Chance of Hit as Function of Launch Angle, Exit Velocity and Spray Angle

By Jim Albert on January 15, 2018

Introduction

In previous posts, I've considered how the probability of a hit varies as a function of the Statcast variables launch angle and exit velocity. I haven't looked carefully at the spray angle, although it is an important aspect of hitting. Here I'll explore how the probability of hit varies as a function of all three variables, focusing on the spray angle effect.

General Thoughts About Spray Angle

How does spray angle affect hitting? Some initial thoughts …

- For a ground ball, the direction is pretty important. We all know that ground balls up the middle tend to be hits, also ground balls hit along the first and third base lines. In contrast, ground balls hit towards the infielders tend to be outs.
- For balls hit in the air, the impact of the direction is different. Since the outfielders are positioned in left, center, and right fields, it is desirable (from the hitter's perspective) to hit away from the fielders which means along the first and base lines or in the gaps between the fielders. Balls hit in the air towards the fielders tend to be caught for outs.
- We are familiar with "bloop hits" — these are the ones hit at a low speed that seem to be away from the fielders. Certainly the spray angle is relevant in this situation.

Obtaining the Spray Angle

The Statcast data does not give a direct measurement of spray angle. But it does contain variables hc_x and hc_y which relate to the location of the batted ball. I've plotted values of these variables for 2000 batted balls.

2000 Values of (hc_x, hc_y)
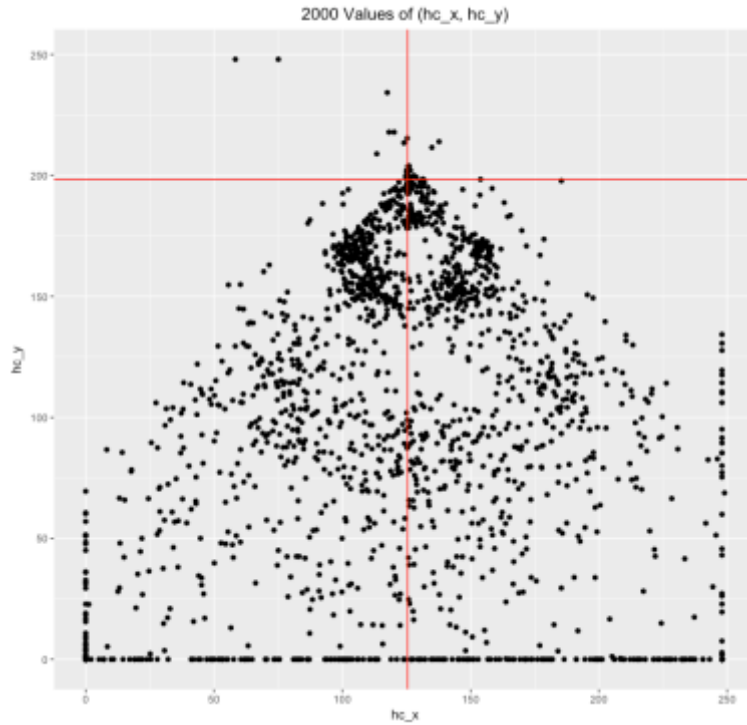
Following Bill Petti's post, I'll do an initial transformation

$$x = hc\_x - 125.42, \quad y = 198.27 - hc\_y$$

which flips these points around and makes the origin home plate. Here are 2000 values of (x, y). (I've added lines corresponding to the 1st and 3rd base lines which indicate that this reexpression is reasonable.)

Locations of 2000 Batted Balls
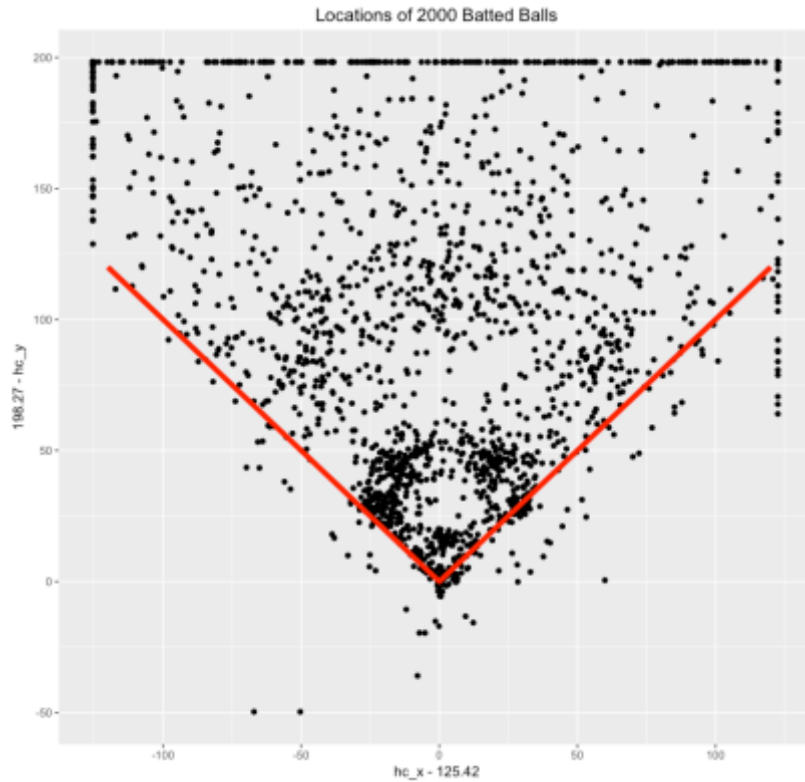
We want to convert the (x, y) field location to a spray angle. I've drawn a right triangle on the field below and labeled the batted ball point (x, y) and the spray angle phi. Using basic trig knowledge, we have that

$$\text{phi} = \text{atan}(x / y) = \text{atan}((hc\_x\text{-}125.42)/(198.27\text{-}hc\_y))$$

where atan is the inverse tan function.
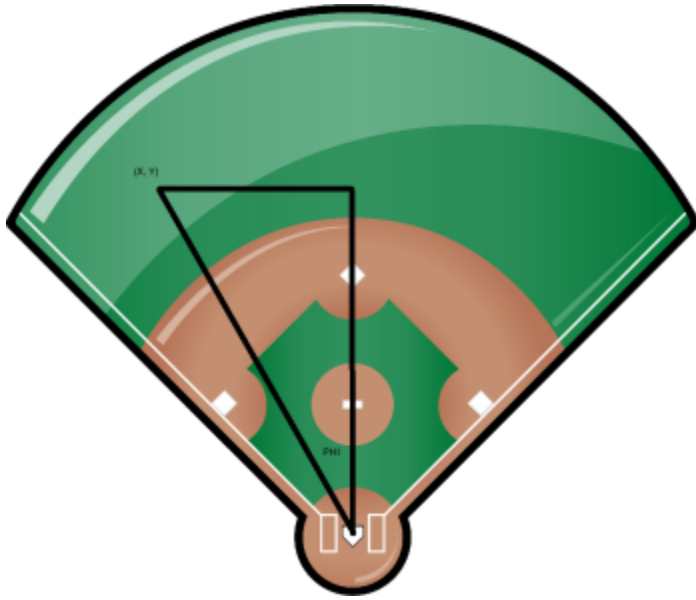
Last, we adjust for the side of the batter — if the batter is left-handed, we let phi1 be the negative value of phi, otherwise we let ph1 = phi (We call this an adjusted spray angle.) So a negative adjusted spray angle corresponds to a batted ball that is pulled, and a positive spray angle is a batted ball hit to the opposite side. An adjusted spray angle of 0 degrees is a ball hit up the middle or towards dead center.

The Model

Now I can use my current favorite toy, a generalized additive model, to fit a model of the form

$$logit(prob(hit)) = s(launch\_speed, launch\_angle, adjusted\_spray\_angle)$$

where s() is a smooth function of the three variables. I fit this model to the 129,365 batted balls for the 2017 season. Using this model, I can predict the probability of a hit for any values of the three variables.

Graph

Graphing hit probabilities when you have three input variables can be a challenge. Here's what I did for the graphs below.

- I chose a wide range of adjusted spray angles from -45 degrees (pulling the ball along the line) to +45 degrees (hitting a ball along the "opposite" line). The horizontal axis for my plot will be the adjusted spray angle.
- I chose three representative launch speeds 80, 90, and 100 mph.
- I fixed the launch angle (a value between -20 and 40 degrees) and graphed the probability of a hit as a smooth curve against the adjusted spray angle for the three launch speeds.

The graphs provide interesting insight on the affect of spray angle on hitting.

Ground Ball (Launch Angle of 0 Degrees)

A launch angle of 0 degrees corresponds to a ball hit along the ground. In this situation, the red line in the graph below (corresponding to the largest launch speed) has the highest hit probabilities, followed by the blue line (90 mph) and the brown line (80 mph). Here spray angles of plus and minus 32 degrees are more likely to be hits, but the most likely location for a hit is a ball up the middle (spray angle of zero). (By the way, a useful reference value on the vertical scale is .33, the overall proportion of batted balls that are hits.)
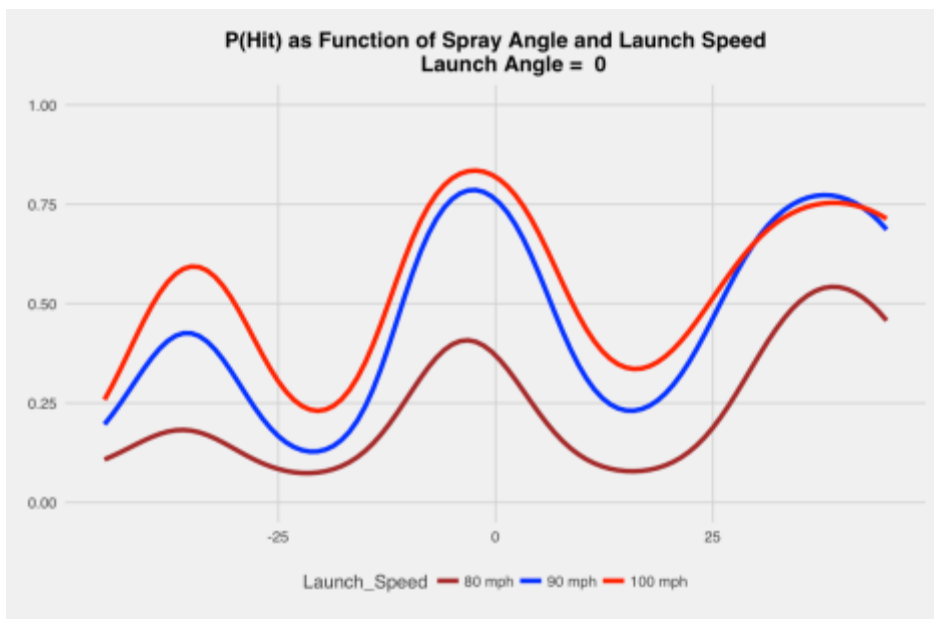
Liner (Launch Angle of 20 Degrees)

A batted ball of 20 degrees is classified by MLB as a liner. Here the impact of spray angle is dramatically different from the first case. For balls that are pulled (negative adjusted spray angles), hitting at a higher launch speed is better. But there is an interval about 0 (balls up the middle) where a 80 mpg batted ball is most likely (among the three launch speeds) to be a hit. Generally, a high fraction of balls hit at this "liner" launch angle are hits.
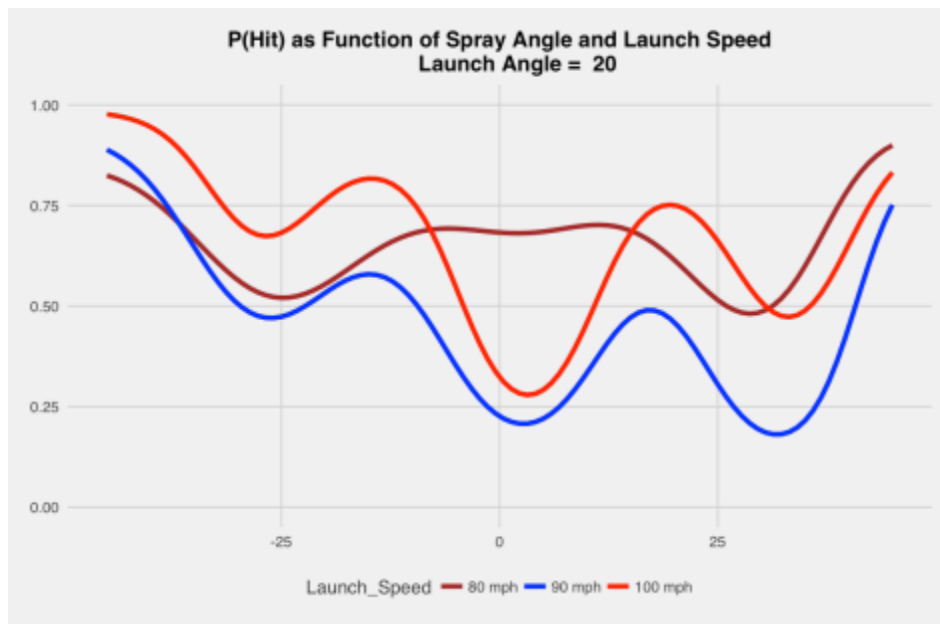
Fly Ball (Launch Angle of 30 Degrees)

A batted ball of 30 degrees is classified by MLB as a fly ball. Generally these types of batted balls tend to be outs for launch speeds of 80 or 90. Harder hit balls (100 mph) with this launch angle are likely be hits for spray angles less than minus 15 degrees and greater than 15 degrees. We are seeing a home run effect here, especially for the balls hit at 100 mph.

P(Hit) as Function of Spray Angle and Launch Speed
Launch Angle = 30
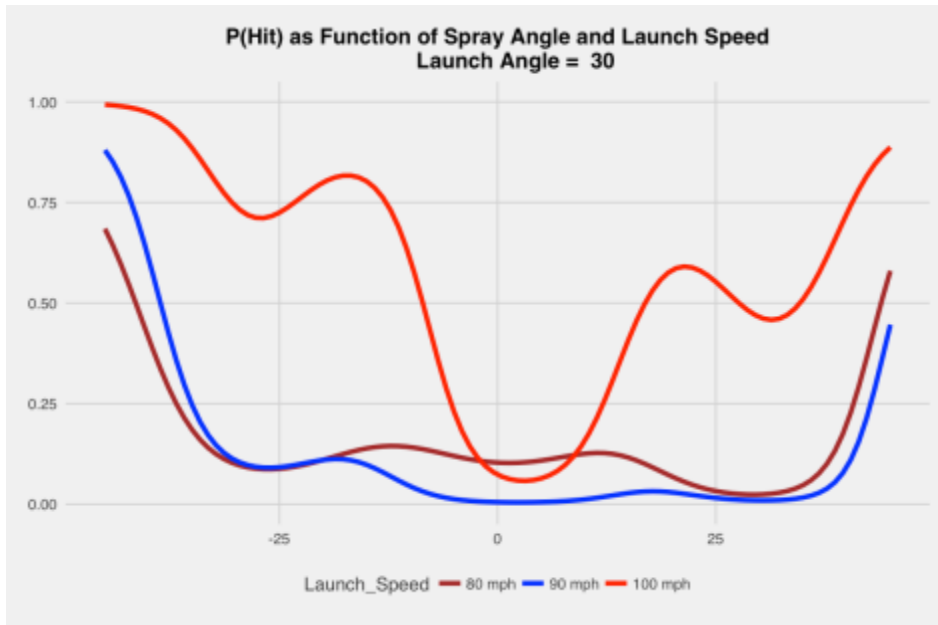
Look at Many Angles

To better understand the pattern of these hit probabilities, I built a Shiny app where one can input the launch angle and three launch speeds of interest and this probability of hit display is produced. Playing with the Shiny app, one gets a better handle on the importance of spray angle in hitting. I recently figured out how to embed a YouTube video in a post. The video below shows using the Shiny app to show the change in the hit probability (again with launch speeds of 80, 90, 100 mph) as the adjusted spray angle moves from minus 20 to 40 degrees.

https://youtu.be/6R0642MvyJQ

Looking Ahead

- I think the use of a generalized additive model is useful in this application since the relationship of the three variables to the probability of a hit is complicated — a simple additive regression predictor of the form launch_angle + launch_speed + spray_angle won't do.

- Certainly there are better outcomes of a batted ball such as linear weights, but MLB will be preoccupied with batting average, so using hit/out as the outcome still is important. This approach could be used with alternative response variables.

- I'd be interested next in exploring the use of this type of model at the player level. Certainly teams want to know the distribution of (launch_angle, launch_speed, spray_angle) for players for defensive purposes. But does this model produce reasonable estimates at the number of hits for all players, and if not, why not?

- To reproduce these plots, you first need to download the Statcast data from the baseballr package, define the new variables and fit the generalized additive model using the gam function from the mgcv package, and then use the Shiny app and the ggplot2 package for the graphing.