Building MLB Matchup Data Using Pitch F/X

Michael Pursey

Jalgos Data Science Agency

`michael@jalgos.com`

July 27, 2015

Note to self: checkout kpsewhich command

Table 1: Glossary of Pitch F/X fields by Mike Fast

Field	Description
des	Brief text description of the result of the pitch: Ball; Ball In Dirt; Called Strike; Foul; Foul (Runner Going); Foul Tip; Hit by Pitch; In play, no out; In play, out(s); In play, run(s); Intent Ball; Pitchout; Swinging Strike; Swinging Strike (Blocked).
id	Unique identification number per pitch within a game. The numbers increment by one for each pitch but are not consecutive between at bats.
type	One-letter abbreviation for the result of the pitch: B, ball; S, strike (including fouls); X, in play.
x, y	Horizontal and vertical location of the pitch as it crossed home plate as input by the Gameday stringer using the old Gameday coordinate system. I'm not sure what units are used or where the origin is located. Note that the y dimension in the old coordinate system is now called the z dimension in the new Pitch F/X coordinate system detailed below.
start_speed	Pitch speed, in miles per hour and in three dimensions, measured at the initial point, y0. Of the two speeds, this one is closer to the speed measured by a radar gun and what we are familiar with for a pitcher's "velocity" .
end_speed	Pitch speed measured as it crossed the front of home plate.
sz_top	Distance in feet from the ground to the top of the current batter's rulebook strike zone as measured from the video by the Pitch F/X operator. The operator sets a line at the batter's belt as he settles into the hitting position, and the Pitch F/X software adds four inches up for the top of the zone.
sz_bot	Distance in feet from the ground to the bottom of the current batter's rulebook strike zone. The Pitch F/X operator sets a line at the hollow of the knee for the bottom of the zone.
pfx_x	Horizontal movement, in inches, of the pitch between the release point and home plate, as compared to a theoretical pitch thrown at the same speed with no spin-induced movement. This parameter is measured at y=40 feet regardless of the y0 value.
pfx_z	Vertical movement, in inches, of the pitch between the release point and home plate, as compared to a theoretical pitch thrown at the same speed with no spin-induced movement. This parameter is measured at y=40 feet regardless of the y0 value.
px	Left/right distance, in feet, of the pitch from the middle of the plate as it crossed home plate. The Pitch F/X coordinate system is oriented to the catcher's/umpire's perspective, with distances to the right being positive and to the left being negative.

Table 1: (continued)

Field (ctd.)	Description (ctd.)
pz	Height of the pitch in feet as it crossed the front of home plate.
x0	Left/right distance, in feet, of the pitch, measured at the initial point.
y0	Distance in feet from home plate where the Pitch F/X system is set to measure the initial parameters. This parameter has been variously set at 40, 50, or 55 feet (and in a few instances 45 feet) from the plate at different times throughout the 2007 season as Sportvision experiments with optimal settings for the Pitch F/X measurements. Sportvision settled on 50 feet in the second half of 2007, and this value of y0=50 feet has been used since. Changes in this parameter impact the values of all other parameters measured at the release point, such as start_speed.
z0	Height, in feet, of the pitch, measured at the initial point.
vx0, vy0, vz0	Velocity of the pitch, in feet per second, in three dimensions, measured at the initial point.
ax, ay, az	Acceleration of the pitch, in feet per second per second, in three dimensions, measured at the initial point.
break_y	Distance in feet from home plate to the point in the pitch trajectory where the pitch achieved its greatest deviation from the straight line path between the release point and the front of home plate.
break_angle	Angle, in degrees, from vertical to the straight line path from the release point to where the pitch crossed the front of home plate, as seen from the catcher's/umpire's perspective.
break_length	Measurement of the greatest distance, in inches, between the trajectory of the pitch at any point between the release point and the front of home plate, and the straight line path from the release point and the front of home plate, per the MLB Gameday team. John Walsh's article "In Search of the Sinker" has a good illustration of this parameter.
sv_id	A date/time stamp of when the Pitch F/X tracking system first detected the pitch in the air, it is in the format YYMMDD_hhmmss.
pitch_type	Most probable pitch type according to a neural net classification algorithm developed by Ross Paul of MLBAM.
type_confidence	Value of the weight at the classification algorithm's output node corresponding to the most probable pitch type, this value is multiplied by a factor of 1.5 if the pitch is known by MLBAM to be part of the pitcher's repertoire

Field (ctd.)	Description (ctd.)
zone	Appears to correspond to the location of the pitch based on the boxes into which the Gameday app divides the strike zone for its hot/cold zone graphics
nasty	The "nasty" field is presumably a crude attempt to calculate how hard to hit a particular pitch was, on a scale of 0-100. My initial cursory look at the data indicates that they are calculating the "nasty" factor mostly based on the location of the pitch, a linear calculation of how close it is to the edges and away from the heart of the zone. For the fastball, MLBAM does not appear to be including anything related to the movement or speed of the pitch into the "nasty" factor. For the curveball, they appear to be rating sweeping curveballs as significantly more nasty than 12-to-6 NA more than a curiosity
cc	comment field that appears to my[Mike Fast] highly-trained eye to be auto-generated, probably also based on the hot/cold zone information that MLBAM tracks

Ideas for physical pitcher measures:

- Release points (c.f. Mike Fast article on calculation method): consistency within games& accross games (clustered by pitch type?), location by pitch type