is then computed as the inverse of α. The diagonal elements of ϵ correspond to the variances of the fitted coefficients, from which the uncertainties are obtained:

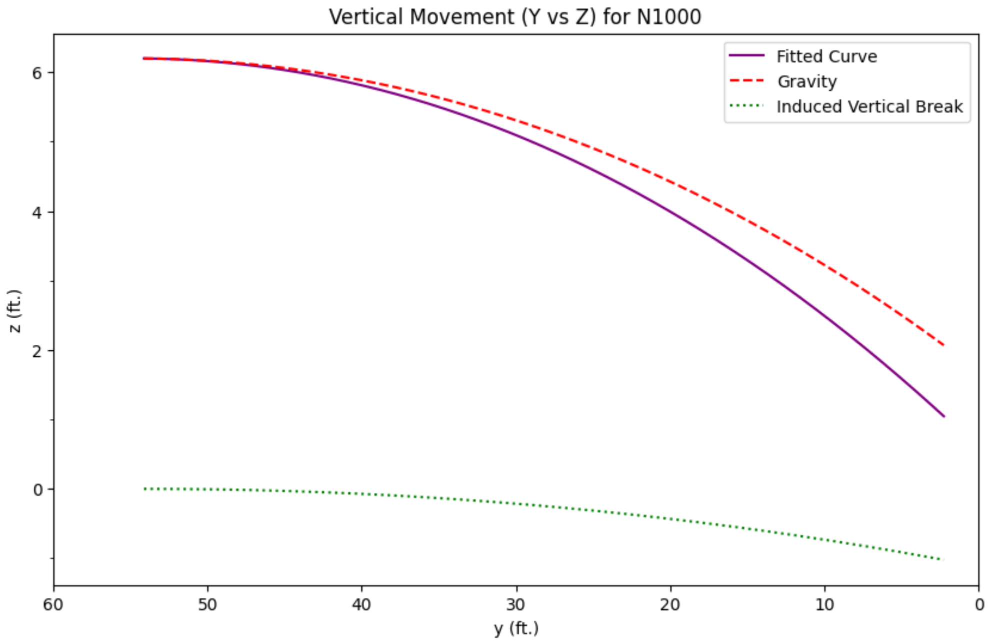

Figure 3: Induced vertical break visualization

$$\epsilon = \alpha^{-1} \tag{5}$$

Relative uncertainties for each fit parameter are found by dividing the standard uncertainties by their corresponding coefficients. Finally, the residuals are computed using Eq. 6 and compared with the uncertainties derived in Section 3.2.

$$r_y^2 = s_y^2 \approx \frac{1}{N-3}\sum_{i=1}^{N}(y_i - A_1 - A_2 x_i - A_3 x_i^2)^2 \tag{6}$$

3.5 Monte Carlo Method

With the uncertainties determined in the previous sections, a total uncertainty propagation will be conducted using the Monte Carlo Method. An algorithm will be developed to take as input the measured or calculated values of x_f, y_f, x_0, y_0, $v_{0,x}$, $v_{0,y}$, t_0, and t_f, along with their respective uncertainties. Each variable will be assigned either a uniform or Gaussian distribution, and the user will be able to specify the number of samples between 1,000 and 1,000,000. The resulting output will include a histogram showing the spread and shape of the propagated uncertainty distribution, which should reflect the underlying probability density function of the combined uncertainties. Equations 1 and 2 will serve as the Data Reduction Equations (DREs) for the Monte Carlo Method with Eq. 3 substituted in for Δt.