Analysis of Softball Pitch Trajectories by the Cluster Method

Caitlin Krebs
Physics Department at Drury University

ABSTRACT

Softball and physics may not seem like they overlap, but physics can be helpful in analyzing pitches. Firstly, we used a least-squared fit on the smooth data, in order to obtain the nine parameters(initial positions, initial velocities, and constant accelerations) necessary to utilize kinematics. Upon figuring the specifications, we calculated the deviation from a straight line trajectory caused by the Magnus force.

Using data from the 2011 Women's College World Series, we successfully classified the pitches into types using the cluster method, in which types of pitches are classified based on velocity, horizontal movement, and vertical movement. Through analysis, it is concluded that the cluster method is effective, and there is a possible connection between initial position and pitch type.

I. Background and Introduction

On October 4, 2006, Sportvision's PITCHf/x system made its debut in Major League Baseball. Two cameras were placed with one above home plate and the other near first base. This system tracks the trajectory by identifying the pitches' speed, break and location in real time. This data is used to inform broadcasters and in MLB's Gameday web application.[2]

Using kinematics, this paper is a real world application of physics analysis to softball pitching knowledge. It studies the Women's College World Series(WCWS), the NCAA Division 1 championship for softball. We analyzed games from the 2011 WCWS whose data set contained over 3500 pitches from thirteen pitchers. In order to protect the pitchers' anonymity, they are numbered 1-4.

The data set's numbers come from the Pitchf/x system that is used in ESPN broadcasts. To get the data, two cameras with roughly orthogonal axes are set up. From the cameras, this system determines the position versus time for each pitch[1].

There are three forces acting on the softball: gravitational, drag, and Magnus forces. The gravitational force points downwards and changes the direction and magnitude of the velocity. The drag force opposes the velocity and decreases its magnitude. The Magnus force is perpendicular to the velocity, on the leading edge of the ball based on the direction of the spin, and changes the direction of the velocity. The focus of this paper is the spin force responsible for the movement of the softball.

The ultimate goal is to calculate the break caused by the Magnus force. In order to do this, we must calculate the components of the Magnus acceleration. By doing a least-squares fit, we obtain the initial velocity, initial position, and acceleration for each axis[1].

In this paper, we will explain how the deviation from a straight line(movement) is calculated, how that movement and the cluster method is then used to categorize different pitches, and finally what can be learned through this classification.

II. Method

When doing a real world application, we need to have a coordinate system. The origin of our system is the back corner of home plate as displayed in Figure 1.

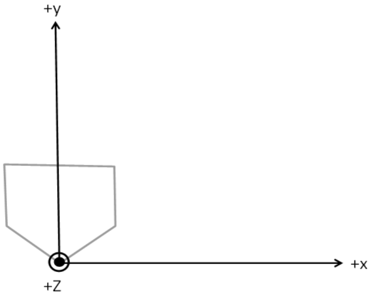

Figure 1 Positive y points towards the pitcher. Positive x points to the catcher's right. Positive z points upward.

Ultimately, all of the following work is done to calculate the deviation from a straight line trajectory caused by the Magnus force.

The nine parameters necessary to calculate the trajectories are initial positions, initial velocities, and accelerations (assumed to be constant) all in x, y and z vector components. We can calculate quantities in order to analyze the trajectories. There is a 2nd degree polynomial that is least-square fitted to the data. It comes in the form:

$$(1)\quad y = y_0 + v_{yo}t + \frac{1}{2}a_y t^2$$

From the initial parameters, the final velocity, where the final position is the front edge of the plate(1.417 ft), is then found:

$$(2)\quad v_{yf} = -\sqrt{v_{yo}^2 - 2a_y\,(y_0 - 1.417)}$$

Next, the time it takes the ball to travel from the release point to crossing the plate is calculated using the final velocity found in equation (2):

$$(3)\quad t = (v_{yf} - v_{y0})/a_y$$

The y velocity and the time it takes the ball to go from release to 39 feet is calculated in a similar way:

$$(4)\quad v_{y39} = -\sqrt{v_{yo}^2 - 2a_y\,(y_0 - 39)}$$

$$(5)\quad t_{39} = (v_{y39} - v_{y0})/a_y$$

The total acceleration includes contributions from the drag, Magnus, and gravitational forces. In order to calculate movement, we only want to focus on the Magnus acceleration. First, we must calculate the drag acceleration. Since the drag is opposite $\vec{v}$ and the Magnus is perpendicular to $\vec{v}$, we take the drag to be the projection of $\vec{a}$ in the -$\vec{v}$ direction.

$$(6)\quad \begin{aligned} \bar{v}_x &= \frac{1}{2}(2v_{x0} + a_x\,t) \\ \bar{v}_z &= \frac{1}{2}(2v_{z0} + a_z\,t) \\ \bar{v}_y &= \frac{1}{2}(2v_{y0} + a_y\,t) \\ \bar{v} &= -\sqrt{\bar{v}_x + \bar{v}_y + \bar{v}_z} \end{aligned}$$

$$(7)\quad a_{drag} = -(a_x\,\bar{v}_x + a_y\,\bar{v}_y + (a_z + 32.179)\,\bar{v}_z)/\,\bar{v}$$

Then, we remove the drag and gravitational acceleration from total acceleration; we are left with the Magnus acceleration.

$$(8)\quad a_{magx} = (a_x + \frac{a_{drag}\,\bar{v}_x}{\bar{v}})$$

$$(9)\quad a_{magz} = (a_z + \frac{a_{drag}\,\bar{v}_z}{\bar{v}} + 32.179)$$

Finally, we are able to calculate the horizontal and vertical break that is caused by the Magnus force acting on the ball from 39 feet to the edge of home plate. We convert it from feet to inches:

(10) $p_{fxx} = \frac{1}{2} a_{magx} (t - t_{39})^2 * 12$

(11) $p_{fxz} = \frac{1}{2} a_{magz} (t - t_{39})^2 * 12$

The total movement is then calculated using the x and z components:

(12) $p_{fx} = \sqrt{p_{fxx} + p_{fxz}}$

From these calculations, we formulate ROOT[3] plots to use cluster analysis to classify the different pitches that each pitcher throws. The three criteria utilized to divide pitches are horizontal movement, vertical movement, and speed. First, pitches are split into groups based upon speed. This results in an off-speed or change up category for pitches with slower initial velocities. Then, the remaining pitches are divided based on the results and analysis of the horizontal and vertical movement ROOT[3] plots.

The movements expected for each type of pitch are displayed in Figure 2. Curveballs move to the catcher's right and screwballs to her left. Drops move downward towards the plate and rises move upwards away from the plate.

Pitches	X-Movement	Z-Movement
Rise	∅	Up
Drop	∅	Down
Curve	Right	∅
Screw	Left	∅

Figure 2 Pitch Types.

A pitch is identified by the cluster, which is a group of pitches that have similar speed, horizontal movement, and/or vertical movement. The red circles in Figure 3 represent the different clusters identified. For all thirteen pitchers in the data set, there were ROOT[3] plots created that display the different pitch clusters.

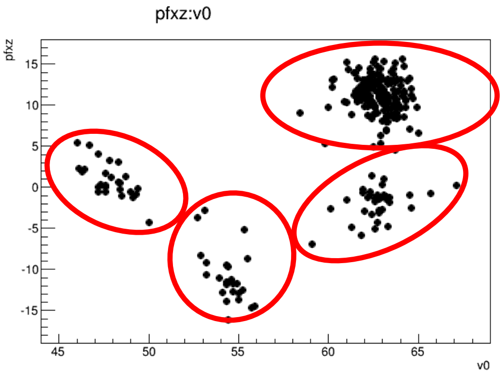

Figure 3 Pitch Clusters for Pitcher 1. A visual on pitch clusters where pitches differ in velocity and up/down movement.

Then, the resulting data was analyzed again using a color scheme to identify each pitch. Thirteen individual plots were made representing each pitcher. For every pitcher, four plots were color coded for pitch type in order to analyze all criteria. These plots were initial velocity, vertical vs. horizontal movement, vertical movement vs. initial velocity, and vertical movement vs. initial velocity. This method allowed a closer look at whether the pitches were correctly classified into the appropriate cluster type.

Lastly, ROOT[3] graphs were made with various other data quantities to determine possible patterns. This includes a graph of the initial position(z0 vs. x0) for each pitcher color coded by pitch type

III. Results

Of the 3505 pitches collected, 3493 of them were classified into perspective pitches using the cluster method. This resulted in a .34% error rate. Two of the resulting classifications are displayed in Figure 4 and Figure 5.

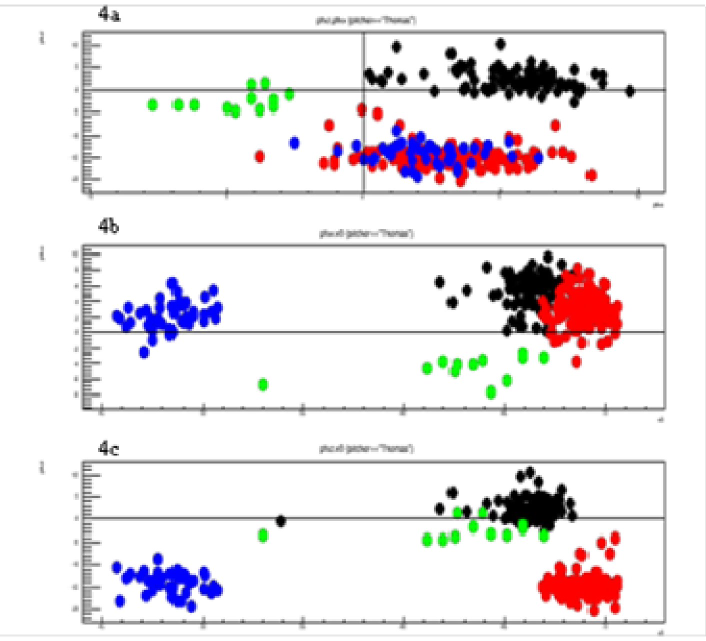

Figure 4 Cluster analysis for Pitcher 2.
4a) Vertical movement vs. horizontal movement
4b) Horizontal movement vs. initial velocity
4c) Vertical movement vs. initial velocity

Pitcher 1 has four different pitch clusters represented in Figure 4. In the top graph, there is a cluster with negative horizontal movement, so it is classified as a screw. The cluster that has positive horizontal movement is identified as a curve. Upon looking at the bottom cluster it appears to be one type of pitch, a drop ball, but when taking speed into account, the third and fourth clusters are drop and an off-speed drop.

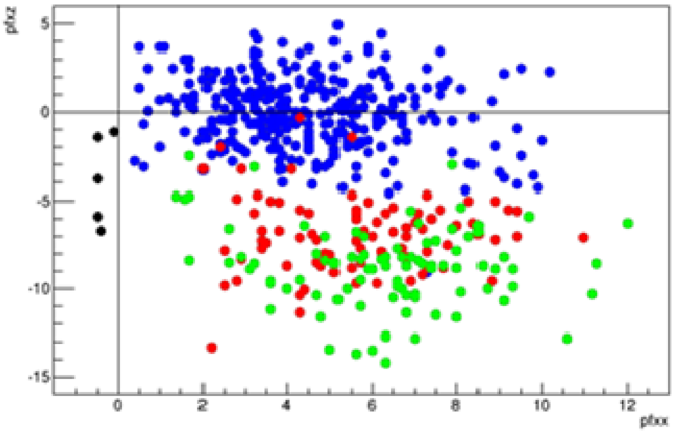

Figure 5 Vertical movement vs horizontal movement color coded by pitch type for pitcher 2.

In Figure 5, Pitcher 2 was found to have three different clusters. She had a pitch with mainly only positive x movement, classified as a curve. Also, she had pitches with positive x movement and negative z movement. These were broken up into two clusters based on speed, a curve drop and off-speed curve drop.

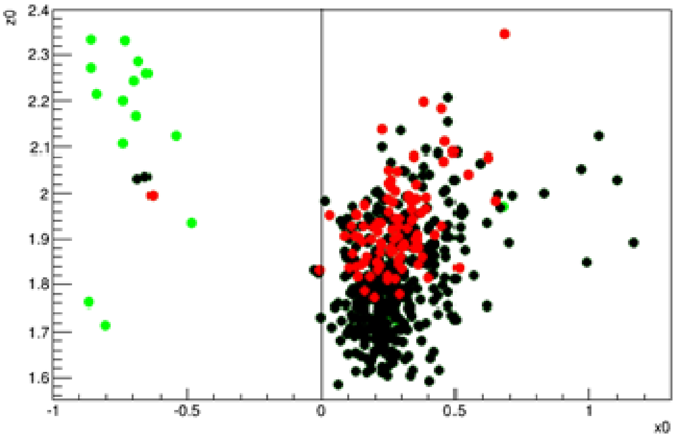

Figure 6 Initial Position color coded by pitch type for Pitcher 3

Another item we researched is variance in release point depending on the type of pitch. Figure 6 is the initial position of Pitcher 3's pitches.

IV. Discussion

When we use the least-squared fit for the trajectory, we assume that the accelerations are constant. This is a reasonable assumption since it does narrowly match the actual trajectory of the ball.

Figure 4a shows Pitcher 1's z-movement vs. x-movement. This best displays the cluster we identify as a screw. It has movement to the left and from Figure 4b and 4c it can be seen that is has a slightly lower initial velocity.

Figure 4b is Pitcher 1's x-movement vs. the initial velocity. From this it can be seen that she throws two different speeds.

The slower pitches move down and are classified to be drop/change.

Finally, Figure 4c displays the z-movement vs. initial velocity. This graph shows the difference between the faster clusters. The pitches that have negative z movement are classified as drops. The yellow pitches have mostly positive z movement and all have positive x movement.

Figure 5 represents Pitcher 2’s vertical movement vs. horizontal movement in her pitches. This plot shows it is not easy to identify the clusters by simply looking at the movement plot. Upon studying the movements vs. velocity, is it possible to see that there are three distinct pitches: curve, curve drop, and off-speed curve drop.

These results demonstrate that the cluster analysis of softball pitches is possible. Of significant note is how remarkable it is that the pitches do come in clusters. A small shift in the spin axis can actually greatly alter the movement. Of all of the different directions the ball could move and the various velocities possible, the fact that the pitches do come in clusters is unique.

Pitchers have different releases, body position, and snaps when attempting to throw the various pitches. This cluster analysis does not take into account what the pitcher is attempting to throw, but rather the actual deviation from the straight line trajectory. For instance, a pitcher may be attempting to throw a curve ball(positive x movement), but the ball may actually have negative x movement and therefore be classified as a screw.

An additional idea for analysis is whether or not there is a relationship between pitch type and release point. In Figure 6, it appears that Pitcher 3 does show some difference in release point. Most of her screws have negative x initial position. It can also be seen that her off-speed curve drop is released higher than her curve. Not every one of her pitches follows this relationship, such as the green screws that are located in the red and yellow cluster on the right in Figure 6. Of importance to note was that not every pitcher in this data set had a difference in initial position according to various pitch type.

Since the movement we are talking about depends on the Magnus(spin) force, we expect there to be a correlation between the rpm and the movement. In Figure 7, we can see the faster the ball spins the more movement occurs. It does not appear to be a linear relationship, but there is some correlation.

V. Conclusions

Using simple kinematics and softball knowledge, we calculated the deviation from a straight line caused by the Magnus force. Then the cluster method was used to successfully categorize the pitches into types. The criteria used to identify the pitches are speed, horizontal movement, and vertical movement. Lastly, ROOT plots were made to make a visual analysis of these pitches.

Also, we determined that at least when considering some pitchers, there could be a connection between pitch release and type. There was no definite conclusion on this correlation yet.

The next plausible addition to my research would be to use the varying pitch classifications to look into pitch sequencing, or the order of pitch types. In addition, it would be helpful to look at more data with known pitch types to find definite evidence on the connection of initial position and pitch type.

This was an exciting project which enabled the connection between and the real world kinematics of softball.

VI. Acknowledgments

Special thanks to Dr. Alan Nathan from the University of Illinois-Champaign Urbana for advising me in this project.

Caitlin Krebs' summer research was funded by National Science Foundation Grant No. 1062690. The opinions, results, and conclusions presented are those of the researchers and not necessarily those of the National Science Foundation.

VII. References

1. The Anatomy of a Pitch: Doing Physics with PITCHf/x Data. David Kagan, Phys. Teach. 47, (2009).
2. What the Heck is PITCHf/x?. Mike Fast. The Hardball Times Baseball Annual, (2010).
3. ROOT - An Object Oriented Data Analysis Framework. http://root.cern.ch/.