

# Fast Balls

August 2, 2007

## Glossary of the Gameday pitch fields

Posted by Mike Fast under [gameday](#), [pitchfx glossary](#).

[11] [Comments](#)

At Major League Baseball's [Gameday](#) data website, the PITCHf/x data is included in the inning and pbp/pitchers XML files. What follows is an explanation of the attributes of the *pitch* element within these XML files.

At stadiums without the PITCHf/x system installed, the *pitch* element includes only five attributes:

- **des**: a 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**: a unique identification number per pitch within a game. The numbers increment by one for each pitch but are not consecutive between at bats.
- **type**: a one-letter abbreviation for the result of the pitch: B, ball; S, strike (including fouls); X, in play.
- **x, y**: the 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 PITCHf/x coordinate system detailed below.

Stadiums with the PITCHf/x camera system have an additional twenty attributes recorded in the *pitch* element:

- **start\_speed**: the 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**: the pitch speed measured as it crossed the front of home plate.
- **sz\_top**: the distance in feet from the ground to the top of the current batter's rulebook strike zone as measured from the video by the PITCHf/x operator. The operator sets a line at the batter's belt as he settles into the hitting position, and the PITCHf/x software adds four inches up for the top of the zone.
- **sz\_bot**: the distance in feet from the ground to the bottom of the current batter's rulebook strike zone. The PITCHf/x operator sets a line at the hollow of the knee for the bottom of the zone.
- **pfx\_x**: the 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.

- **px\_z**: the 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  $y_0$  value.
- **px**: the left/right distance, in feet, of the pitch from the middle of the plate as it crossed home plate. The PITCHf/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.
- **pz**: the height of the pitch in feet as it crossed the front of home plate.
- **x0**: the left/right distance, in feet, of the pitch, measured at the initial point.
- **y0**: the distance in feet from home plate where the PITCHf/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 PITCHf/x measurements. Sportvision settled on 50 feet in the second half of 2007, and this value of  $y_0=50$  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**: the height, in feet, of the pitch, measured at the initial point.
- **vx0, vy0, vz0**: the velocity of the pitch, in feet per second, in three dimensions, measured at the initial point.
- **ax, ay, az**: the acceleration of the pitch, in feet per second per second, in three dimensions, measured at the initial point.
- **break\_y**: the 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**: the 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**: the 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.

Three new fields were added to the pitch element for 2008:

- **sv\_id**: a date/time stamp of when the PITCHf/x tracking system first detected the pitch in the air, it is in the format `YYMMDD_hhmmss`.
- **pitch\_type**: the most probable pitch type according to a neural net classification algorithm developed by Ross Paul of MLBAM.
- **type\_confidence**: the 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.

Resources for this glossary included the following:

- Comments, including those by MLB.com's Director of Stats Cory Schwartz at Tom Tango's blog [THE BOOK](#).
- The [MLB Gameday blog](#).
- The article "[How to Use MLB Gameday Data](#)" by Anthony at Friar Watch.
- Comments to Joe P. Sheehan's article "[Enhanced Gameday](#)".

Note: Shortly after I wrote this, I found that Dr. Alan Nathan has published a [better glossary](#) than mine at his excellent [Physics of Baseball](#) site.

His freshman physics [lectures on the Physics of Baseball](#) at the University of Illinois is also an excellent primer to understanding the calculations surrounding baseball trajectories.

If you need to convert MLB.com's player ID's into names or Lahman database player ID's, you can consult my [list of player ID's](#).

Another common question is about the meaning of the BRK and PFX numbers reported in the Gameday application. Here's what I wrote about the subject on another website:

*There are three main forces acting on a spinning baseball: gravity, drag, and the spin force (also called the Magnus or lift force).*

*The drag force mainly acts to slow a pitch down, it doesn't have much effect on the movement/break of a pitch, except for very, very slowly spinning pitches (i.e., knuckleballs).*

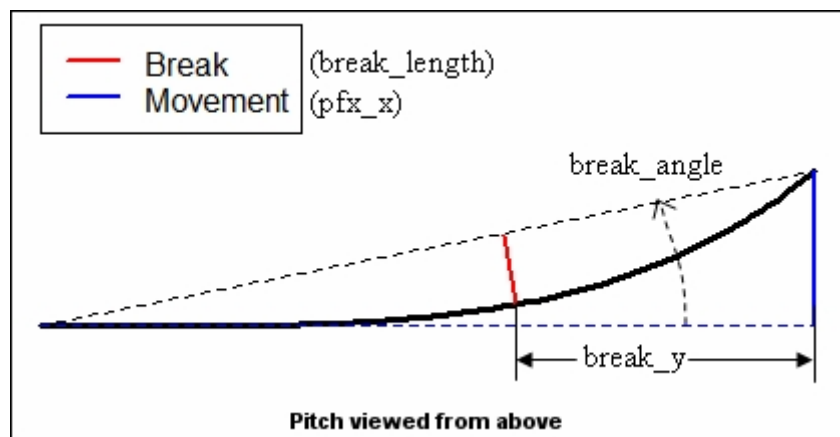
*The force of gravity is the same on all pitches, but it has a greater effect on the movement of slow pitches because it has longer to act on them before they reach the plate. Curveballs and changeups drop more due to gravity than fastballs do because they are slower pitches.*

*Finally, the spin force acts differently on fastballs and curveballs, as the Gameday folks described. Because a fastball is thrown with backspin, the spin force pushes the ball up, counteracting to some extent the force of gravity that is pulling the ball down. This makes the fastball trajectory straighter. Because a curveball is thrown with topspin, the spin force pushes the ball down, reinforcing gravity which is also pushing it down. This makes the curveball drop even more.*

*Thus, the curveball trajectory has a big bend and the fastball trajectory is relatively straight. The amount of bend in the trajectory is what is being measured by the BRK parameter on Gameday.*

*The amount of deflection by the spin force is what is being measured by the PFX parameter on Gameday. This PFX deflection is mostly upward for a fastball, meaning that it counteracts roughly 10 or so inches of the drop due to gravity, and the PFX deflection is mostly downward for a curveball, meaning that it adds an additional 6 or so inches of drop in addition to that from gravity.*

Additionally, here is a diagram, adapted from [John Walsh](#), that illustrates the break parameters:





[Report this ad](#)



[Report this ad](#)

## 11 Responses to “Glossary of the Gameday pitch fields”

1. [Alan Nathan](#) Says:



August 7, 2007 at 3:30 pm

Thanks for the plug about the glossary. I have just written a brief article that is in the “nearly ready for prime time” category. You can download this article at <http://webusers.npl.uiuc.edu/~a-nathan/pob/Analysis.pdf> Before I go public with it, I would appreciate comments from any of you who are interested.

[Reply](#).

2. [mrflip](#) Says:



November 25, 2007 at 9:38 am

I made a [drawing illustrating the trajectories and the various data fields](#) with Sketchup (Google’s free 3-d program) ([PNG preview](#)). It’s free for all to use as they please.

I believe that the pfx\_x and pfx\_z offset endpoints are not at the strike zone plane ( $y=1.417\text{ft}$ ) but rather at the dragless ball's position when the actual pitch crosses the strike plane. (That is, their positions at the same \*time\*, not the same \*final y\* value). The ending y position for the dragless path is somewhere near or even behind home plate: if you parameterize by y you have to find the dragless ball's terminal position from the actual pitch's strikezone-crossing time. I left a [messy mathematica notebook](#) (PDF) with my calculations up there, if that's useful. This, at least, made my calculations come out to agree with the data files where nothing else would.

[Reply](#).

3. Mike Fast Says:

[November 25, 2007 at 1:44 pm](#)

Flip, thanks for the link to your drawing.

I can't see anything in the PDF file, and I don't have Mathematica, so I can't see your calculations.

The calculations all worked out for me to replicate the pfx\_x and pfx\_z numbers from the other PITCHf/x data when they were calculated at  $y=1.417$  feet. However, you have to include the effect of drag, i.e., the y acceleration.

[Reply](#).

4. [Back to the Grind « The Mockingbird](#) Says:

[July 3, 2008 at 12:00 pm](#)

[...] the string that attaches both ends and the middle of the bow (here's a graphic stolen from Mike Fast, who adapted it from Mike [...])

[Reply](#).

5. [The Curious Case of Manny Parra | FanGraphs Fantasy Baseball](#) Says:

[December 19, 2009 at 5:46 pm](#)

[...] Does Parra throw "fat" strikes? I'm not Pitch F/X maestro, but I did my best to try and quantify what might constitute a fat pitch. Pitch F/X keeps track of a number called "px." Here's Mike Fast's definition: [...]

[Reply](#).

6. [Working with the MLB Gameday dataset: Setup and Tools « Mining the Diamond](#) Says:

[March 3, 2013 at 10:13 am](#)

[...] pitch: This is where most of the magic happens. Contains the location and speed of every pitch, as well as whether it was called a ball/strike or put in play. There are many other fields here worth exploring, such as the estimated pitch type. For more information about the pitch fields, you can check out this diagram and this glossary. [...]

[Reply](#).

7. [PitchFx: Pitcher DNA – Tim's Prior](#) Says:

[March 31, 2013 at 11:09 pm](#)



[...] For my first exploration into the data of baseball, I wanted to take a look at pitchFx data, which gives an incredible wealth of data on pretty much every pitch thrown in a major league ballpark. The data is plentiful, descriptive, multidimensional, and gives us an unprecedented amount of quantitative information about each and every pitch thrown. If you've ever followed a game on GameCast, you have seen the pitchFx data: in addition to delivering first order stats about the pitch like velocity, location, and pitch type – pitchFx goes deeper and measures crazy stuff like horizontal movement, break, spin and acceleration. For a good glossary of the full observations you get from this data you can have a look at this blog post. [...]

Reply.

8. Pitch Classification with Mclust | Exploring Baseball Data with R Says:

January 12, 2015 at 9:36 am

[...] these measures would help us to identify what this would look like in terms of pitch types. You can start here at Mike Fast's blog with a glossary of what each of these variables [...]

Reply.

9. Analyzing Jake Arrieta's Perfect Game – Baseball Data Science Says:

August 7, 2016 at 3:35 pm

[...] of pitch movement is tricky and, in my opinion, not strongly intuitive, although it is useful. This glossary – which is somewhat dated but still appears to be accurate – provides an explanation of [...]

Reply.

10. In Praise of the Rich Hill Curveball – Sitting Low 80's Says:

November 27, 2016 at 5:28 pm

[...] Some rather technical definitions are necessary before we proceed. According to Mike Fast's PITCHf/x glossary, break angle is the angle at which the ball breaks from the catchers perspective; the greater the [...]

Reply.

11. Predict the Pitch ! Fun with Baseball Research | Data Central - With Harini Kannan Says:

March 23, 2017 at 1:53 pm

[...] <https://fastballs.wordpress.com/2007/08/02/glossary-of-the-gameday-pitch-fields/> [...]

Reply.



[Blog at WordPress.com.](https://fastballs.wordpress.com/2007/08/02/glossary-of-the-gameday-pitch-fields/)