

Pitch Tunneling

Introduction

This project uses data collected during Major League Baseball games during the 2016 and 2017 seasons. The data is collected by Statcast, a system composed of radars and cameras that collects the movements of the ball as it is pitched to home plate. One branch of research stemming from this data is “Pitch Tunneling”. Pitch Tunneling is outlined in this article and the videos embedded therein:

<https://www.baseballprospectus.com/news/article/31030/prospectus-feature-introducing-pitch-tunnels/>.

Essentially, Pitch Tunneling is the idea that pitches that look similar out of the pitcher’s hand are likely to be more effective than those that look different. More specifically, pitches that are well-tunneled will be in similar locations at the “Tunneling Point”, but will vary widely in their location when they cross home plate. The Tunneling Point is the point at which the batter must decide whether or not to swing. **Throughout this project, we will be using the distance 23.8 feet from home plate as the "Tunneling Point".**

Goal

The goal of this project is to develop a metric or metrics to measure the effect of Pitch Tunneling on a pitcher’s performance as measured by each pitch’s Run Value. The Run Value of a pitch is the empirical estimate of how much that pitch is worth in runs. This value is provided for each pitch in the data. **The Pitch Tunneling metric or metrics are required to be included in whatever method or model that is created.**

Data

As previously mentioned, the data used in this project is MLB’s Statcast pitching data. This data contains every pitch for the 2016 and 2017 seasons for pitchers that threw more than 500 pitches during that span. A large portion of this project will be wrangling the data into a workable form to achieve the goal of the project. This data has been split into a training set and a test(validation) set where each pitcher has pitches in both sets. The variables available are as follows:

1. **PitchGuid** - Unique pitch identifier.
2. **GamePk** - Unique game identifier.
3. **batterID** - Unique batter identifier.
4. **pitcherID** - Unique pitcher identifier.
5. **PitchType** - Type of pitch thrown with the following encoding:
 - CH = Changeup
 - CU = Curveball
 - FC = Cutter
 - FF = Four-seam fastball
 - FO = Forkball

- FS = Splitter
 - FT = Two-seam fastball
 - KC = Knuckle curve
 - KN = Knuckleball
 - SC = Screwball
 - SI = Sinker
 - SL = Slider
6. **VenueID** - Unique identifier for the stadium in which the pitch is thrown.
 7. **islhp** - Whether or not the pitcher is left-handed.
 8. **islbh** - Whether or not the batter is left-handed.
 9. **Inning** - The inning in which the pitch is thrown.
 10. **IsTop** - Indicates if the pitch is thrown in the top or bottom of the inning.
 11. **Outs** - Number of outs in the current inning immediately preceding the batted ball.
 12. **Balls** - Number of balls in the count immediately preceding the batted ball.
 13. **Strikes** - Number of strikes in the count immediately preceding the batted ball.
 14. **ReleaseExtension** - The distance from the pitching mound towards home plate in which the pitch leaves the pitcher's hand. The mound is 60'6" away from home plate.
 15. **ReleaseAngle** - The vertical angle relative to the horizon at which the pitch was released.
 16. **ReleaseSpinAxis** - The axis of rotation for the ball at release given as an angle that reflects how the spin will influence the ball trajectory. Pure back-spin is 180 degrees, pure side-spin that puts the ball to the 1b side is 90 degrees, pure-side spin that pulls the ball to the 3b side is 270 degrees, and pure top-spin is 0 or 360 degrees.
 17. **ReleasePositionX** - Horizontal distance in feet left-right across home plate at which the pitch leaves the pitcher's hand. 0 is the center of home plate, negative numbers are inside on right-handed batters, positive numbers are outside on right-handed batters, and vice-versa for left-handed batters.
 18. **ReleasePositionY** - Horizontal distance in feet from the pitcher's mound to home plate at which the pitch leaves the pitcher's hand. Higher numbers indicate pitches further from home plate.
 19. **ReleasePositionZ** - Vertical distance in feet from the ground at which the pitch leaves the pitcher's hand
 20. **ReleaseSpeed** - Speed in MPH in which the pitch leaves the pitcher's hand.
 21. **ReleaseDirection** - The horizontal angle relative to the line from the center of the rubber to the tip of home plate at which the pitch was released. Positive towards 3B.
 22. **ReleaseSpinRate** - The spin rate in Rotations per Minute (RPM), of the pitch, measured at the time the ball is released.
 23. **TrajectoryHorizontalBreak** - The amount of horizontal break of the Pitch. Measured in the vertical plane at the front of home plate as the difference in the vertical position of the ball crossing and where the ball would have crossed if it had moved on a straight line following the direction in which it was released (i.e. if no acceleration acted on the ball). Positive for break towards 3B.
 24. **TrajectoryVerticalBreakInduced** - The amount of vertical break that the player induced on the pitch trajectory. This is found by compensating VerticalBreak by the amount of vertical break that can be attributed to gravity.

25. **TrajectoryZoneSpeed** - The speed of the ball as it crossed the vertical plane at the front face of home plate.
26. **TrajectoryZoneTime** - Time after the release when the ball passes the vertical plane at the front face of home plate.
27. **TrajectoryPolynomial[X,Y,Z][0, 1, 2]** - Polynomial coefficients for the [X,Y,Z] components of the pitch's trajectory. The following formula calculates the X-coordinate of a pitch at a given time t in the pitch's flight towards home plate:

$$X = TrajectoryPolynomialX0 + TrajectoryPolynomialX1 \times t + TrajectoryPolynomialX2 \times t^2$$

This same formula can be used to find the Y- and Z-coordinates of a pitch.
28. **TunnelTime** - The time it takes for the pitch to reach the Tunneling Point from the point the pitch is released.
29. **TunnelLocation[X,Y,Z]** - The coordinates of the pitch at the Tunneling Point.
30. **TrajectoryVerticalApproachAngle** - The angle relative to the horizon at which the ball crosses the vertical plane at the front face of home plate.
31. **TrajectoryHorizontalApproachAngle** - The angle relative to the line from the center of the rubber to the tip of home plate at which crosses the vertical plane at the front face of home plate. Positive for direction towards 3B.
32. **VerticalBreak** - The total amount of vertical break of the pitch. Vertical break is measured in the vertical plane at the front face of home plate as the difference in the vertical position of the ball crossing and where the ball would have crossed if it had moved on a straight line following the direction in which it was released (i.e. if no acceleration acted on the ball).
33. **TrajectoryLocation[X,Y,Z]** - The X-,Y-, and Z-coordinates of the pitch as it crossed the vertical plane at the front of home plate.
34. **AtBatNumber** - The number of the at bat for the game. This number will run from 1-the number of at-bats in the game sequentially.
35. **PitchNumber** - The pitch number of the at-bat.
36. **RunValue** - The Run Value associated with the pitch. This will ultimately be the dependent variable in this project.

Data Notes

- In this analysis, batterID, pitcherID, and the teams that they play for are not to be used beyond simple data organization.
- It is recommended that the data be split into the 4 different possible combinations of pitcher and batter handedness for analysis.
- It may be beneficial to split the data into two groups: 1.) Fastballs and sinkers, 2.) All other pitches.

Additional Materials

X, Y, Z Orientation

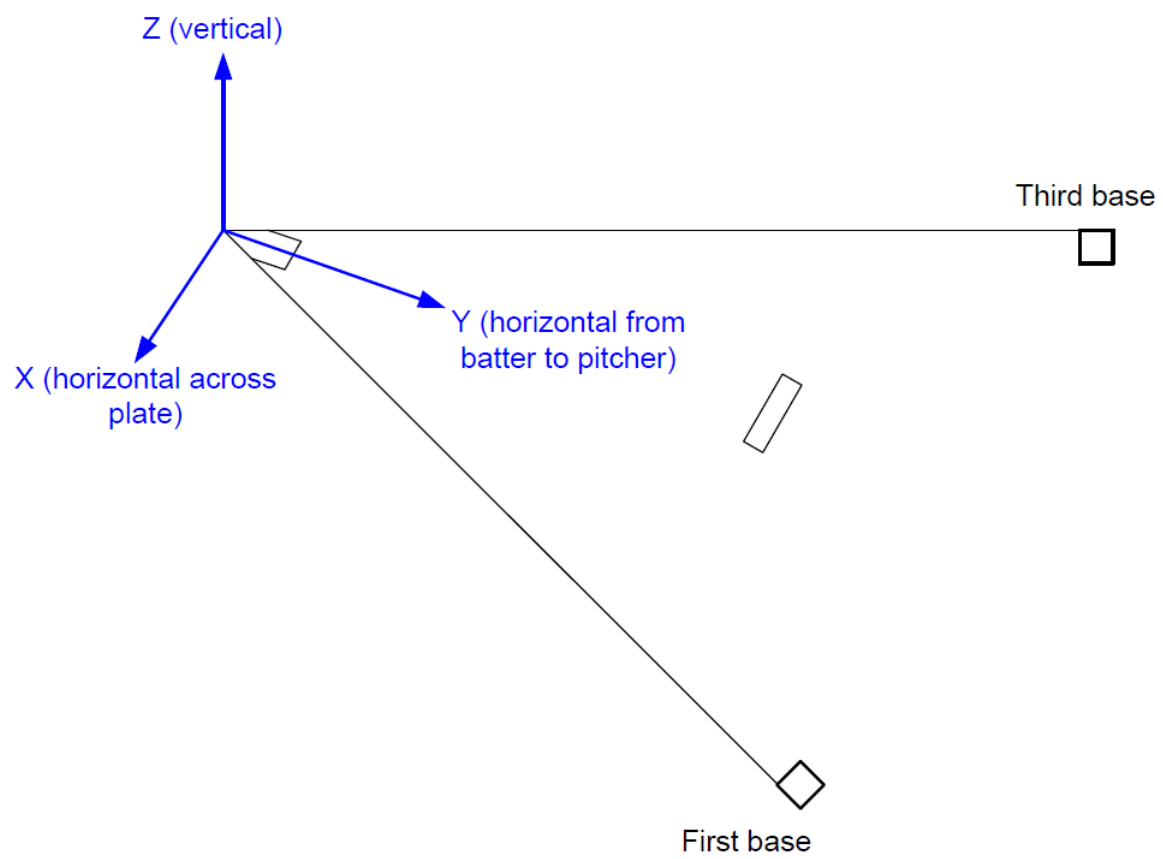


Figure 1: