Deriving Spin Signatures and the Components of Movement from Trackman Data for Pitcher Evaluation

Glenn Healey
Electrical Engineering and Computer Science
University of California, Irvine, CA 92617
Email: ghealey@uci.edu
Research Paper for 2019 SABR Analytics Conference

1 Introduction

The 3-D pitch trajectories and total spin measurements acquired by the Trackman radar enable the recovery of the spin axis for individual pitches. We leverage a computational process proposed by Alan Nathan [6] in combination with fine-grained weather data to compute this recovery for every MLB pitch over a full season. This analysis allows the creation of a new spin signature representation for characterizing and comparing pitchers.

As shown in Figure 1, three forces act on a baseball traveling through the air with a velocity vector $\overline{v}$. Gravity pulls the ball down, drag acts opposite the velocity direction, and the Magnus force causes the ball to change direction due to spin.

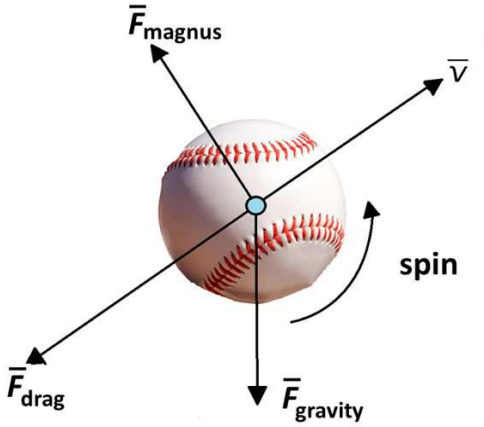

Figure 1: Forces on a spinning baseball in flight

The Magnus force depends on the spin vector $\overline{\omega}$ and causes movement [5] [7] on a pitch. The spin vector $\overline{\omega}$ has a magnitude defined by how fast the ball is spinning, e.g. 2400

revolutions per minute (rpm), and a direction defined by the spin axis and the right-hand rule as shown in Figure 2. A pitcher controls the spin axis with the orientation of his hand and fingers when he releases a pitch. The direction of $\overline{\omega}$ can be represented by the unit vector $\widehat{\omega} = \overline{\omega}/|\overline{\omega}|$.

Figure 2: Spin axis and spin vector direction

The magnitude of the Magnus force [8] is given by

$$|\overline{F}_M| = \frac{1}{2}\rho A C_L |\overline{v}|^2 \tag{1}$$

where ρ is the air density, A is the ball cross-sectional area, and C_L is the dimensionless lift coefficient. The direction of the Magnus force is the direction of $\widehat{\omega} \times \widehat{v}$ where $\widehat{v} = \overline{v}/|\overline{v}|$. Thus, while the drag force is parallel and opposite to the velocity direction, the Magnus force is perpendicular to the velocity direction as shown in Figure 1.

The lift coefficient C_L is zero when $|\widehat{\omega} \times \widehat{v}| = 0$ and increases with $|\overline{\omega}||\widehat{\omega} \times \widehat{v}|/|\overline{v}|$. Therefore, the fraction of the spin magnitude $|\overline{\omega}|$ that gets transferred into Magnus force $\overline{F}_M$ and movement on a pitch depends on $|\widehat{\omega} \times \widehat{v}| = \sin\theta$ where θ is the angle between the spin axis direction $\widehat{\omega}$ and the velocity direction $\widehat{v}$. Figure 3, for example, shows a fastball with only backspin where the spin axis direction $\widehat{\omega}$ is perpendicular to $\overline{v}$ so that $|\widehat{\omega} \times \widehat{v}| = 1$ and all of the spin is transferred to Magnus force. On the other hand, Figure 4 shows a gyroball (sometimes called a backup slider) where the spin axis direction $\widehat{\omega}$ is parallel to $\overline{v}$ so that $|\widehat{\omega} \times \widehat{v}| = 0$ and none of the spin is transferred to Magnus force.

The spin vector $\overline{\omega}$ can be written as the sum of two components

Figure 3: Fastball with only backspin

Figure 4: Gyroball

$$\overline{\omega} = \overline{\omega}_{\parallel} + \overline{\omega}_{\perp} \tag{2}$$

where the gyrospin $\overline{\omega}_{\parallel}$ is parallel to the velocity direction $\widehat{v}$ with $\overline{\omega}_{\parallel} = (\overline{\omega} \cdot \widehat{v})\widehat{v}$ and the transverse spin $\overline{\omega}_{\perp}$ is perpendicular to $\widehat{v}$. The magnitudes of the components are then given by

$$|\overline{\omega}_{\parallel}| = |\overline{\omega}| \cos\theta \tag{3}$$

$$|\overline{\omega}_{\perp}| = |\overline{\omega}| \sin\theta \tag{4}$$

where θ, as defined previously, is the angle between $\widehat{v}$ and $\widehat{\omega}$.

Since the lift coefficient C_L which determines the movement on a pitch is an increasing function of $|\overline{\omega}| \sin\theta/|\overline{v}|$, the transverse spin $\overline{\omega}_{\perp}$ is referred to as the useful spin. The spin efficiency E is defined as the ratio of the magnitude of the useful spin to the magnitude of

the total spin

$$E = \frac{|\overline{\omega}_{\perp}|}{|\overline{\omega}|} \tag{5}$$

Figures 5 and 6 show the effect of the spin axis direction $\widehat{\omega}$ for pitches thrown by right-handers Craig Kimbrel and Dellin Betances in a game on 13 August 2017. Figure 5 is a scatterplot of release speed $|\overline{v}|$ and spin rate $|\overline{\omega}|$ for the four-seam fastballs thrown by the pitchers in this game. Even though the pitchers had similar speed and spin rate for this pitch, Kimbrel's four-seam fastball had a higher swing-and-miss rate and yielded weaker contact than Betances' four-seam fastball during 2017. This was due to Kimbrel obtaining more movement on this pitch as shown in Figure 6 where b_x and b_z are the horizontal and vertical movement components from the release point. This additional movement is due to a spin axis direction $\widehat{\omega}$ that is more nearly perpendicular to the velocity direction $\widehat{v}$ which enables Kimbrel's fastball to achieve a larger Magnus force for a given speed and spin rate.

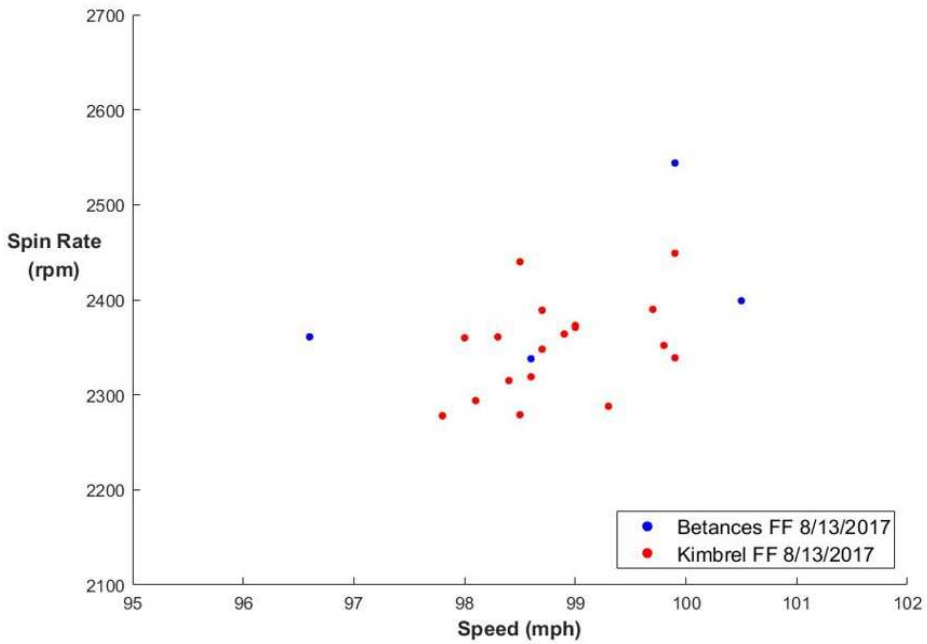

Figure 5: Release speed $|\overline{v}|$ and spin rate $|\overline{\omega}|$ for four-seam fastballs, 13 August 2017

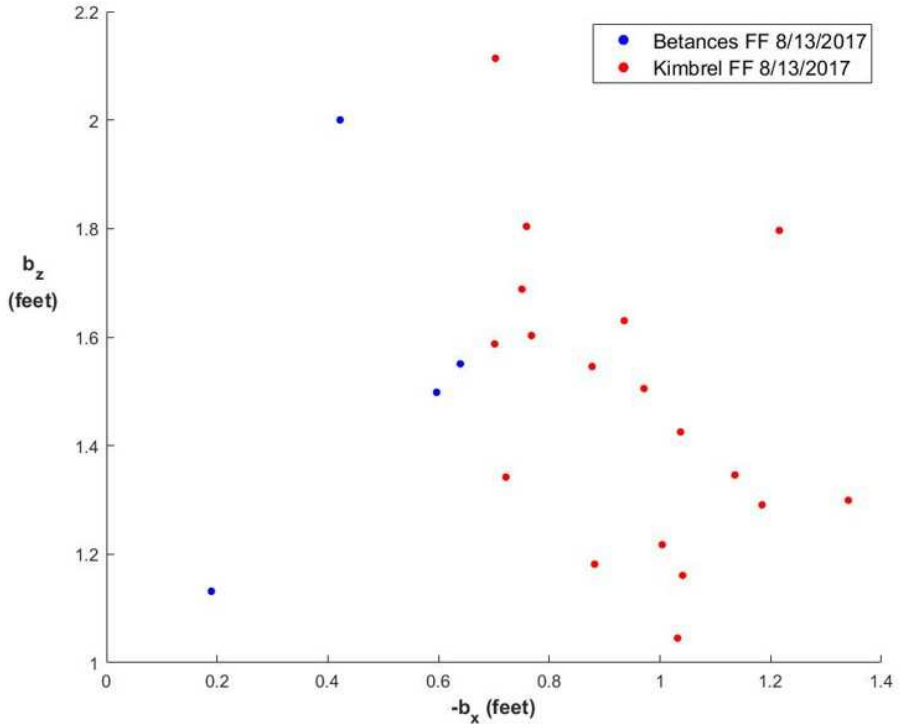

Figure 6: Horizontal b_x and vertical b_z movement, four-seam fastballs, 13 August 2017

2 Data Processing Methodology

2.1 Computing the Lift Coefficient

Equation (1) can be used to recover the lift coefficient C_L from the nine-parameter constant-acceleration pitch trajectory model generated using the Trackman radar. The 3-D acceleration vector $\overline{a}$ for the model can be represented by

$$\overline{a} = \overline{a}_D + \overline{a}_M + \overline{a}_G \tag{6}$$

where $\overline{a}_D, \overline{a}_M$, and $\overline{a}_G$ are the accelerations corresponding to the drag, Magnus, and gravitational forces depicted in Figure 1. Since the drag force is parallel and opposite to the velocity direction and the Magnus force is perpendicular to the velocity direction, we can compute the magnitude of $\overline{a}_D$ as the projection of $\overline{a} - \overline{a}_G$ onto the velocity direction so that

$$\overline{a}_D = -\left[(\overline{a} - \overline{a}_G) \cdot \widehat{v}_\mu\right] \widehat{v}_\mu \tag{7}$$

where $\widehat{v}_\mu$ is a unit vector in the direction of the average velocity vector $\overline{v}_\mu$ over the pitch trajectory. Therefore the Magnus acceleration is computed using

$$\overline{a}_M = \overline{a} - \overline{a}_D - \overline{a}_G = \overline{a} + \left[(\overline{a} - \overline{a}_G) \cdot \widehat{v}_\mu\right] \widehat{v}_\mu - \overline{a}_G \tag{8}$$

From (1) the average lift coefficient over the pitch trajectory is then given by

$$C_L = \frac{2m|\overline{a}_M|}{\rho A |\overline{v}_\mu|^2} \tag{9}$$

where m is the mass of the baseball and Newton's second law is used to relate the Magnus force $\overline{F}_M$ and the Magnus acceleration $\overline{a}_M$.

2.2 Computing Air Density

While $|\overline{a}_M|$ and $|\overline{v}_\mu|$ in (9) can be computed from the nine-parameter model for a pitch, the air density ρ is computed from the altitude, temperature, relative humidity, and the barometric pressure. We obtained weather data for the time and location of each MLB pitch thrown in 2017 from wunderground.com. This data was used to compute the air density ρ in kg/m^3 for each pitch using the model from [1] given by

$$\rho = \frac{(1.2929 * 273.0)(P - 0.01VH)}{760.0(T + 273.0)} \tag{10}$$

where H is relative humidity in percent and T is temperature in degrees Celsius. P is the absolute atmospheric air pressure given by

$$P = b * \exp\left[(-gME)/(RT + 273.15)\right] \tag{11}$$

where b is the barometric pressure in millimeters of mercury, g is the earth's gravitational acceleration in m/sec^2, M is the molecular mass of air in kg/mole, E is the elevation in meters, and R is the universal gas constant in joules/(° K mole). V is the saturation vapor pressure in millimeters of mercury which is computed using the model in [2] given by

$$V = 4.5841 * \exp\left[\frac{(18.687 - T/234.5)T}{257.14T}\right] \tag{12}$$

From equation (1) we see that the magnitude of the Magnus force and hence the movement on a pitch increases with increasing ρ. As an example of this effect, Figure 7 plots the measured vertical movement b_z on four-seam fastballs thrown by German Marquez as a function of ρ in 2017. We see that movement tends to increase with air density.

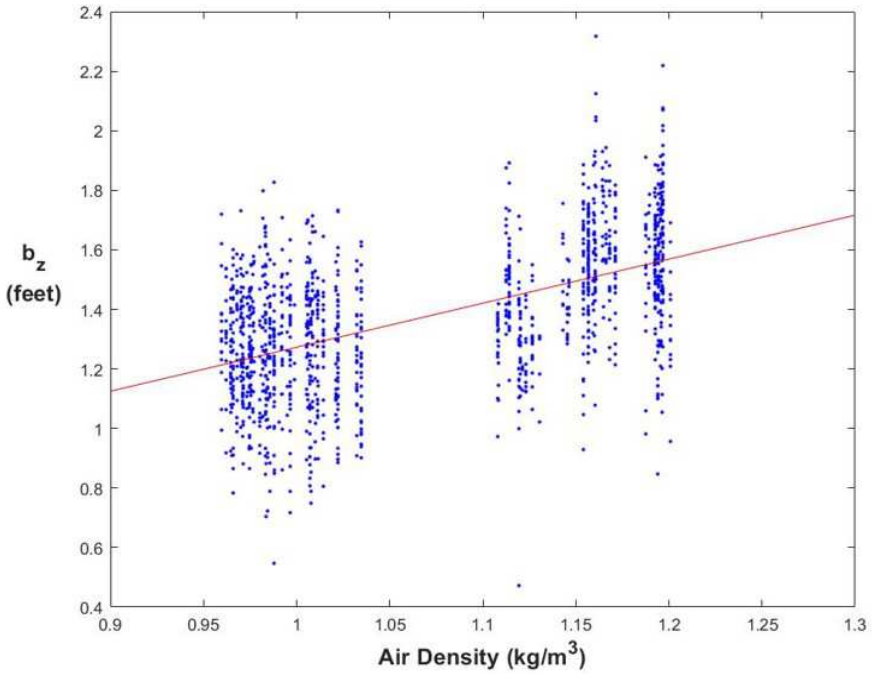

Figure 7: Vertical movement and air density, German Marquez, four-seam fastballs, 2017

2.3 Computing the Spin Parameter

The dimensionless spin parameter S [4] for a baseball traveling through the air is defined by

$$S = \frac{R\Omega}{|\overline{v}|} \tag{13}$$

where R is the ball radius in feet, Ω is the angular velocity in radians per second, and $|\overline{v}|$ is the speed in feet per second. Trackman reports the spin rate $|\overline{\omega}|$ in revolutions per minute

(rpm) which allows the average S over the pitch trajectory to be computed using

$$S = \frac{2\pi R|\overline{\omega}|}{60|\overline{v}_\mu|} \tag{14}$$

where $|\overline{v}_\mu|$ is in feet per second. We note that S is a constant times Bauer Units ($|\overline{\omega}|/|\overline{v}|$) [9].

2.4 The Relationship Between C_L and $S|\widehat{\omega} \times \widehat{v}|$

As described in Section 1, C_L is an increasing function of $S|\widehat{\omega} \times \widehat{v}|$. This function is important because it describes the dependence of the magnitude of the Magnus force which determines pitch movement on pitch parameters through equation (1). In this work we consider estimating this function $f(\cdot)$ from Trackman data for a full MLB season. This relationship has been previously modeled using smaller data sets [8] obtained using laboratory measurements from which the pitch trajectory and spin vector can be measured directly.

Since the direction of the spin axis $\widehat{\omega}$ is not directly measured by the Trackman system, we can consider using the heuristic that $|\widehat{\omega} \times \widehat{v}|$ is typically close to one for fastballs. If this were exactly true, then we would expect a scatterplot of C_L versus S for fastballs to generate a curve that gives the function $f(\cdot)$. Figure 8 is a scatterplot of C_L versus S for four-seam fastballs in 2017 along with the best-fit line. We see that the trend of the relationship is increasing as expected, but that there is significant scatter in the values of C_L for a given value of S.

There are several sources of scatter in Figure 8. Several ballparks in 2017 had biases in the recovered trajectories which led to larger values of the movement parameters (b_x, b_z) for pitches thrown in those ballparks [11]. This tends to give larger estimated values of C_L for a given value of S. Figures 9 and 10 give examples of these biases. Figure 9 is a scatterplot of the horizontal and vertical movement parameters for four-seam fastballs thrown by Jacob deGrom in 2017. We see that the pitches thrown at Citi Field have a significantly larger mean in absolute horizontal movement. Figure 10 is a scatterplot of the horizontal and vertical movement parameters for four-seam fastballs thrown by Ervin Santana in 2017. We see that the pitches thrown at Target Field have a larger mean vertical movement.

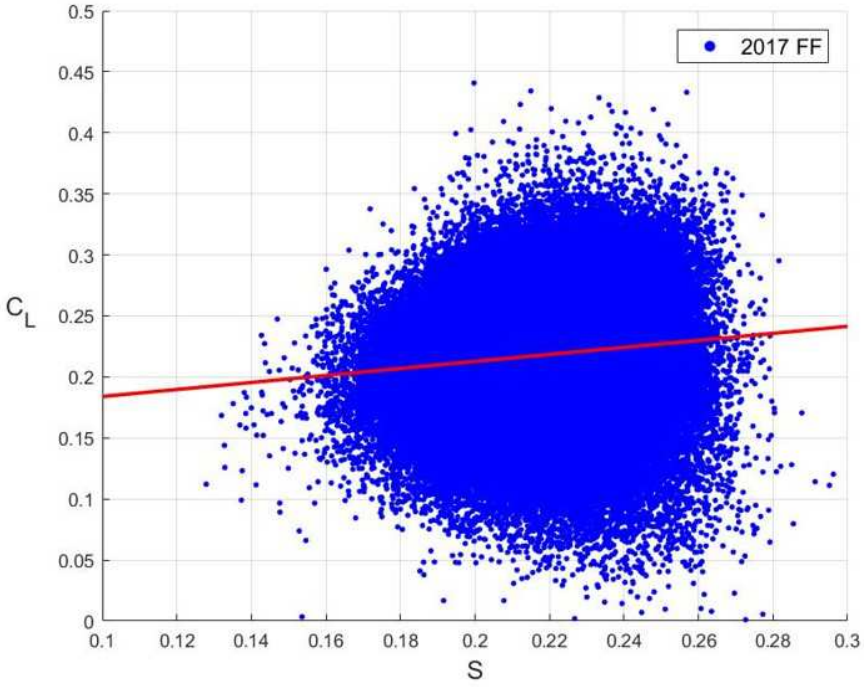

Figure 8: C_L and S, four-seam fastballs, 2017

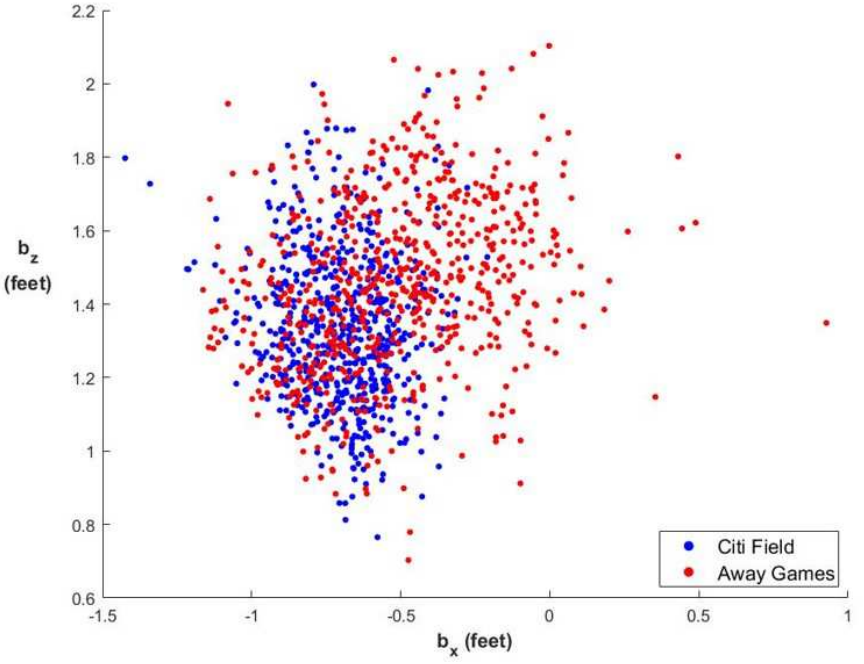

Figure 9: Horizontal and Vertical movement, Jacob deGrom, Four-Seam FB, 2017

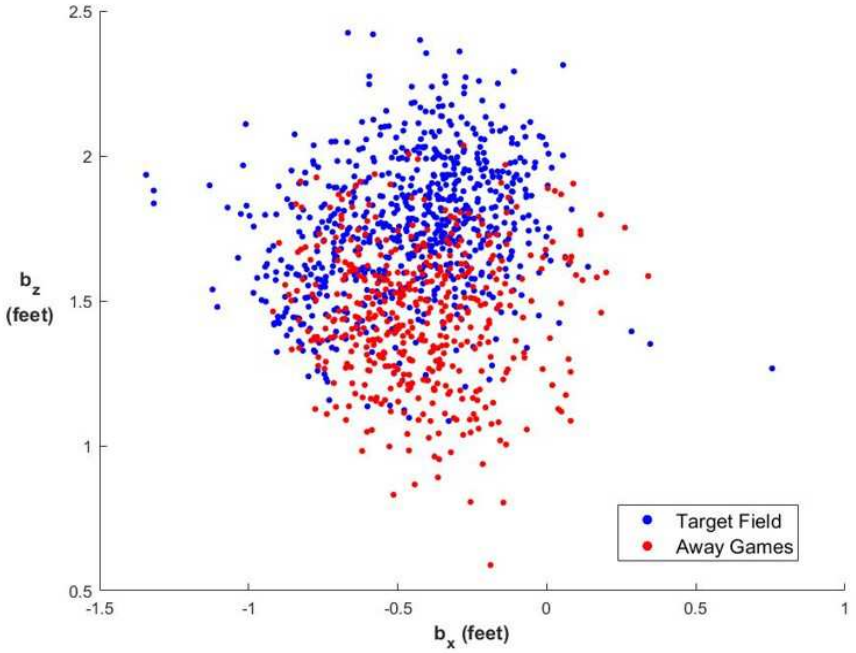

Figure 10: Horizontal and Vertical movement, Ervin Santana, Four-Seam FB, 2017

Another source of scatter is random variation from pitch-to-pitch for a given pitcher and pitch type. Figure 11, for example, is a scatterplot of C_L versus S for four-seam fastballs thrown by Justin Verlander in 2017. We we see that there is substantial pitch-to-pitch variation. For a single pitcher and pitch type, therefore, we can replace the distribution shown in Figure 11 with a robust mean for C_L and S. The robust mean pair is determined by using the Orthogonalized Kettenring-Gnanadesikan (OKG) algorithm [3] to remove outliers as shown in Figure 11 and then computing the sample mean $(\overline{S}, \overline{C}_L)$ of the core points.

We computed the $(\overline{S}, \overline{C}_L)$ robust means separately for every MLB pitcher and pitch type for 2017 after removing data from ballparks with significant biases. Figure 12 is a scatterplot of the $(\overline{S}, \overline{C}_L)$ points for four-seam fastballs (FF) for pitchers with at least 900 four-seamers in 2017 and also for all points for two-seam fastballs (FT) for pitchers with at least 900 two-seamers in 2017. This figure shows that there is still significant variation in C_L for a given S. This variation is largely due to differences in $|\widehat{\omega} \times \widehat{v}|$ across pitchers for fastballs. If we have multiple points with the same S in Figure 12 then the point with the largest C_L has the largest $|\widehat{\omega} \times \widehat{v}|$. Thus, we can estimate $f(\cdot)$ using an upper bound to the

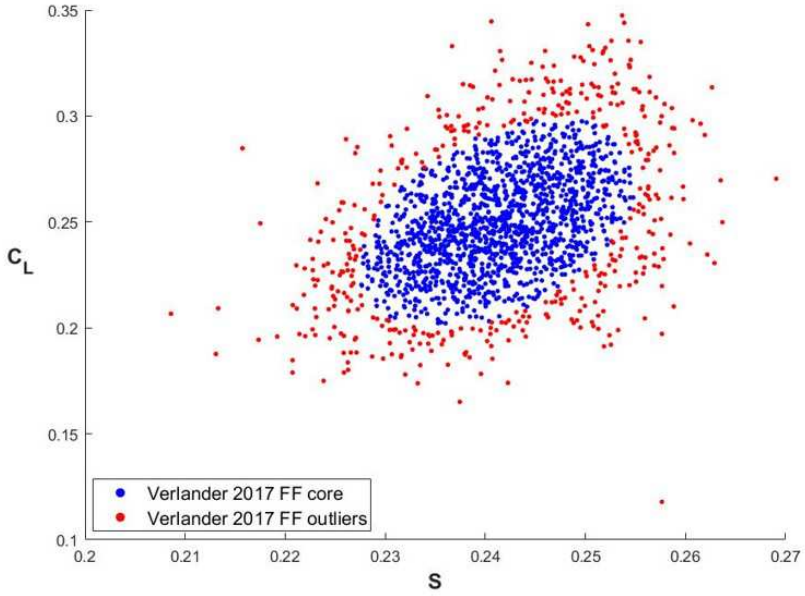

Figure 11: OKG outlier removal, Verlander four-seam fastballs, 2017

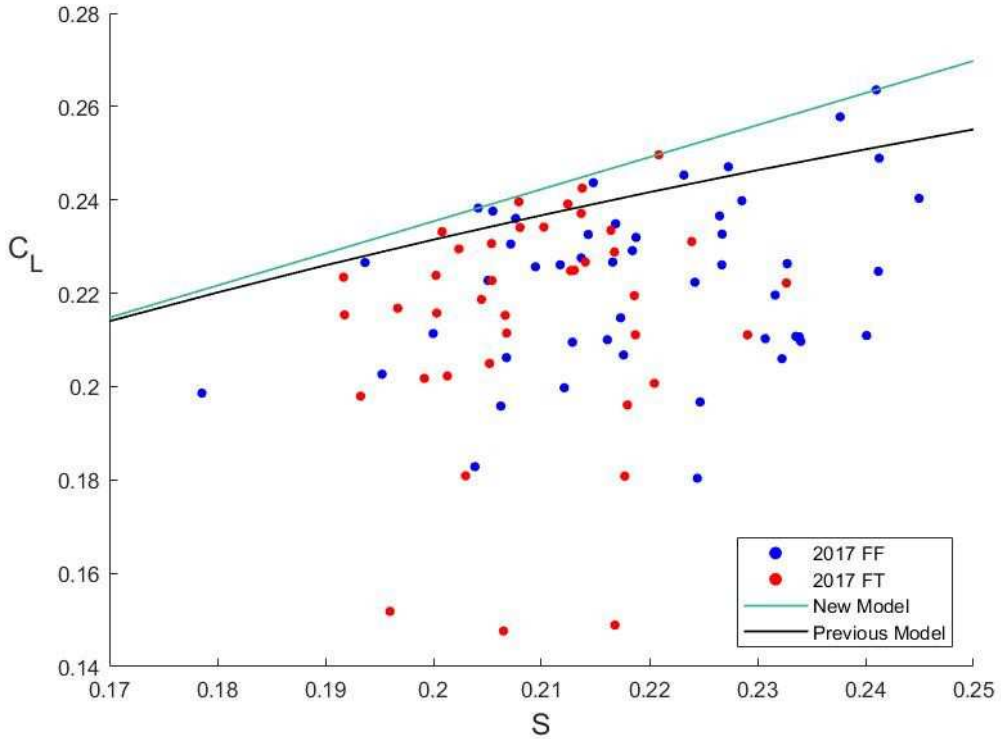

Figure 12: C_L and S, two-seam and four-seam fastballs for individual pitchers, 2017

set of points in the figure. We approximate the upper bound by the green line in the figure which intersects the four-seam fastball for Gerrit Cole, the two-seam fastball for Miguel Gonzalez, and the four-seam fastball for Marco Estrada. The previous nonlinear model [8] for this range is plotted in black in Figure 12.

2.5 Recovering the Spin Vector $\overline{\omega}$

In Sections 2.1 and 2.2 we showed that the lift coefficient C_L can be recovered from the nine-parameter pitch trajectory model and the ambient weather conditions. Using the inverse of the function $f(\cdot)$ estimated in Section 2.4 we can recover $S|\widehat{\omega} \times \widehat{v}|$ from C_L. The useful spin magnitude $|\overline{\omega}_\perp| = |\overline{\omega}||\widehat{\omega} \times \widehat{v}|$ in equation (4) is then computed using equation (14) by

$$|\overline{\omega}_\perp| = \frac{S|\widehat{\omega} \times \widehat{v}|60|\overline{v}_\mu|}{R2\pi} \tag{15}$$

which allows computation of the spin vector using equation (2)

$$\overline{\omega} = |\overline{\omega}_\perp| \left(\widehat{v} \times \widehat{a}_M\right) \pm \sqrt{|\overline{\omega}|^2 - |\overline{\omega}_\perp|^2}\, \widehat{v} \tag{16}$$

where the ambiguous sign on the gyrospin component is positive for a right-handed pitcher and negative for a left-handed pitcher.

There is significant scatter in the components of the spin vector for a single pitcher and pitch type. Figure 13, for example, is the distribution of $|\overline{\omega}_\perp|$ values for four-seam fastballs for Zack Greinke in 2017. In addition to the scatter we see that the distribution is asymmetric with a heavier tail to the right with a mean of 1584.3 rpm and a median of 1479.8 rpm. Thus, we use robust estimates based on the median to recover the components of the spin vector for individual pitch types for individual pitchers.

3 Evaluating and Comparing Pitchers

3.1 Spin Diagram

A pitch can be represented by a point in the unit circle in the $x-z$ plane from the pitcher's point-of-view where its direction from the origin is given by the direction of the Magnus force

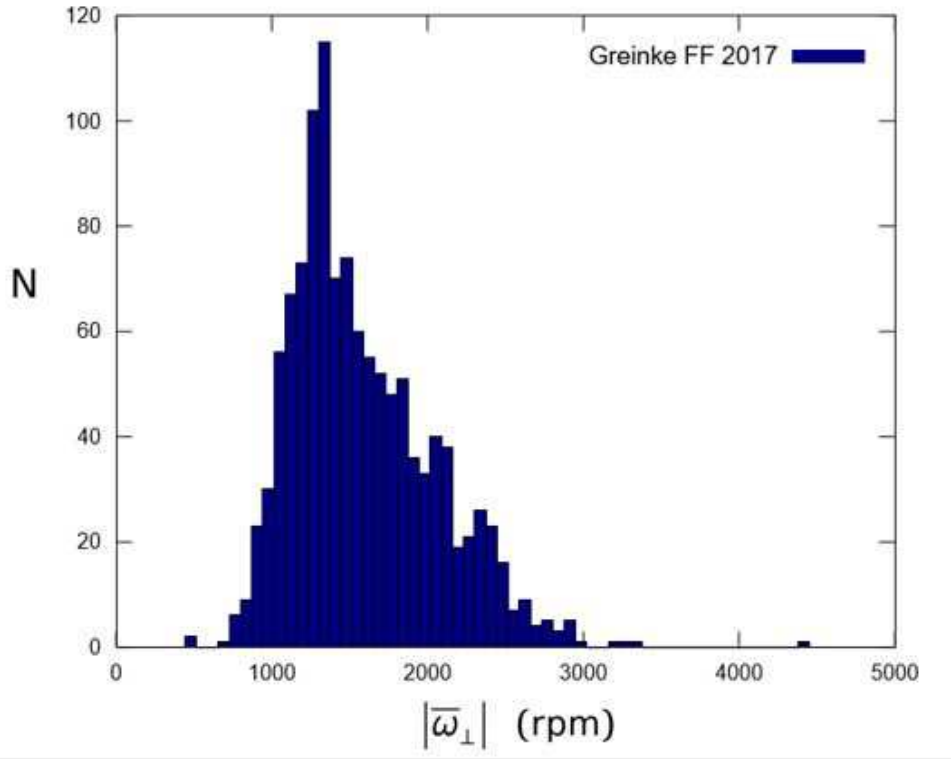

Figure 13: Useful spin magnitude distribution, Zack Greinke four-seam fastballs, 2017

and its distance from the origin is given by its spin efficiency E. We call this representation the spin diagram. As an example, Figure 14 is the spin diagram for the four-seam fastball for Craig Kimbrel and Dellin Betances in 2017. As seen in Figure 6, the Kimbrel fastball has more horizontal movement and a somewhat higher spin efficiency.

The spin diagram can be used to track pitchers over time. Figure 15 is the spin diagram for Justin Verlander in 2017 for four-seam fastballs with Detroit and with Houston. We see that Verlander increased his spin efficiency by making an adjustment with the Astros which led to a significant improvement in performance.

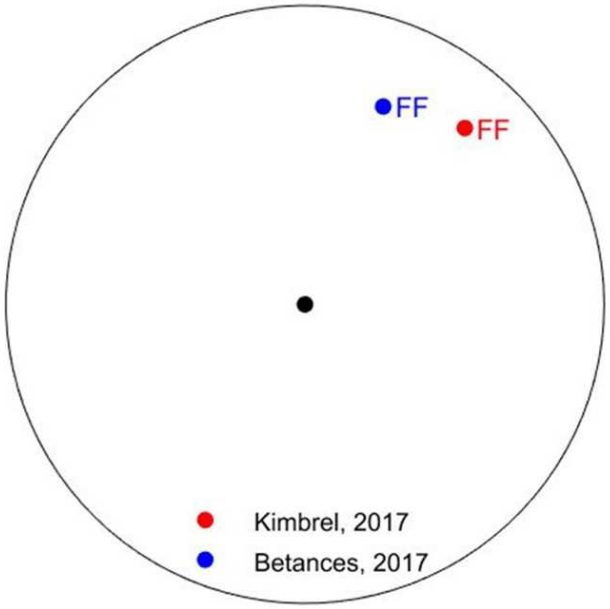

Figure 14: Spin diagram for Kimbrel and Betances four-seam fastball, 2017

Figure 15: Spin diagram for Verlander four-seam fastball, 2017

Figure 16 is the spin diagram for the averages over pitch types for right-handed and left-handed pitchers in 2017. We see that the four-seam fastball has the most upward force and that the curveball has the most downward force. We also see that the fastballs and changeup have the highest spin efficiency.

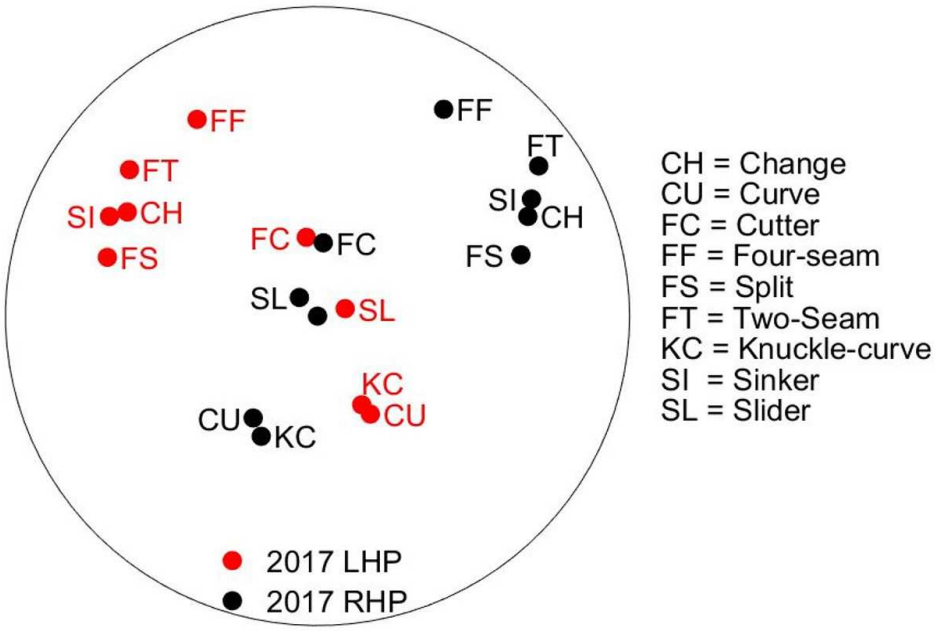

Figure 16: League averages by pitch type, 2017

3.2 Spin Signature

We define the spin signature as a spin diagram for a single pitcher that includes all of his pitch types. The Earth Mover's Distance [10] can be used to compare the spin signatures for two pitchers. A pitcher's degree of uniqueness is defined by the distance of his spin signature to the nearest other pitcher. Tables 1 and 2 present the pitchers with the most unique spin signatures in 2017 and Figures 17 through 20 present the spin signatures for four of the most unique pitchers.

Figure 17 is the spin signature for Charlie Morton for 2017 along with the league averages for the same pitch types. From the clockwise rotation of the FF, FT, and FS pitches we see that Morton has a lower arm slot than the average right-hander. We also see that he has a much higher spin efficiency for these pitches. As a result, he had a higher groundball rate

Unique RHP	Distance to nearest RHP
Charlie Morton	0.2687
Aaron Nola	0.2575
Zack Godley	0.2522
Lance McCullers	0.2415
Hector Neris	0.2379

Table 1: Most unique spin signatures, right-handed pitchers, 2017

Unique LHP	Distance to nearest LHP
Brad Hand	0.4039
Alex Claudio	0.3945
Rich Hill	0.3682
Clayton Kershaw	0.3625
Brent Suter	0.2552

Table 2: Most unique spin signatures, left-handed pitchers, 2017

on these pitches. The spin signature also shows that Morton had a higher spin efficiency on his curve ball than the average right-handed pitcher.

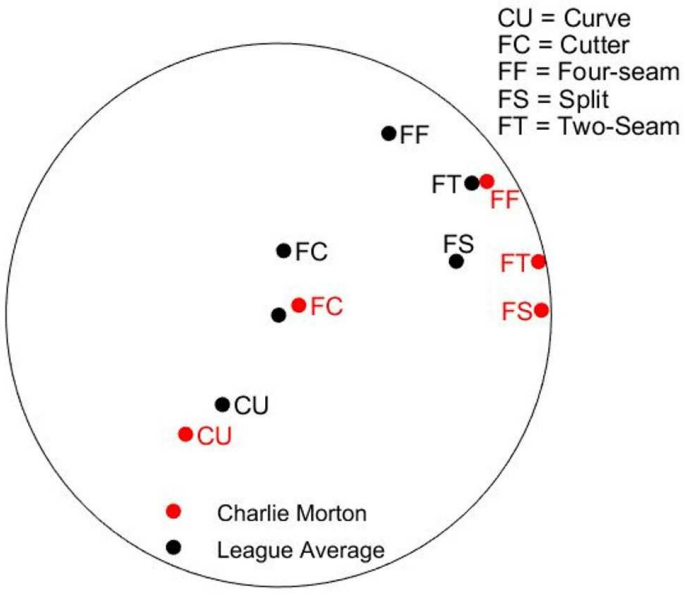

Figure 17: Spin signature, Charlie Morton, 2017

Figure 18 is the spin signature for Aaron Nola for 2017 along with the league averages. As with Morton, we see that Nola had a lower than average arm slot and that his changeup, in particular, had much more downward force than league average. This led to Nola achieving a much higher than average groundball rate on his changeup. We also see that Nola's curve had a much higher spin efficiency than average leading to more glove-side movement and more swings and misses than the average curveball.

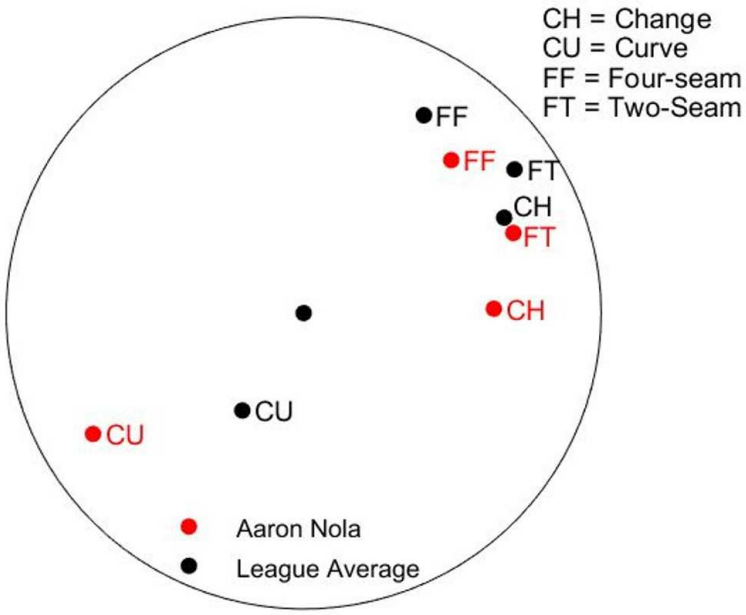

Figure 18: Spin signature, Aaron Nola, 2017

Figure 19 is the spin signature for Brad Hand for 2017 along with the league averages. We see that Hand's uniqueness is largely due to the spin efficiency of his slider which gives this pitch exceptional sweep and dive.

Figure 20 is the spin signature for Alex Claudio for 2017 along with the league averages. We see that Claudio uses a much lower arm angle and obtains much more downward force for his sinker and changeup. This leads to a much higher than average groundball rate for these pitches.

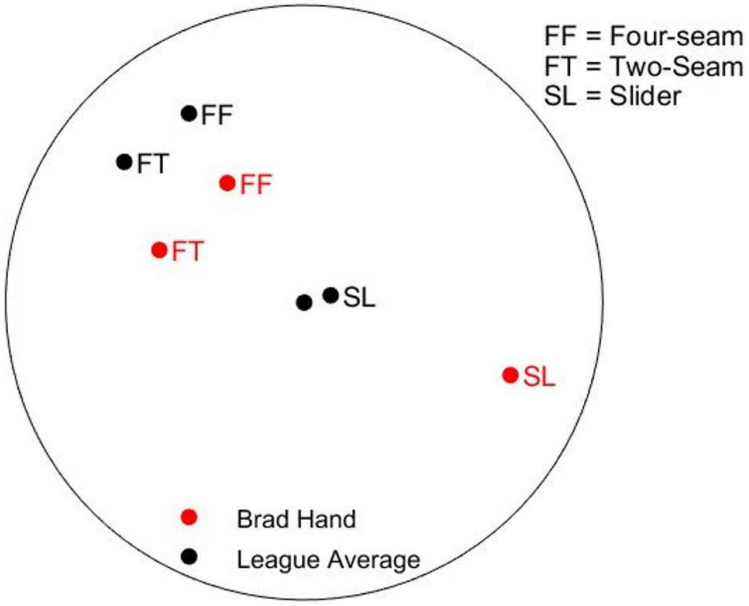

Figure 19: Spin signature, Brad Hand, 2017

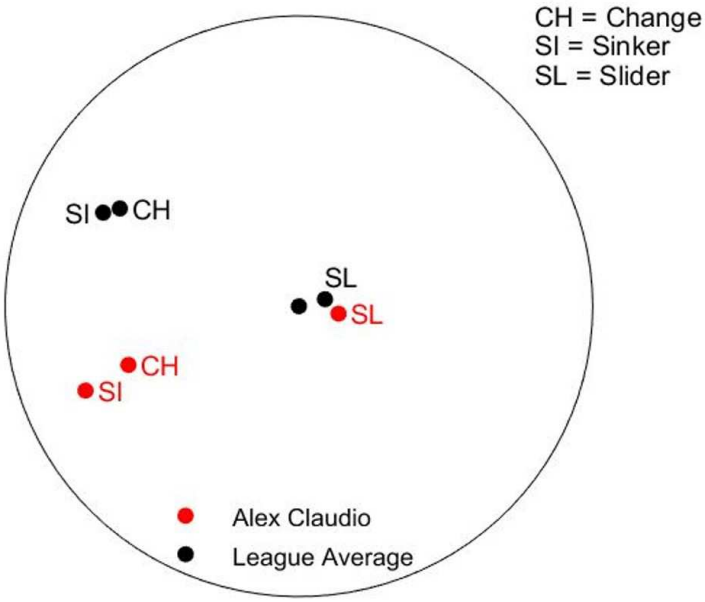

Figure 20: Spin signature, Alex Claudio, 2017

4 Contributions to Baseball Analytics

Pitch movement is a complicated function of the forces on a baseball when it leaves the pitcher's hand and the weather conditions. We have constructed a Trackman pitch database using measurements archived at Baseball Savant. Temperature, relative humidity, and barometric pressure data from Weather Underground for the time and location of each pitch have been combined with altitude data to compute the air density associated with each pitch. By joining the Trackman and weather databases, we can separate the spin components and compute the spin axis for each pitch. After the removal of biases due to Trackman calibration issues, a robust means process is applied to the derived data to relate the lift coefficient and spin parameter for individual pitch types for each pitcher. These data points have led to an improved model for the relationship between the lift coefficient and the scaled spin parameter (Bauer units). This model is then used to recover the useful spin, spin efficiency, and spin vector for pitchers and pitch types. We use these parameters to define the spin diagram for an individual pitch type in terms of the direction of the Magnus force and the spin efficiency and also use this information to define the spin signature for a pitcher's collection of pitches. These signatures are used to compare pitchers by applying the earth mover's distance to the associated distributions.

This work makes several important contributions to baseball analytics. We have developed an improved physical model for pitch movement in terms of the relationship between lift coefficient, transverse spin, and velocity. This model is critical to the accurate calculation of spin efficiency and has led to the definition of the spin diagram and spin signature for pitcher evaluation and comparison. The new model allows the direct comparison of pitch characteristics across various contexts including MLB, MiLB, amateur, and foreign leagues. The identification of similar pitchers increases the sample sizes that can be used to forecast the outcome of batter/pitcher matchups and supports regression to more appropriate population means by projection models. The measure can also be used to monitor pitchers over time and to develop improved models for the aging characteristics associated with different pitcher types. The new approach can also be used to improve pitch classification accuracy and to analyze the effect of altitude and weather on pitch characteristics.

Acknowledgment

I thank Alan Nathan, Lequan Wang, Nikka Mofid, Vu Le, Jason Wang, Kun Han, Selina Wang, Yuanke Zhang, and Dan Brooks for their help.

References

[1] A.T. Bahill, D. Baldwin, and J. Ramberg. Effects of altitude and atmospheric conditions on the flight of a baseball. *International Journal of Sports Science and Engineering*, 3(2):109–128, 2009.

[2] A. Buck. New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology*, 20(2):1527–1532, 1981.

[3] R. Gnanadesikan and J. Kettenring. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28:81–124, 1972.

[4] R. Mehta and J. Pallis. Sports ball aerodynamics: effects of velocity, spin, and surface roughness. In F. Froes and S. Haake, editors, *Materials and Science in Sports*, pages 185–197. TMS, 2001.

[5] A. Nathan. (August 27, 2018). Pitch movement, spin efficiency, and all that [Online]. Available: tht.fangraphs.com/pitch-movement-spin-efficiency-and-all-that.

[6] A. Nathan. (Mar. 31, 2015). Determining the 3D spin axis from TrackMan data [Online]. Available: baseball.physics.illinois.edu/trackman/SpinAxis.pdf.

[7] A. Nathan. (Oct. 21, 2012). Determining pitch movement from PITCHf/x data [Online]. Available: baseball.physics.illinois.edu/Movement.pdf.

[8] A. Nathan. The effect of spin on the flight of a baseball. *Am. J. Phys.*, 76(2):119–124, 2008.

[9] M. O'Connell. (March 29, 2017). Bauer Units and pitch comparison [Online]. Available: www.drivelinebaseball.com/2017/03/bauer-units-pitch-comparison.

[10] Y. Rubner, C. Tomasi, and L. Guibas. The Earth Mover's Distance as a metric for image retrieval. *Int. J. Comp. Vision*, 40(2):99–121, 2000.

[11] G. Schifman. (March 27, 2018). The lurking error in statcast pitch data [Online]. Available: www.fangraphs.com/tht/the-lurking-error-in-statcast-pitch-data.