What the Heck is PITCHf/x?

by Mike Fast

There is a revolution underway in baseball. It is a quiet revolution, not as controversial perhaps as 100-pitch limits for starting pitchers, but in the end, the changes it brings may be more influential.

The revolution debuted on Oct. 4, 2006, in an American League Division Series game between the Oakland Athletics and Minnesota Twins at the Metrodome. Most of us had no idea that, as Esteban Loaiza threw a first pitch strike to Twins leadoff batter Jason Kendall, Sportvision's PITCHf/x cameras were tracking the trajectory of the pitch, identifying its speed, break and location in real time and transmitting this information for use by broadcasters and Major League Baseball's Gameday web application.

Sportvision is a TV sports broadcast effects company based in Mountain View, Calif. Its engineers are most famous for creating the glowing FoxTrax hockey puck and the yellow first-down line superimposed on football broadcast video. Its current bread and butter product is GPS tracking of race cars during NASCAR races, and it provides this information to broadcasters and online to race fan subscribers. So when Sportvision turned its attention to baseball, it brought a strong pedigree in applying technology to sports broadcasts to bring fans new insight into their sports.

To track the flight of each major league pitch precisely, Sportvision has installed a pair of cameras in each stadium, in the stands above home plate and first base. Twenty-eight of these installations occurred during the 2007 season, and the other two, in Baltimore and Washington, were completed before to the 2008 season. Detailed tracking data were recorded for about a third of the pitches thrown in the major leagues in 2007 and more than 95 percent of the pitches in 2008 and 2009. Sportvision made these data available in real-time to Major League Baseball Advanced Media (MLBAM) for use in its online Gameday application and to broadcasters such as ESPN for its K-Zone strike zone graphic.

How Do We Get The Data?

Sportvision's tracking software takes the approximately 20 images of the baseball acquired during its flight from the pitcher's hand to home plate, determines the 3-D ball location on the baseball field from each image, and finds the pitch trajectory that best fits these 20 or so points along the path of the pitch. The software assumes that the baseball experiences constant acceleration; i.e., that the forces on the ball do not change in any meaningful way during its flight. This is a good assumption for rapidly spinning baseball pitches, and almost all pitches fit this criterion. The notable exception is the knuckleball, which spins very slowly, causing the drag force on the baseball to change depending on the seam orientation. However, it turns out that the PITCHf/x constant-acceleration model works fairly well for knuckleballs, too, with only slightly larger errors than for other pitch types.

What information does Sportvision and MLBAM collect about the game? We already mentioned that the PITCHf/x tracking software calculates the position, velocity and acceleration of the pitched baseball in three dimensions. It also puts a date-time stamp on each pitch. The PITCHf/x operator is a Sportvision employee on site at each baseball game. The operator is responsible for monitoring the calibration of the system during the game and for recording the top and bottom of the strike zone for every batter from the center field camera video.

Other information is entered in real time by a stringer, an MLBAM employee also at the game, into software for transmission to Gameday. That person records the result of each pitch (for example, ball, called strike, swinging strike, foul ball or in play). The stringer notes any pinch hitters or defensive substitutions and records the result of each at-bat (for example, single or fly out), the movement of baserunners, and the location on the playing field where the batted ball is fielded. MLBAM software attempts to classify the type of pitch that was thrown (for example, fastball or slider).

All of this data is available free on the MLBAM GameDay website. As a result, between the PITCHf/x data on the trajectory of the pitch and the MLBAM stringer's record of game events, we have a wealth of information on every baseball game.

Differences Between Coarse And Fine Data

With this abundance of fine-grained data, particularly the detailed record of the speed and movement of every pitch, a new analytical approach must be crafted and new common language developed for discussion among fans and analysts. Our current language of baseball has its roots in the 19th century with Henry

Chadwick, inventor of the box score. Games have been described in terms of hits and outs, bases and runs. These are very clear but coarse measuring sticks, but they are the ones we have come to understand. We can converse in that language without needing to explain the context. Debates about the merits of various analytical approaches do not require arguments about foundational principles. Instead, the validity of the base-out counting system and much of the long analytical history built atop it is assumed.

Within the last decade, the availability of detailed fielding data has begun to shift this paradigm, if only perhaps at the margin. It is widely recognized that fielding is not well measured by simply recording which fielder recorded each out. More detailed data on the path of the batted ball are desirable for determining which fielder should be assigned credit or blame for fielding or not fielding a ball and how much credit or blame he should get based on the difficulty of the play. However, since most of this data is proprietary, the impact on the language of the game has so far been small.

With the wide availability of PITCHf/x data, however, a seismic shift in the language of baseball and its analysis is beginning. The description of a pitch is now on a continuum between fast and slow, breaking left or right, up or down. How do we make sense of these detailed data? Where is the context?

Radar gun readings have established some context for the speed of a pitch; as a result, that information from PITCHf/x has been the most quickly adopted and easily understood, although no one really knows yet what a difference 1 mph of speed on his fastball means to a pitcher's effectiveness. These fundamental concepts will have to be investigated, understood and communicated over time. Almost every baseball fan has an intuitive understanding of the difference between a .260 hitter, a .290 hitter and a .320 hitter. One is average, one is good, and one is among the best in the game. I suspect some day we will have a similar intuitive understanding about pitch speeds. However, that context is only now emerging. We have so much new information that it is difficult to assimilate it quickly. We should remember that our current analytical framework of understanding was developed over more than a century.

The new detail we have about pitch movement is particularly challenging in that we have almost no common lexicon or understanding to draw from. What does it mean, if anything, when the movement on a pitcher's fastball changes by a few inches from one game to another? What does "movement on a pitcher's fastball" even mean? This context, too, must be developed over time.

Spin Deflection

I'm not a fan of the term "movement" to describe a pitch. How a ball moves depends on its initial velocity and the forces on the ball in flight, which include gravity, drag, and the Magnus, or spin, force. It has become popular for analysts to use the term "movement" to refer to the deflection of the ball due to the spin force. I prefer to refer to it as spin deflection.

Speed and the spin characteristics of a pitch are our primary ways to identify what type of pitch a pitcher threw. A pitcher's grip and wrist motion on release apply spin to the baseball, and this spin produces force on the baseball that causes it to veer. The topspin on a curveball causes it to veer downward more than it would from gravity alone. The backspin on a fastball causes it to drop less than it would due to gravity. Most pitches also have some sidespin, which causes the ball to veer to the left or right.

The Physics Of Baseball

One of the more fascinating results of the PITCHf/x data set is the extent to which it has expanded our knowledge of the physics of baseball. Dr. Robert Adair is well known for his excellent book, *The Physics of Baseball*, which he wrote in 1987 as the result of research he undertook at the behest of then National League president Bart Giamatti. In fact, Adair served as an adviser to Sportvision in the early days of PITCHf/x development.

However, our understanding of the physics of pitched and batted baseballs has improved immensely simply through the availability of such a large, high-quality data set. We have learned a great deal about the drag and spin forces on a baseball. Dr. Alan Nathan has lead the investigation into many topics of interest, along with a crowd of others, and you can read more at his Physics of Baseball website (see References section).

Gameday Application

Many people's first exposure to PITCHf/x is through MLBAM's Gameday application. (For those who are unfamiliar with Gameday, it lets fans follow the game online. It falls somewhere between a TV broadcast and box score, attempting to model and describe the game in detail without showing actual game video or audio.)

At the center of the Gameday application is a display of pitch trajectories from the pitcher to the batter for the current at-bat. These pitch trajectories are based on the data taken in real time from PITCHf/x. Below the pitch trajectory display, Gameday gives additional information about each pitch, including the speed, BRK (break), PFX (spin deflection) and the pitch type.

The speed of the pitch 50 feet from home plate is displayed here. This is not the maximum speed of the pitch. That occurs as the ball leaves the pitcher's hand, which typically happens around 55 feet from home plate. However, the 50-foot distance for reporting pitch speed was chosen so that the PITCHf/x speeds would match most closely to typical radar gun readings.

The most confusing aspects of the PITCHf/x data displayed in Gameday are the two parameters, BRK and PFX, each measured in inches. Break refers to the maximum bend in the pitch. The arc of a curveball bends much more than a fastball. Break is useful in that sense, and perhaps that sense only, but it has found little use in the analytical community.

PFX is related closely to the parameters preferred by the analytical community for describing and identifying pitches, but it has a fatal shortcoming. PFX, as presented on Gameday, refers to the total amount of spin deflection, but it tells the viewer nothing about the direction of the spin deflection, which is a critical piece of information. Totally different pitches may have similar PFX values. I recommend ignoring the PFX value while watching Gameday.

The pitch type displayed on Gameday is produced by a neural-net algorithm MLBAM developed to attempt to identify pitch types in real time. It's correct much more often than it's wrong, but it's far from perfect. Consider it sufficient for entertainment value and even for surface-level analysis of pitchers. Deeper and more accurate analysis usually requires more accurate pitch types.

The Strike Zone

One of the most popular uses of PITCHf/x data is to evaluate umpires' ball-strike calls. Fans love to blame the umpire when he hurts their team with a call at a critical juncture of the game. Now PITCHf/x, by identifying the exact location where the ball crossed the plate, gives them the ammunition they need to fire off a half-educated missive against the men in blue.

A common misconception among novice umpire evaluators is that a strike call a couple of inches past the outside edge is a sure sign that the umpire was out to get the batter and his team. In an ideal world, umpires might be able to call the exact zone described in the official rules. In reality, it's a complex classification task, one that umpires typically address by setting up behind the inside edge of home plate. This helps them call the inside edge of the strike zone fairly accurately but at the expense of inaccuracy on the outside edge. A typical strike zone called by most umpires will include strike calls a few inches off the outside edge of the plate.

Major League Baseball uses a human-audited version of the PITCHf/x data to grade the umpires, but this version of the data is not available publicly.

Pitch Classification

An important area of ongoing research is in the classification of pitch types. As previously mentioned, MLBAM classifies pitch types based on its own algorithm for use in Gameday; it's been doing this classification since the beginning of the 2008 season. One approach that is suitable for much research that requires pitch types is to use the MLBAM classifications. It is accurate enough for most work that involves differentiating between fastballs and off-speed pitches. Another approach is to group the pitches into bins by speed and spin deflection and use these bins without regard to a particular label, such as "two-seam fastball" or "slider." Both approaches have merit in ease of use and probably do a decent job of approximating what the hitter sees, particularly at the season level.

However, in examining pitching strategies, evaluating pitchers or examining a single game in detail, pitch classifications need to be more attuned to what the pitcher is actually doing rather than taking a "close enough" approach. It can be fascinating to attempt to get inside the pitcher's head and understand his craft. His various pitch types are his tools, and correctly identifying them is key to understanding his strategy, his strengths, and his weaknesses.

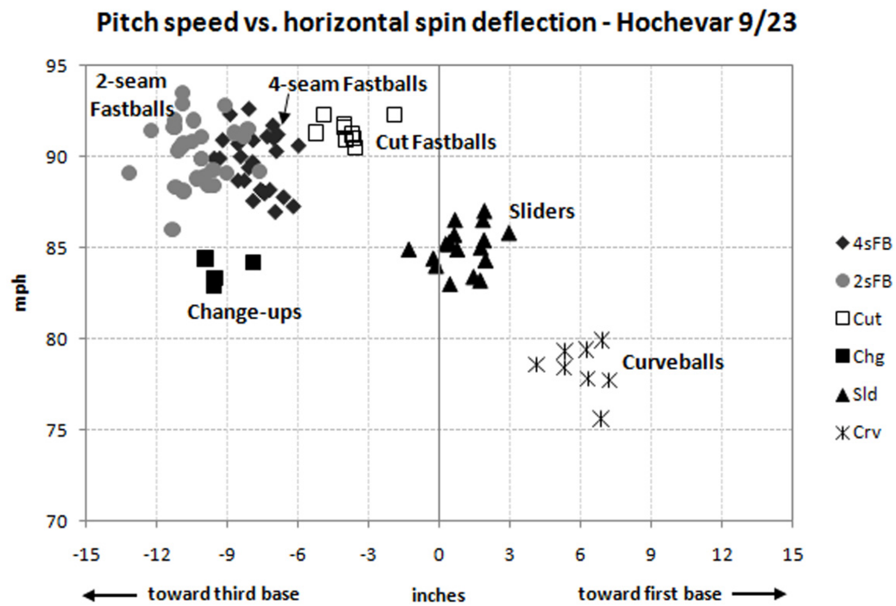
How does one identify pitch types? Because pitchers are so different, there's no perfect one-size-fits-all approach, but there are some helpful guidelines. The graphs of Luke Hochevar's pitches from his September 23rd start against the Red Sox on the next page illustrate some useful guidelines.

I am a strong advocate of looking at a pitcher's data at the game level rather than the season level when classifying pitches. Due to inconsistencies (usually small but sometimes big) between PITCHf/x systems at different parks and sometimes at the same park at different points in the season, it can be difficult to distinguish

(continued on page 5)

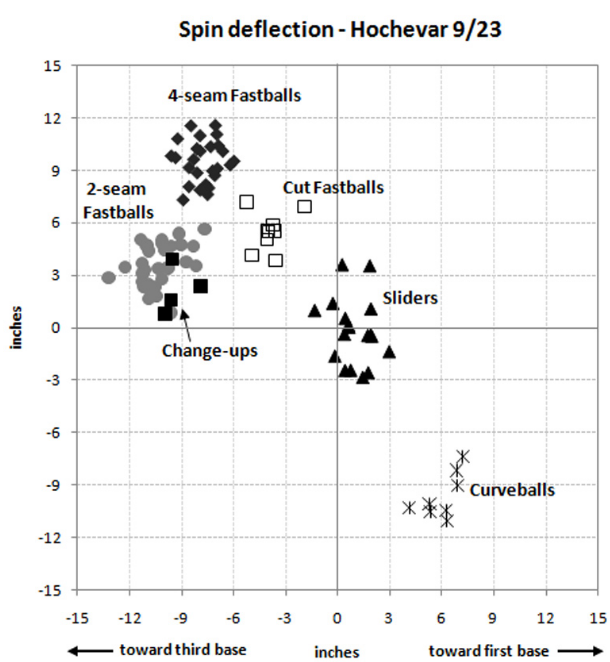
The Graphs of Hochevar

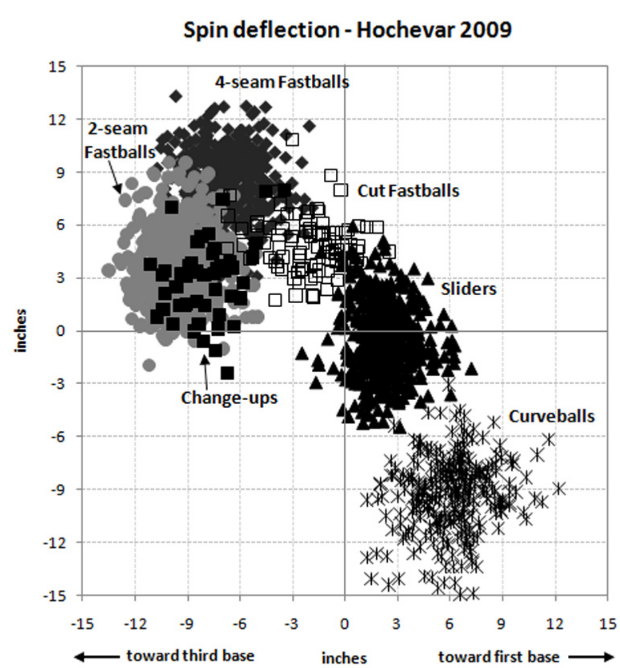
Two ways of looking at the pitch data are particularly useful when classifying pitches into types. The first is to plot pitch speed versus horizontal spin deflection. Speed helps us separate fastballs from off-speed pitches. The horizontal spin deflection helps us separate fastballs and change-ups, which break in toward a same-handed batter, from curveballs and sliders, which break away from a same-handed batter.



It is also often useful to plot spin deflection in both the vertical and horizontal dimensions. For example, this can help us separate four-seam fastballs from two-seam fastballs. Two-seam fastballs have less vertical spin deflection (i.e., they "sink" relative to the four-seamer), and they break in more toward a same-handed batter.

Due to inconsistencies between PITCHf/x systems at different parks (and sometimes at the same park at different points in the season), it can be difficult to distinguish between pitch types on a season-level graph That's why I prefer to identify pitches on a game level.





between pitch types on a season-level graph, whereas the distinctions between pitches may be obvious at the game level.

Some pitchers' pitch types may be particularly difficult to distinguish, but in those cases there are additional tools in the classifier's arsenal. For example, pitchers often throw a few pitch types primarily to one handedness of batter. Change-ups and split-finger fastballs tend to be thrown to opposite-handed hitters. Sliders are thrown more frequently to same-handed hitters. So handedness of the batter can be one clue.

Pitch location is another place to look for help. Some pitchers tend to throw their four-seam fastball to one side of the plate and their two-seam fastball to the other side of the plate. It also may be useful to graph pitches in terms of spin axis and spin rate to tease out subtle differences between pitch types. Sometimes something as simple as looking at the sequence of pitch speeds over the course of a game may help identify the boundary between fastballs and change-ups.

Really, the tools for pitch classification are as endless as your understanding of the game of baseball and the tendencies of the pitcher in question. The discipline of pitch classification is as fine a way as any to increase your understanding of the art of pitching.

HITf/x, FIELDf/x and Trackman

The growing success of PITCHf/x has only increased the thirst for similarly detailed knowledge about additional parts of the game. In May 2008, Sportvision held its First Annual PITCHf/x Summit for analysts, scientists and team personnel. At the summit, Peter Jensen presented a proposal for using existing PITCHf/x camera footage to track batted balls coming off the bat. By spring 2009, Sportvision had brought this HITf/x system to fruition, and participants in the Second Annual PITCHf/x Summit in July 2009 were given access to a month's worth of HITf/x data.

This HITf/x data set contained the initial velocity (speed and direction) for nearly all the batted balls hit in every major league game during April 2009. Ground balls that hit the dirt area near home plate proved to be difficult to track in the PITCHf/x video, although Sportvision provided some trajectory data based on the flight of the ball after the first bounce for these grounders.

Analysis of the HITf/x data is still in its nascent stages. Sportvision is fine-tuning its data production procedures and is considering whether or how to make additional HITf/x data available to the public. Nonetheless, this data holds considerable promise for evaluation not only of hitting and pitching but also a better system of fielding evaluation. Knowing exactly how hard and in what direction each ball was hit allows us to distinguish between a screaming line drive into the hole at short and a Texas Leaguer dropping just over the shortstop's glove. Most of our currently available data sets consider both of those batted balls simply as singles to left field.

This past summer, Sportvision and MLBAM announced their plans to take the digital tracking of game events to another level altogether. Sportvision has installed a camera system in AT&T Park in San Francisco that allows it to track the movement of the players and the baseball across the whole field throughout the game. Plans are to install and test this system in the other parks around the league during the 2010 season, similar to the test and installation phase for PITCHf/x in 2007. This system, popularly labeled FIELDf/x, promises to give us an unprecedented window into baserunning, defensive positioning, full batted ball trajectories, and detailed assessments of fielder capabilities.

Sportvision has not announced plans to share any of this data publicly, although it can be expected that MLBAM will want to incorporate much of the information from FIELDf/x into its Gameday application to make the viewer's experience even more immersive. Other broadcasters likely will want to make use of this fascinating storehouse of information, and many major league clubs will no doubt want to tap this analytical gold mine. It will be interesting to see how FIELDf/x unfolds and how far-reaching its impacts on the game will be.

Another tracking technology that has gotten less press but is no less interesting is Trackman Baseball, a portable phased array Doppler radar system for tracking pitched and batted balls. Trackman is widely used for tracking golf shots, and the parent company, Interactive Sports Games, has adapted the technology for baseball. The Trackman Baseball system is able to reconstruct the full trajectory of batted balls. Its portability makes it promising for scouting and coaching applications, where the fixed nature and expense of the PITCHf/x and FIELDf/x systems prevent their use.

Conclusions

Baseball and baseball analysis is undergoing exciting transitions due to these new tracking technologies. The wide availability and discussion of much of this information among the online sabermetric community is leading to an explosion in our understanding of the game at a much deeper level than previously possible.

Thanks to a number of enterprising individuals, it is now possible for a baseball fan or researcher to interact with this data at any number of levels. Several of them are listed in the References section. I hope this article will help you engage at a level that interests you and increase your enjoyment of the game.

References

Several websites present PITCHf/x data in graphical format. Brooks Baseball (http://www.BrooksBaseball.net), created by Dan Brooks, offers a variety of detailed pitcher and umpire charts for each game, available in real time. Fangraphs (http://www.fangraphs.com—look for the PitchFx tab under each pitcher's page)offers basic speed, spin deflection and release point graphs for every pitcher, usually within a couple days after the game. Texas Leaguers (http://pitchfx.texasleaguers.com), created by Trip Somers, also has basic spin deflection and release points graphs, usually available the day after the game, but offers useful flexibility in graphing between dates of the user's choice.

If you are interested in further reading or resources on PITCHf/x and related topics, you may find the following websites useful. Dan Brooks has created a PITCHf/x tutorial at the Sons of Sam Horn discussion board (http://www.sonsofsamhorn.net/wiki/index.php/Pitchfx). My Fastballs blog contains a glossary of XML fields in the Gameday data (http://fastballs.wordpress.com/2007/08/02/glossary-of-the-gameday-pitch-fields/). Dr. Alan Nathan has a site with all manner of information about the physics of baseball (http://webusers.npl.illinois.edu/~a-nathan/pob/).

If you want to get your hands dirty, you can find the data in raw XML format at the MLB Gameday website (http://gd2.mlb.com/components/game/mlb). If you want to build your own database, you can download the necessary files from Baseball on a Stick, or you can get your hands even a little dirtier and build it using the instructions from my Fastballs blog (http://fastballs.wordpress.com/2007/08/23/how-to-build-a-pitch-database). For those who don't want the headache of Python or Perl code, Jeff and Darrell Zimmerman have provided a database of PITCHf/x data for download, and they appreciate donations for the bandwidth needed (http://www.wantlinux.net/category/baseball-data/).

Acknowledgments

Special thanks go to Alan Nathan and Marv White for their support, encouragement, and willingness to share time and data; to Dave Studenmund and the team at Hardball Times for providing a platform that fosters good baseball research and specifically for encouraging me in my work; and to the other PITCHf/x analysts, too many to list here, for sharing so much of their work and ideas.