

# Using Statcast Data To Study The Flight of a Baseball

Alan M. Nathan

Department of Physics

University of Illinois at Urbana-Champaign

Over the years in which I have been studying the physics of baseball, I have been totally fascinated with baseball aerodynamics. In a simple Physics 101 world, where the effects of the atmosphere are neglected, baseball trajectories are pretty boring. But we don't live in such a simple world and the atmospheric effects of drag and lift play a crucial role in the flight of a baseball.

When PITCHf/x data first became public in 2007, that produced a veritable bonanza of information that has helped considerably with our quantitative understanding of the effects of drag and lift on a pitched baseball. But after a while, those kinds of trajectories get pretty boring too, since they mostly follow a straight line, with just a little bit of deviation due to the combined effects of gravity and spin. Far more varied, and therefore more interesting, are the trajectories of batted baseballs, which run the gamut from line drives (which are sort of like pitches), to fly balls, to popups. If the goal is to understand the atmospheric effects with quantitative precision, it is necessary to investigate all these varied kinds of trajectories. With the advent of Statcast, we now have the opportunity to do just that.

In this article, I will use Statcast data to take a first look at batted ball trajectories, with the goal of developing an aerodynamics model, including the effects of drag and lift, based on the variation of fly ball distances with exit speed, launch angle, and air density. The data used in this study consisted of exit speeds, launch angles, and distances of approximately 80,000 batted balls from the 2015 season. Although the spray angle was also part of the data set, it was not used in the analysis I present here. Since the home field and game time temperature for each batted ball were also known, it was straightforward to calculate the air density, assuming standard atmospheric pressure and 50% relative humidity, neither of which I had access to.

Let me start with the overview shown in Figure 1. For this analysis, I considered only fly balls hit for which the air density was close to the MLB average. In particular, extreme elevations (e.g. Denver) and temperatures were excluded. In effect, I am trying to get an overview without the added complication of extreme atmospheric conditions. The plot shows average fly ball distances and their standard error as a function of launch angle for various values of the exit velocity. These data show quantitatively what we probably already knew at least qualitatively. Namely, fly ball distance reaches a maximum at launch angles in the vicinity of  $25^{\circ}$ - $30^{\circ}$ , with the angle decreasing slightly as the exit speed increases. Moreover, distances increase with exit speed at the rate of about 5 ft for each mph increase in exit speed.

I have done this kind of [analysis previously](#) using HITf/x data, combined with independently measured [home run distances](#), and found that an exit speed/launch angle of 100 mph/ $26^{\circ}$  leads to a mean distance of 405 ft, over 20 ft greater than found in the present analysis. The problem is a well-known issue with HITf/x exit speeds, which are measured at a distance somewhat removed

from the ball-bat impact point, resulting their being systematically underestimated. A 20 ft discrepancy corresponds to an underestimation of exit speed by about 4 mph.

These data are extremely valuable in developing and fine-tuning an aerodynamics model for the flight of the baseball. The important components of such a model are drag (i.e., air resistance) and lift (which results from the backspin). I utilize a model with five parameters that can be adjusted to best fit the data shown in the plot. Three of these parameters relate to drag and how it depends on the speed and spin of the baseball. The other two are used to specify the rate of backspin as a function of exit speed and launch angle. The resulting model is shown by the dashed curves, which faithfully reproduce many of the features of the data. In particular, the model accounts for the slight shift in the peak of the distributions to smaller launch angle as the exit speed increases, a consequence of the increase of drag with speed. A notable exception to the good agreement is at the highest exit speed and angles below about  $22^\circ$ , where the data fall distinctly below the curve and even appear to be discontinuous. Given that most things in nature behave smoothly, the data look suspect to me, but any stronger conclusion will have to await more data.

Figures 2 and 3 show mean distances for fly balls hit with exit speed in the range 101-105 mph and with launch angle in the range  $25^\circ$ - $30^\circ$ . Figure 2 plots the mean distance versus air density along with a dashed line showing the model calculation. Figure 3 plots the mean distance for each MLB stadium, with Denver the clear winner at 430 ft, compared with 401 ft from the average of the other stadiums, indicated by the red dashed line. The Denver effect is huge!

Since the model is an excellent representation of the data, we can use it to draw some interesting conclusions about how fly ball distance depends on the various atmospheric effects. Some of these effects are shown in the table below, all calculated relative to 401 ft, which is the MLB average distance (Denver excluded) for exit speed 101-105 mph and launch angle  $25^\circ$ - $30^\circ$ .

<b>Atmospheric Effect</b>	<b>Change in Distance</b>
$10^\circ$ increase in temperature	3.3 ft
1000 ft increase in elevation	5.9 ft
50% increase in relative humidity at $75^\circ$	0.9 ft
5 mph out-blowing wind	18.8 ft

I next want to examine the Denver effect in more detail. To that end, Figures 4 and 5 compare distances in Denver with those at sea level, where the latter actually refer to air densities in the range  $1.15$ - $1.20 \text{ kg/m}^3$ . Figure 4 shows distance versus exit speed for launch angles in the  $25^\circ$ - $30^\circ$  range, while Figure 5 shows distance versus launch angle for exit speeds in the range 101-105 mph. As before, the lines are the model calculation. From Figure 4 we learn that the slope of distance versus exit speed is larger for Denver than at sea level, so that the Denver effect increases from about 19 ft at 91 mph to about 32 ft at 110 mph. From Figure 5 we learn that the distance peaks at a bit larger launch angle in Denver than it does at sea level. These results make sense physically, as reducing the air density at higher elevations pushes the trajectories closer to those expected in a vacuum, where distances increases much more rapidly with exit speed and peak at  $45^\circ$ . The aerodynamics calculation nicely accounts for both of these features.

Another interesting comparison is Arizona and San Francisco, shown in Figure 6. Arizona is about 1000 ft higher in elevation than San Francisco and has an average temperature about 17<sup>0</sup> warmer, both of which contribute to a lower air density and therefore a longer distance, just as shown in the plot. Once again, the calculation agrees with the general trend of the data.

But not everything is as well understood. As an example, consider Figure 7, which compares Tropicana with Wrigley. These two venues have mean air densities that are nearly identical, yet the data show the ball carrying measurably better at the Trop, by an average of over 10 ft. Perhaps we are seeing the net effect of an in-blowing wind at Wrigley, noting that no wind is expected at the covered Trop.

Finally, I want to take advantage of the fact that we have an aerodynamic model that accounts for most of the features of the data to investigate how fly ball distance depends on the amount of backspin, for a fixed exit speed of 103 mph and launch angle of 27<sup>0</sup>. The results are given in the table. They show that distance increases rapidly as the backspin increase from zero but eventually saturates, with very little gain in distance for spin rates exceeding about 1500 rpm. The reason for the saturation is partly because air drag increases with increasing spin, essentially cancelling the increase in lift.

<b>Backspin rate (rpm)</b>	<b>Distance (ft)</b>
0	336
500	368
1000	386
1500	395
2000	400
2500	403
3000	403

Before concluding, it is useful to remind the reader that the analysis only considers average distances for given values of exit speed and launch angle and that actual distance may vary. One reason for variation might be wind. Another might be variation in the drag properties of individual baseballs, which is a topic I addressed in a [previous article](#) and which can lead to a significant variation in distance.

I very much look forward to continuing my analysis to fine-tune the aerodynamics model. The analysis presented here was “two-dimensional”, in that the spray angle was ignored. Including the spray angle, both at impact and at the landing point, allows for the determination of the rate of sidespin on the batted ball. Moreover, using the spin measured directly from the Trackman device, an integral part of Statcast, as well as the hang time should allow better determination of the lift properties of the trajectory. There is still lots to do and, hopefully, lots of data to help do it.

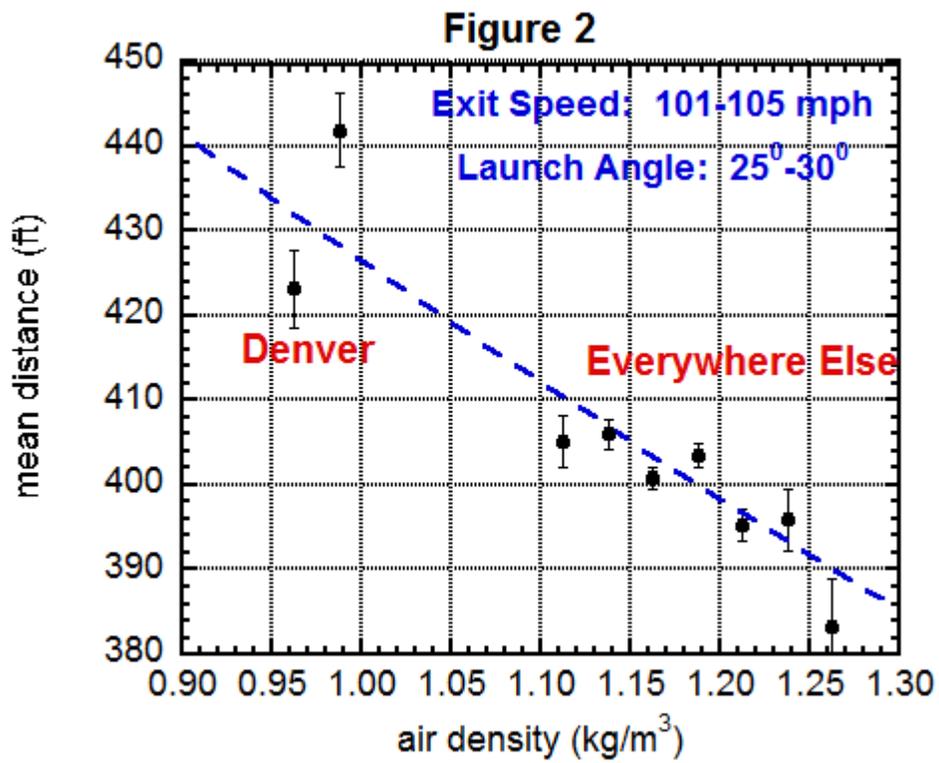
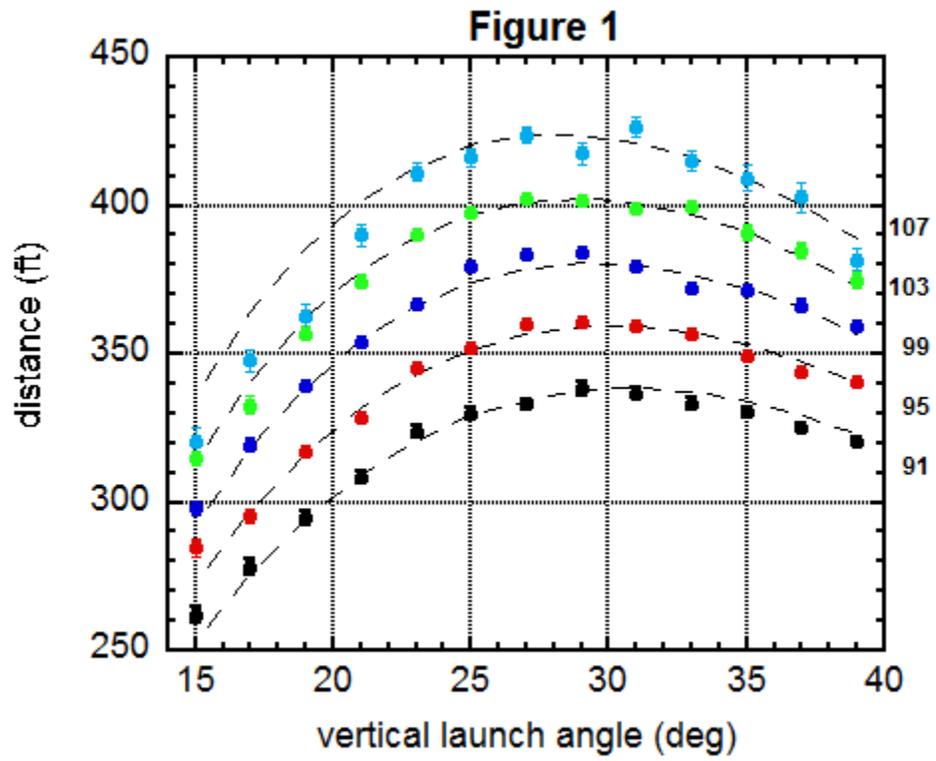


Figure 3

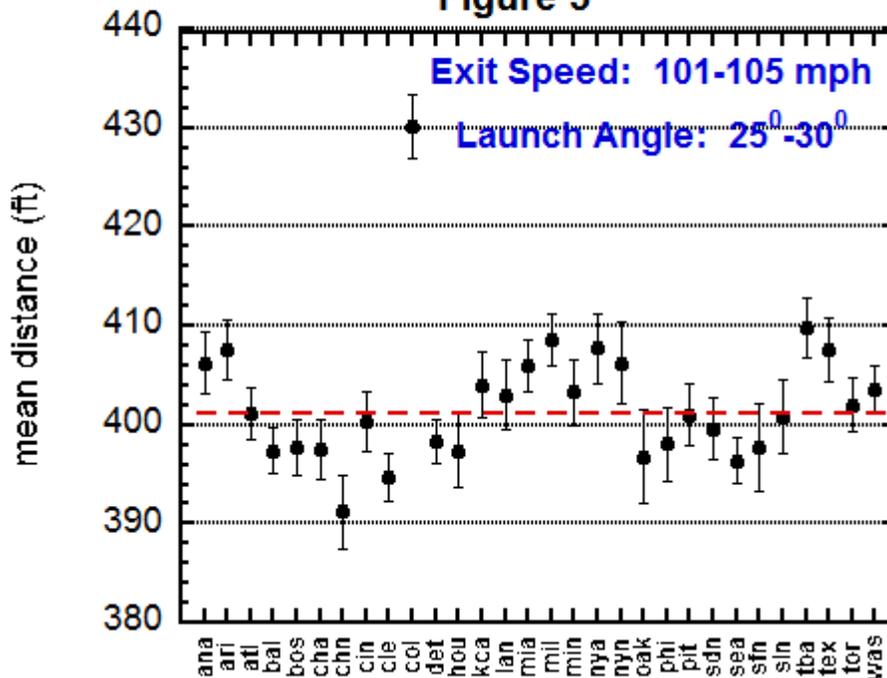


Figure 4

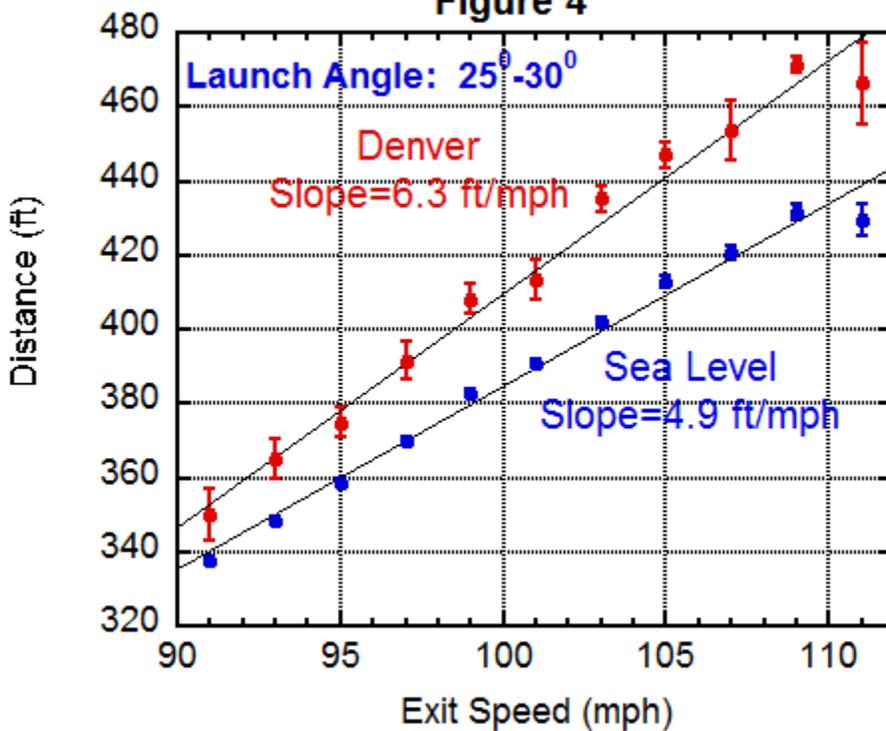


Figure 5

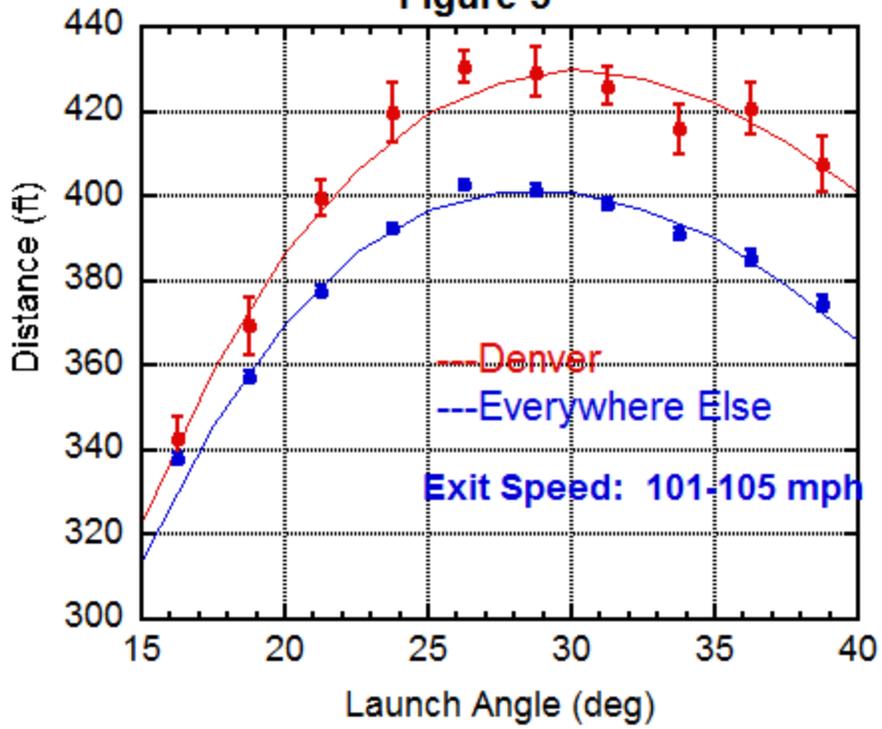


Figure 6

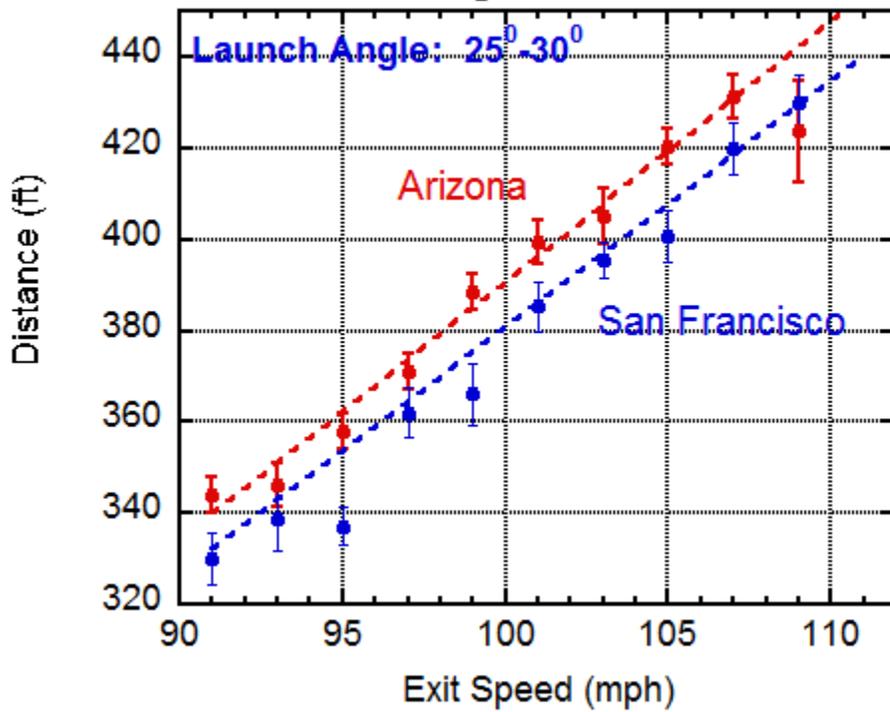


Figure 7

