Combining Radar, Weather, and Optical Measurements to Model the Dependence of Baseball Lift on Spin and Surface Roughness

Glenn Healey and Lequan Wang
Electrical Engineering and Computer Science
University of California, Irvine, CA 92617

Abstract

An accurate model for the lift force on a baseball is important for several applications. The precision of previous models has been limited by the use of small samples of measurements acquired in controlled experiments. The increased prevalence of ball-tracking radar systems provides an abundant source of data for modeling, but the effective use of these data requires overcoming several challenges. We develop a new model that uses this radar data and is constrained by the physical principles and measurements derived from the controlled experiments. The modeling process accounts for the uncertainty in different data sources while exploiting the size and diversity of the radar measurements to mitigate the effects of systematic biases, outliers, and the lack of geometric information that is typically available in controlled experiments. Fine-grained weather data is associated with each radar measurement to enable compensation for the local air density. We show that the new model is accurate enough to capture changes in lift due to small changes in surface roughness which could not be discerned by previous models.

1 Introduction

Baseball is a multibillion dollar industry that is popular in many countries around the world. The mechanics governing many facets of the sport can be represented using physical models [1] [11]. Of particular interest is the flight of a pitch which is a complicated function of the forces on the ball when it leaves a pitcher's hand and the weather conditions. Small changes in pitch trajectory can significantly impact results which underscores the need for accurate models in processes for pitcher evaluation and development [14]. Detailed physical models for pitch trajectories can also support several other applications including the monitoring of sensor calibration systems and properties of the baseball itself.

Key to understanding pitch trajectories is an accurate model for the dependence of the lift force on the velocity and spin of the ball [21]. Given the complex geometry of a baseball with a surface composed of leather pieces that are stitched together to generate a pattern of raised seams, this dependence cannot be derived from first principles but must be measured. The relevant parameters are easiest to measure in controlled laboratory settings using wind tunnels [5][8][34], multi-camera high-speed video systems [2][16][22][26], or light gates [17] [18]. Each of these experiments, however, has generated measurements for fewer than 200 pitches which limits the accuracy of the recovered models.

In recent years, an array of sensors [13] has been deployed that capture several terabytes of data during each Major League Baseball (MLB) game. The Trackman (TM) radar, for example, has captured data for more than 700,000 pitches per year since being introduced as MLB's primary pitch-tracking technology in 2017. The availability of large data sets provides important benefits for modeling especially in the ability to reduce the variance of estimators [28]. But there are also significant challenges. Most MLB games are played outdoors where the weather conditions are uncontrolled. The TM measurements are contaminated by outliers and there are systematic biases in sensor output from site-to-site. In addition, the TM system does not generate the full set of parameters that can be measured in the laboratory.

The effective use of data acquired during MLB games requires that constraints and sources of uncertainty are carefully incorporated into the modeling process. To this end,

the principles and measurements generated by the previous laboratory experiments are used to frame the integration of the TM data into new models. Systematic biases due to pitchers, pitch types, and site calibration differences are accounted for by partitioning the TM data into pitch groups. A robust estimation process is used to mitigate the effect of outliers. A final model is constructed using an optimization method that considers the uncertainty in different measurement sources and leverages the size and diversity of the radar data to overcome the lack of an explicit spin vector measurement. To compensate for variation in weather conditions, we augment the TM data with measurements from weather sensors near the time and location of each pitch. The new approach provides the ability to examine dependence on variables with effects that can be masked by the uncertainty in less precise models. We show that the new model for lift force is sensitive to small changes in surface roughness due to variation in seam height. These effects are important in the quantification of pitcher performance [14], but have not been detectable by models derived from small samples in controlled experiments [3][18][19].

2 Baseball Aerodynamics

2.1 The Flight of a Baseball

A baseball traveling through the air with a translational velocity vector $\overline{v}$ is acted on by three forces as shown in Figure 1. Gravity pulls the ball down, drag acts opposite the velocity direction, and the Magnus force [7] causes the ball to change direction due to spin. The Magnus force depends on the spin vector $\overline{\omega}$ which 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. The magnitude of the Magnus force [26] 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. If we define the velocity and spin vector directions by the unit vectors $\widehat{v} = \overline{v}/|\overline{v}|$

and $\hat{\omega} = \overline{\omega}/|\overline{\omega}|$, then the Magnus force is in the direction of $\hat{\omega} \times \hat{v}$.

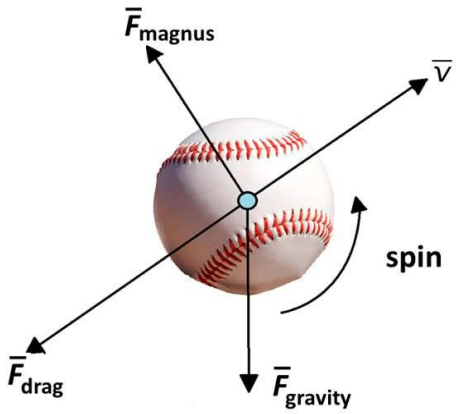

Figure 1: Forces on a spinning baseball in flight

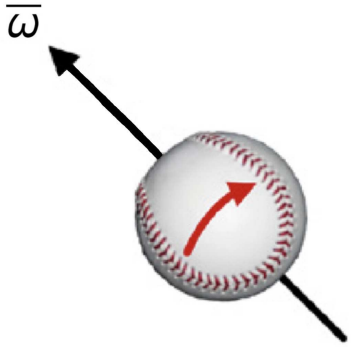

Figure 2: Spin vector $\overline{\omega}$

The spin vector $\overline{\omega}$ can be written as

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

where $\overline{\omega}_{\parallel}$ is parallel to $\hat{v}$ and $\overline{\omega}_{\perp}$ is perpendicular to $\hat{v}$. $\overline{\omega}_{\parallel}$ is known as the gyro component of the spin and does not contribute to the Magnus force [27]. The magnitude of $\overline{\omega}_{\perp}$ is given by

$$|\overline{\omega}_{\perp}| = |\overline{\omega}||\hat{\omega} \times \hat{v}|. \tag{3}$$

The dimensionless spin parameter S [21] plays an important role in determining C_L and is defined as the ratio of the speed of the ball surface relative to its center to the translational speed of the ball center

$$S = \frac{2\pi R|\overline{\omega}|}{|\overline{v}|} \tag{4}$$

where R is the ball radius.

2.2 The Relationship Between C_L and S

Watts and Bahill [33] speculated in 1990 that the lift coefficient C_L depends on the ratio $|\widehat{\omega} \times \widehat{v}|$ of $|\overline{\omega}_{\perp}|$ to $|\overline{\omega}|$ and Jinji and Sakurai [16] later confirmed this using an experiment with measurements for 168 pitches using a set of synchronized video cameras. Nagami et al. [22] used a similar setup to make measurements for 75 pitches and to show experimentally that

$$C_L = f(S)|\widehat{\omega} \times \widehat{v}| \tag{5}$$

where $f(S)$ is an increasing function of S with $f(0) = 0$. They also showed that this conclusion was consistent with previous video-based optical measurements made by Nathan [26] (22 pitches) and Alaways and Hubbard [2] (17 pitches) for the special case where $|\widehat{\omega} \times \widehat{v}| = 1$. The studies reported in [2] and [26] assigned uncertainties to the measurements using methods described in the articles.

A frequently used approximation to $f(S)$ was presented in [11] that is based on a fit of experimental data from several sources including [2][16][26]. This data, however, is based on a relatively small number of pitch measurements as detailed above. These measurements also have significant scatter, particularly in the region $0.1 \leq S \leq 0.3$ which is most relevant for MLB pitches. In this work we consider estimating the function $f(S)$ by combining these optical measurements with a large set of radar measurements collected during MLB games. The function $f(S)$ is important for several applications that we will discuss in Section 6.

3 Radar Measurements

3.1 Pitch Data

The Trackman (TM) phased-array Doppler radar operates in the X-band at approximately 10.5 GHz and has been used to measure 3-D pitch trajectories and spin information for over two million pitches thrown in MLB games between 2017 and 2019. The TM system generates a nine-parameter model for each pitch in terms of the three-dimensional acceleration vector $\overline{a} = (a_x, a_y, a_z)$ which is assumed constant over the pitch trajectory and the three-dimensional velocity and position vectors for a point on the trajectory. These parameters can be used to recover the full path of the pitch from the measured release point using the equations of motion. The system also estimates the magnitude of the spin vector $|\overline{\omega}|$ from the distribution of Doppler shifts.

3.2 Estimating C_L and S

The TM radar data can be used to estimate the lift coefficient C_L and the spin parameter S for each pitch using equations (1) and (4). Since both C_L and S depend on the velocity magnitude $|\overline{v}|$ which is not constant, we use the mean velocity magnitude $|\overline{v}_\mu|$ over each pitch trajectory to construct the estimates. A similar approach has been used in previous studies [24]. The acceleration vector recovered by the TM system 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 $\widehat{v}_\mu = \overline{v}_\mu / |\overline{v}_\mu|$. Therefore the Magnus acceleration is given by

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

and using equation (1) the lift coefficient for the pitch trajectory can be estimated 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 $\overline{a}_M$. Equation (4) can be used to compute the spin parameter for the trajectory using

$$S = \frac{2\pi R|\overline{\omega}|}{|\overline{v}_\mu|}. \tag{10}$$

Each quantity on the right-hand side of equations (9) and (10) is known or can be recovered from the TM measurements except for the air density ρ.

4 Computing Air Density

The air density ρ can be computed from the altitude, temperature, relative humidity, and barometric pressure. MLB provides the temperature for the start of each game and information on whether a retractable roof is open or closed. We obtained additional weather information by identifying the three closest weather stations that report on Weather Underground (wunderground.com) for each MLB stadium. Using the time stamps provided by the TM system, we determined the closest station that reported within thirty minutes of each pitch. For pitches with multiple weather reports from the closest station within this time window, we associated the closest weather data in time. For domed stadiums or cases where a retractable roof was closed, the air density was computed using the MLB game temperature, a relative humidity of 50 percent, and the barometric pressure retrieved from a nearby weather station as described above. The altitude for each MLB stadium was obtained from the Seamheads Ballparks Database (seamheads.com). We used this approach to assign weather data to the 2.164 million pitches analyzed in this study over the 2017 to 2019 MLB seasons. The average time difference between pitches and weather data was

14.06 minutes and the average distance between the stadium and the weather station used for the measurement was 1.93 km.

The air density ρ associated with each pitch was computed in units of kg/m^3 using the model from [6] given by

$$\rho = \frac{(1.2929 * 273.0)(P - 0.01VH)}{760.0(T + 273.0)} \tag{11}$$

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

$$P = b * \exp\left[(-gME)/(RT + 273.15)\right] \tag{12}$$

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 [9] given by

$$V = 4.5841 * \exp\left[\frac{(18.687 - T/234.5)T}{257.14T}\right] \ . \tag{13}$$

The estimated value of ρ obtained using equations (11), (12), and (13) is used in (9) to complete the estimate of C_L.

5 Recovering $f(S)$

5.1 Sources of Scatter in (S, C_L) Data

The goal of this work is to find the function $f(S)$ that defines the mapping $C_L = f(S)|\widehat{\omega} \times \widehat{v}|$. We showed in Sections 3 and 4 that by combining TM measurements and weather data we can estimate S and C_L for each pitch. In contrast to the video setups described in Section 2.2, however, the TM system does not allow direct measurement of $|\widehat{\omega} \times \widehat{v}|$. To alleviate this difficulty, we can consider using the domain knowledge that $|\widehat{\omega} \times \widehat{v}|$ is often 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 $f(S)$. Figure 3 is a scatterplot of C_L

versus S for pitches classified as 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.

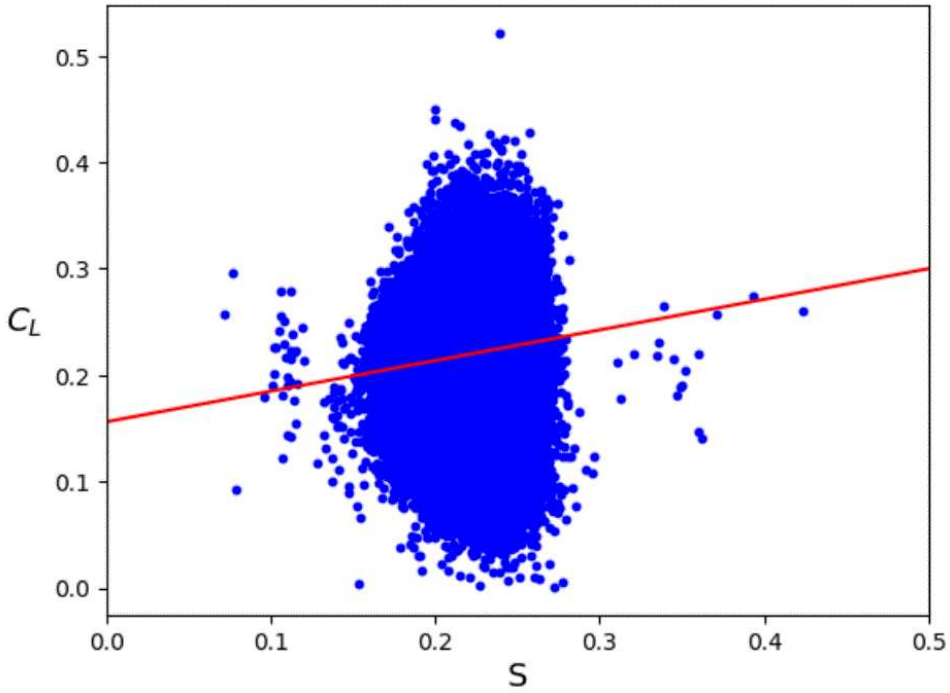

Figure 3: C_L versus S for four-seam fastballs, 2017

There are several sources that contribute to the scatter in Figure 3. We examine a few of these sources in more detail in Figure 4 which plots the S and C_L values for four-seam fastballs thrown by pitcher Ervin Santana in 2017. The pitches in blue were thrown at Santana's home ballpark, Target Field in Minnesota, and the pitches in red were thrown at twelve different ballparks when his team was not playing at home. We see that there is a significant positive bias in the C_L values for Target Field which can be traced to calibration issues with the TM system at that site [31]. If we restrict the analysis to either the Target Field games (blue points) or the away games (red points), we see that there is still substantial scatter from pitch-to-pitch which includes multiple outliers at a significant distance from the center of the distributions. This scatter is due to natural variation in pitches, variation in the physical properties of the baseball [4], sensor noise, and pitch classification errors.

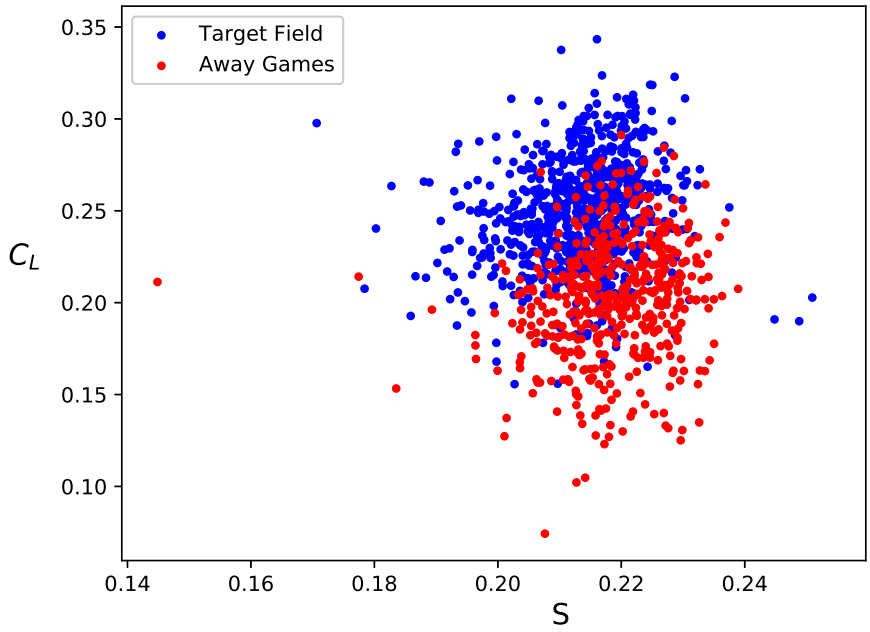

Figure 4: C_L versus S for Ervin Santana, Four-Seam Fastball, 2017

5.2 Robust Estimates and Uncertainty

A specific pitch type thrown by a particular pitcher will have unique velocity and spin characteristics which can change from year-to-year as a pitcher ages and makes adjustments. Thus, we generate a set of (S, C_L) points from the TM data by considering separately pitches corresponding to a specific pitcher, pitch type, and year. We reduce the effects of ballpark bias by only considering pitches thrown by a pitcher in away games. After imposing this constraint, we identify all (pitcher, pitch type, year) pitch groups, e.g. (Ervin Santana, Four-seam fastball, 2017), which include at least 200 pitches. There were a total of 1678 of these groups in our data set which were nearly equally distributed over the three years with 549 in 2017, 565 in 2018, and 564 in 2019.

For a given pitch group, we reduce the measurements to a single estimate of (S, C_L). Since the data is contaminated by outliers, we use robust estimates based on the sample median [15]. Figure 5 demonstrates the action of the sample median as compared to the standard sample mean. This figure plots the histogram of the C_L values for the 294 pitches in the (CC Sabathia, Sinker, 2017) group after restricting to away games. We see that the

distribution includes outliers to the right which contribute to the mean of 0.191 exceeding the median of 0.178. To minimize the impact of outliers, the (S, C_L) estimate for each pitch group is given by $(\widehat{S}, \widehat{C}_L)$ where $\widehat{S}$ is the sample median of the group S values and $\widehat{C}_L$ is the sample median of the group C_L values.

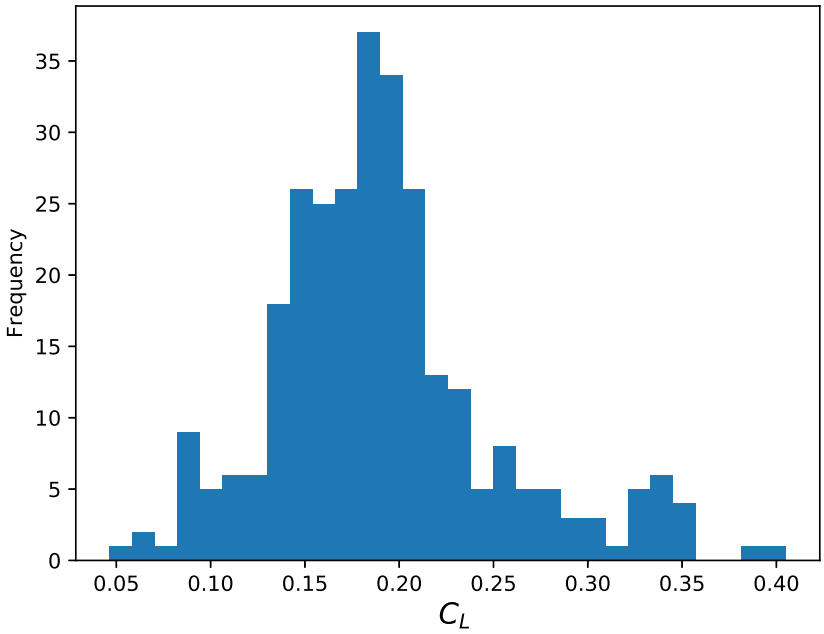

Figure 5: Distribution of C_L for CC Sabathia, sinker, 2017, away games

For a sample of size n derived from a distribution with probability density $p(x)$ and median m, the uncertainty in the sample median $\widehat{m}$ can be approximated by the asymptotic variance of the estimator [29] which is given by

$$\mathrm{Var}(\widehat{m}) = \frac{1}{4np^2(m)} \cdot \tag{14}$$

For a computed $\widehat{m}$, we can approximate the right-hand side of equation (14) by evaluating a kernel density estimate [32] for $p(x)$ at the sample median $\widehat{m}$. Figure 6 illustrates this process for the $\widehat{C}_L$ estimate for two pitch groups. The first group is (Matt Boyd, Changeup, 2017) with a size of 244 pitches and a sample median of $\widehat{C}_{L1} = 0.191$. The second group is (Marco Estrada, Four-seam fastball, 2017) with a size of 908 pitches and a sample median of $\widehat{C}_{L2} = 0.264$. The kernel density estimates $\widehat{p}_1(C_L)$ and $\widehat{p}_2(C_L)$ for these two groups are

plotted in Figure 6 and yield values of $\widehat{p}_1(\widehat{C}_{L1}) = 5.793$ and $\widehat{p}_2(\widehat{C}_{L2}) = 17.975$. Equation (14) then yields a standard deviation for $\widehat{C}_{L1}$ of $1/(2 * \sqrt{244} * 5.793) = 0.00553$ and for $\widehat{C}_{L2}$ of $1/(2 * \sqrt{908} * 17.975) = 0.00092$. Thus, Group 2 has a significantly smaller uncertainty due to both its larger sample size and its more concentrated distribution.

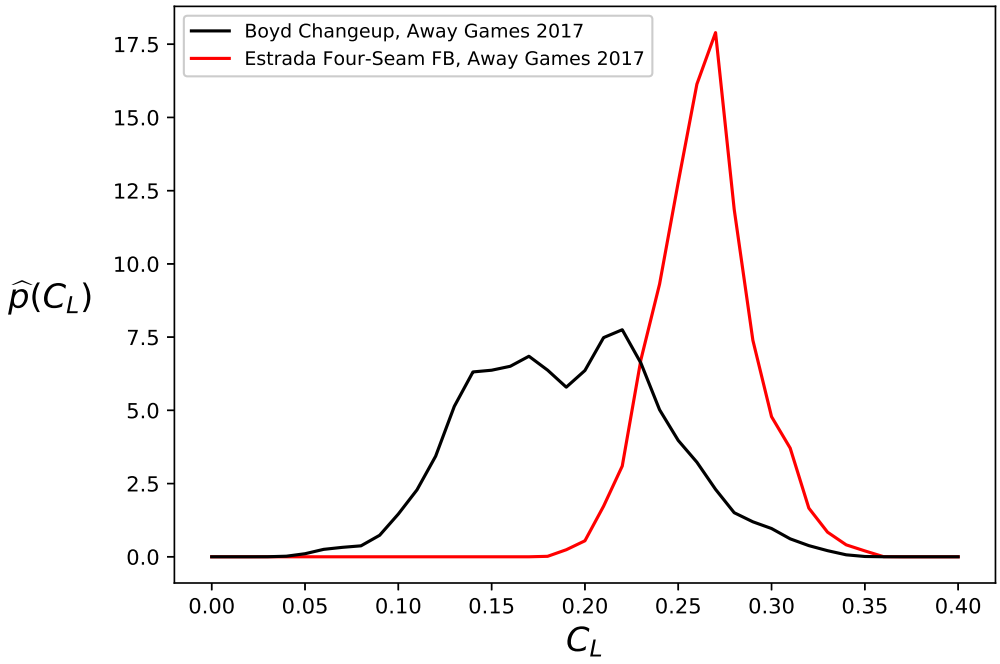

Figure 6: Kernel density estimates for C_L, Boyd and Estrada pitch groups, 2017

5.3 Combining Radar and Optical Measurements

Figure 7 is a scatterplot of the 1678 $(\widehat{S}, \widehat{C}_L)$ points generated using the method described in the previous section. We see that there is still significant variation in C_L for a given S. Since C_L depends on both S and $|\widehat{\omega} \times \widehat{v}|$, this variation is due largely to differences in $|\widehat{\omega} \times \widehat{v}|$ for different pitch groups. If we assume that there are groups for which $|\widehat{\omega} \times \widehat{v}| = 1$, then we can estimate $f(S)$ by finding a curve that is an upper bound to the $(\widehat{S}, \widehat{C}_L)$ points. To improve the accuracy of the estimate, we can also consider the use of the optical video measurements with assigned uncertainties [2][26] that were described in Section 2.2. In addition to having known values for $|\widehat{\omega} \times \widehat{v}|$, the optical data also includes measurements over a wider range of S values.

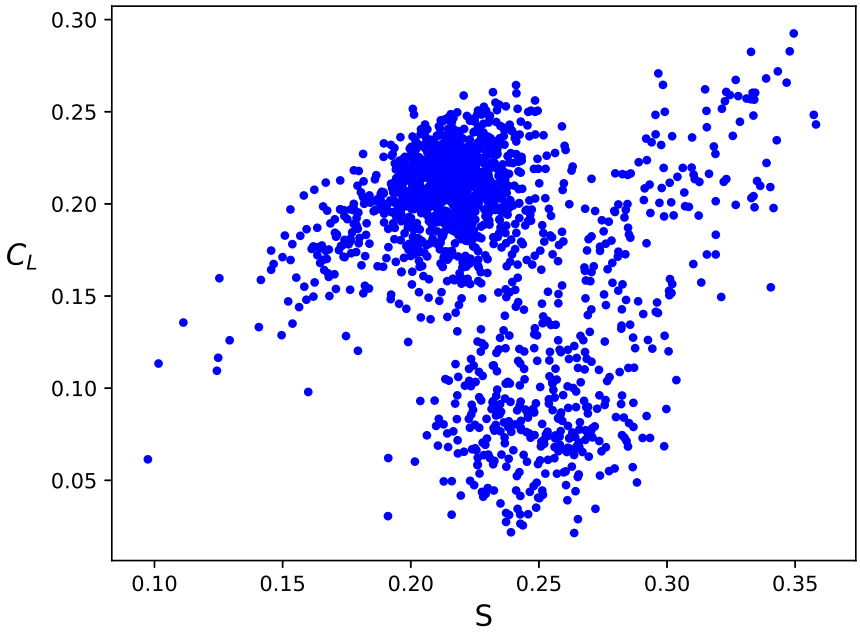

Figure 7: $\widehat{C}_L$ versus $\widehat{S}$ for 1678 TM pitch groups over 2017 to 2019

Let $(S_{O}(i), C_{LO}(i))$ for $1 \leq i \leq N_O$ denote the set of (S, C_L) points estimated using the optical video-based techniques and let $\sigma_{O}(i)$ be the standard deviation of each $C_{LO}(i)$. Let $(S_{T}(i), C_{LT}(i))$ for $1 \leq i \leq N_T$ denote the set of $(\widehat{S}, \widehat{C}_L)$ points recovered from the TM data and let $\sigma_{T}(i)$ be the standard deviation of each $C_{LT}(i)$ as computed using the approximation to equation (14) described in Section 5.2. Given a set of possible approximating functions $f(S)$, we define the optimizing function as the one that minimizes the sum of the absolute errors weighted by the standard deviations of the measurements

$$E = \sum_{i=1}^{N_O} \left(\frac{E_{O}(i)}{\sigma_{O}(i)} \right) + \sum_{i=1}^{N_T} \left(\frac{E_{T}(i)}{\sigma_{T}(i)} \right) \tag{15}$$

where

$$E_{O}(i) = |C_{LO}(i) - f(S_{O}(i))| \tag{16}$$

and

$$E_{_T}(i) = \begin{cases} C_{_{LT}}(i) - f(S_{_T}(i)) & \text{if } C_{_{LT}}(i) \geq f(S_{_T}(i)) \\ 0 & \text{if } C_{_{LT}}(i) < f(S_{_T}(i)) \end{cases} \tag{17}$$

Since the optical measurements were generated for the $|\widehat{\omega} \times \widehat{v}| = 1$ configuration, the error $E_{_O}(i)$ for each point $(S_{_O}(i), C_{_{LO}}(i))$ is considered in E. Since $|\widehat{\omega} \times \widehat{v}|$ may be less than one for the radar measurements, the error $E_{_T}(i)$ in equation (17) only contributes to E if a point $(S_{_T}(i), C_{_{LT}}(i))$ is above the approximating function value $f(S_{_T}(i))$.

We applied this optimization method to the TM data ($N_T = 1678$) in combination with the optical video data ($N_O = 39$) from Alaways and Hubbard [2] and Nathan [26]. After considering a range of increasing parametric functions, we found that a Hill function of the form

$$C_L(S) = \frac{AS^n}{a^n + S^n} \tag{18}$$

with parameters $A = 0.370$, $n = 1.651$, and $a = 0.137$ gave the best fit using the error measure in equation (15). Figure 8 plots this new model $\widehat{f}(S)$ along with the TM data and the optical video data. The Previous Model curve in the figure was presented in [11] as a representation for several sets of optical measurements [26] and is given by

$$C_L(S) = \frac{S}{2.32S + 0.4} \ . \tag{19}$$

We see that there are significant differences between the models represented by equations (18) and (19) particularly in the range $0.20 \leq S \leq 0.35$ which is important for MLB pitches.

5.4 Sensitivity to Additional Variables: Surface Roughness

We have seen that the lift coefficient C_L has a strong dependence on the spin parameter S. Watts and Ferrer [34] suggested in 1987 that C_L may also depend on the surface roughness. For a baseball, surface roughness is often defined in terms of the seam height. Laboratory measurements [17] reported in 2011 confirmed the Watts/Ferrer hypothesis by demonstrating that the high-seamed collegiate ball had a lift coefficient that was measurably greater

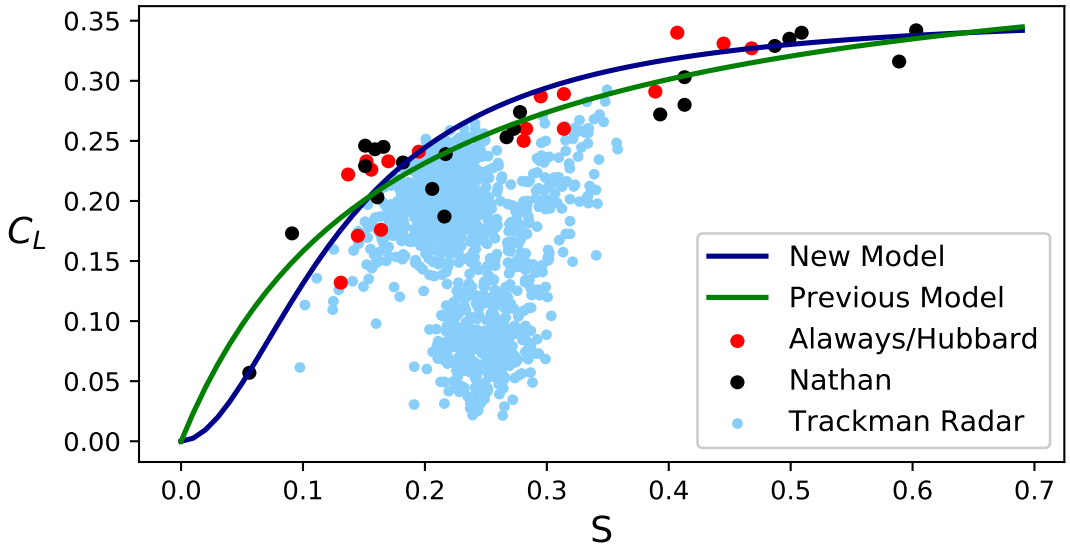

Figure 8: New model $\widehat{f}(S)$ with points recovered from TM and optical systems

than for the lower-seamed MLB ball. Processes have been developed to accurately measure seam height [19] and a more recent laboratory study [18] found no discernible difference in C_L across seam height variation from 0.034 inches to 0.046 inches. Other experiments [3] [19] that also considered fewer than 100 measurements were similarly unable to detect dependence of C_L on seam height. Over the three years in our current study, the average seam height for the MLB ball decreased from 0.0329 inches in 2017 to 0.0305 inches in 2019 with significant variation in seam height during each year [4]. We can ask whether the process for modeling C_L developed in this work is sensitive to these small changes in average seam height.

Figure 9 plots the twenty $(\widehat{S}, \widehat{C}_L)$ points in Figure 8 with the largest ratios $\widehat{C}_L/\widehat{f}(\widehat{S})$ for each of the three years. These are the points with the highest estimated lift coefficient $\widehat{C}_L$ relative to the optimizing function value $\widehat{f}(\widehat{S})$. We see that for a fixed S, the 2017 points tend to have the largest $\widehat{C}_L$ values which is consistent with 2017 having the largest average seam height. If we separately find a model of the form $K\widehat{f}(S)$ for each year that minimizes E in equation (15) by considering only the measurements in the TM range $(0.1 \leq S \leq 0.35)$, we find $K_{2017} = 1.000, K_{2018} = 0.970$, and $K_{2019} = 0.968$. Figure 10 is a scatterplot of K versus average seam height for each year and shows that the model can account for increases

in C_L due to small increases in seam height.

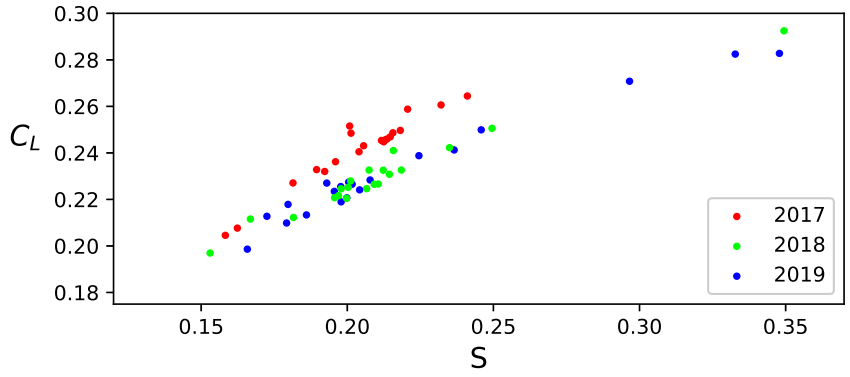

Figure 9: $(\widehat{S}, \widehat{C}_L)$ points with the largest $\widehat{C}_L/\widehat{f}(\widehat{S})$ for each year

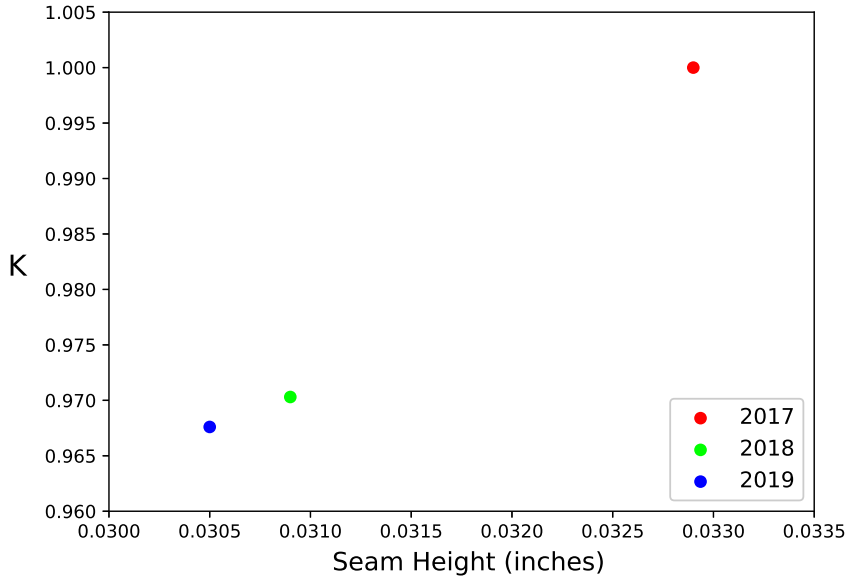

Figure 10: Lift coefficient dependence on average seam height

All other things being equal, we would expect that a decrease in C_L would lead to a corresponding decrease in pitch movement. Figure 11 plots the average length of the (pfx_x,pfx_z) movement vector over all pitches for the three seasons versus average seam height and we see a trend that is similar to the trend in Figure 10. We observe that even these relatively small changes in movement can have a meaningful impact on the effectiveness of pitches [14].

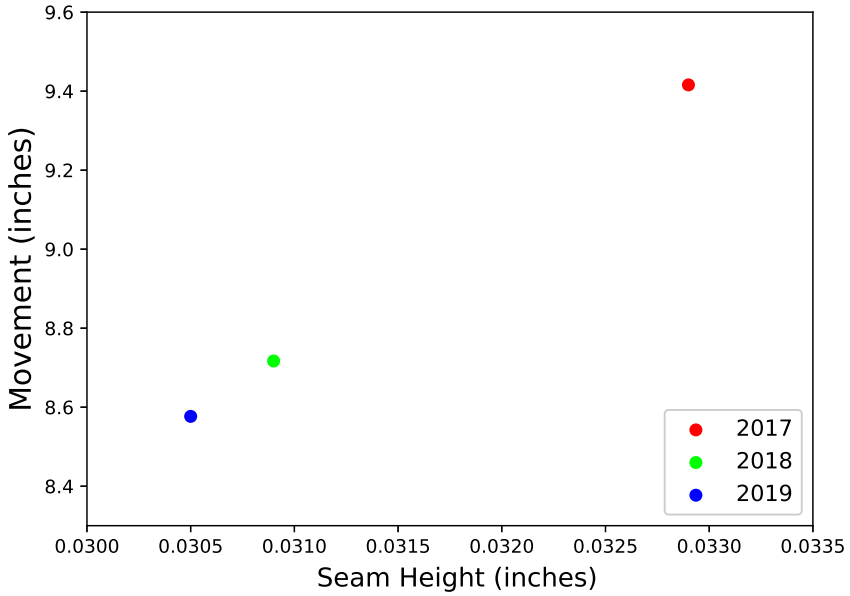

Figure 11: Pitch movement dependence on average seam height

6 Applications

In this section, we present applications that make use of the recovered $\widehat{f}(S)$ function for monitoring TM system calibration and for supporting pitcher evaluation and development.

6.1 Monitoring Sensor Calibration

The function $\widehat{f}(S)$ represents the maximum value of the lift coefficient C_L for spin parameter S. Since C_L and S can be estimated from the TM data for each pitch using the method described in Section 3.2, we can use this function to monitor the accuracy of the TM calibration. For this purpose, we consider the distribution of the ratio $C_L/\widehat{f}(S)$ where C_L and S are estimated for each pitch in a game. As an example, Figure 12 plots the distribution of this ratio for the TM measurements for an MLB game played at SunTrust Park in Atlanta on 11 June 2017. This is a fairly typical distribution with a mean of 0.72 and a maximum of 1.4 with a fraction of the ratios exceeding one due to the sources of scatter described in Section 5.1. Figure 13 plots the distribution of ratios for the next MLB game played at SunTrust Park which occurred on 16 June 2017. We see that the distribution is significantly shifted to the right due to calibration errors with a mean of 1.37 and a number of ratios

exceeding 2.0. Given the large disparity between the distributions in Figures 12 and 13, sensor calibration errors could be rapidly detected in real-time using $\widehat{f}(S)$.

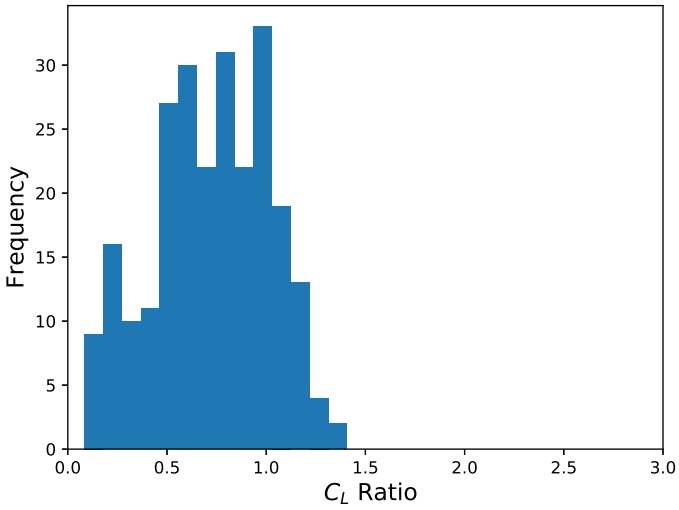

Figure 12: Distribution of $C_L/\widehat{f}(S)$ for SunTrust Park, 11 June 2017

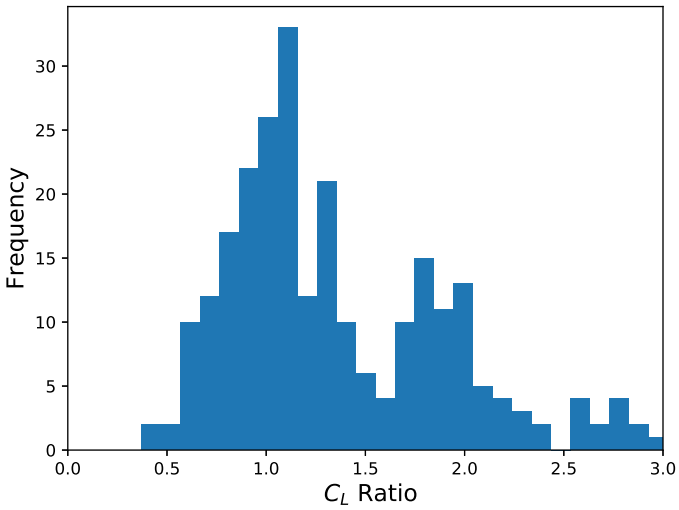

Figure 13: Distribution of $C_L/\widehat{f}(S)$ for SunTrust Park, 16 June 2017

6.2 Recovering the Spin Vector

The function $\widehat{f}(S)$ is the link that enables the spin vector $\overline{\omega}$ to be derived from TM data [24]. This vector and the associated active spin fraction [23] are important tools for pitcher development and evaluation [20]. The active spin fraction $|\overline{\omega}_\perp|/|\overline{\omega}|$ measures the proportion of the spin vector magnitude that is transferred to the Magnus force $\overline{F}_M$. The spin vector $\overline{\omega}$, more generally, determines the magnitude and direction of $\overline{F}_M$ as detailed in Section 2. The Magnus force causes a change in a pitch's location in a predefined plane which is described by a movement vector [25]. The direction and magnitude of the movement vector have been shown to be key determinants of a pitch's effectiveness [14]. Since a pitcher controls the spin vector with the orientation of his hand and fingers when he releases a pitch, the availability of measurements of the spin vector under game conditions can allow for pitcher mechanics to be monitored and can streamline the process of refining pitches to achieve desired results.

We briefly summarize the use of the estimated $\widehat{f}(S)$ function to recover the active spin fraction and the spin vector. The active spin fraction is defined by

$$A = \frac{|\overline{\omega}_\perp|}{|\overline{\omega}|} \tag{20}$$

which simplifies to $A = |\widehat{\omega} \times \widehat{v}|$ using equation (3). Since C_L and S can be estimated from TM data for a pitch as described in Section 3.2, A can be estimated using equation (5) by

$$\widehat{A} = \frac{C_L}{\widehat{f}(S)} \cdot \tag{21}$$

In a similar way, $\widehat{f}(S)$ can be used to estimate the spin vector $\overline{\omega}$. As described in Section 2.1, the $\overline{\omega}_\parallel$ component of the spin vector $\overline{\omega}$ is parallel to $\widehat{v}$ and can be expressed as

$$\overline{\omega}_\parallel = |\overline{\omega}_\parallel| \widehat{v} \tag{22}$$

where since $\overline{\omega}_\parallel$ and $\overline{\omega}_\perp$ are perpendicular we can write

$$\overline{\omega}_\parallel = \sqrt{|\overline{\omega}|^2 - |\overline{\omega}_\perp|^2}\ \widehat{v}. \tag{23}$$

Since the $\overline{\omega}_\perp$ component of the spin vector is perpendicular to both the velocity vector $\overline{v}$ and the Magnus acceleration $\overline{a}_M$, the direction of $\overline{\omega}_\perp$ is specified by the unit vector $\widehat{v} \times \widehat{a}_M$

where $\widehat{a}_M = \overline{a}_M/|\overline{a}_M|$. Using equation (3) we can write

$$\overline{\omega}_\perp = |\overline{\omega}||\widehat{\omega} \times \widehat{v}|(\widehat{v} \times \widehat{a}_M) \tag{24}$$

where the estimate $\widehat{A}$ from (21) can be used for $|\widehat{\omega} \times \widehat{v}|$ and the remaining quantities in equations (23) and (24) are either directly measured by the TM radar or can be derived as described in Section 3. This allows an estimate of $\overline{\omega}$ to be generated by combining the right-hand sides of equations (23) and (24) according to

$$\overline{\omega} = |\overline{\omega}||\widehat{\omega} \times \widehat{v}|\,(\widehat{v} \times \widehat{a}_M) \pm \sqrt{|\overline{\omega}|^2 - |\overline{\omega}_\perp|^2}\;\widehat{v} \tag{25}$$

where the ambiguous sign is positive for a right-handed pitcher and negative for a left-handed pitcher. We observe that measurements derived from individual pitches can have substantial scatter as discussed in Section 5.1. This suggests the use of robust estimates over pitch groups and a careful consideration of uncertainty as described in Section 5.2 when using $\widehat{f}(S)$ to estimate A and $\overline{\omega}$.

7 Conclusion

The use of sensor systems to acquire measurements at sporting events has enabled a range of new applications [10][13][35]. In this work, we have developed a new model for the lift coefficient of a baseball as a function of the spin parameter. The model was built by combining a large set of TM radar measurements made under uncontrolled game conditions with smaller sets of optical measurements made under controlled laboratory conditions. Fine-grained weather data was used to associate an air density estimate with each radar measurement. The modeling process considers the uncertainty in the various measurements and leverages the size and diversity of the radar data to address the presence of biases and outliers. We show that the new model is sensitive to changes in the lift coefficient due to small changes in seam height which cannot be distinguished by previous models. This level of accuracy provides the potential for day-to-day monitoring of the physical properties of the baseball which has become a topic of considerable interest [3] [4]. The new model may also be helpful in quantifying other small effects that have been difficult to measure such

as a side force that can occur if the ball is rougher on one side over a significant fraction of its trajectory [12][33]. We have described several additional applications of this approach in areas that include player evaluation [30] and development [20] as well as system calibration. We expect that the new model can improve the utility of the TM radar system which is currently used at many professional, college, and high school baseball facilities.

Acknowledgment

We thank Alan Nathan and Lloyd Smith for generously providing measurements that have been used in this work. Pitch-tracking data was obtained from baseballsavant.mlb.com and weather data was obtained from wunderground.com.

References

[1] R. Adair. *The Physics of Baseball*. Perennial, New York, 3rd edition, 2002.

[2] L. Alaways and M. Hubbard. Experimental determination of baseball spin and lift. *Journal of Sports Sciences*, 19:349–358, 2001.

[3] J. Albert, J. Bartroff, R. Blandford, D. Brooks, J. Derenski, L. Goldstein, A. Hosoi, G. Lorden, A. Nathan, and L. Smith. (May 24, 2018). Report of the committee studying home run rates in Major League Baseball. Available: www.mlb.com/documents/7/9/4/278128794/Full_Report_of_the_Committee_Studying_Home_Run _Rates_in_Major_League_Baseball_052418.pdf.

[4] J. Albert, A. Hosoi, A. Nathan, and L. Smith. (Dec. 10, 2019). Preliminary report of the committee studying home run rates in MLB. Available: mlb.com/documents/0/5/6/312149056/Report_of_the_Committee_Studying_Home_Run_Rates_in_MLB_121119.pdf.

[5] K. Aoki, Y. Kinoshita, J. Nagase, and Y. Nakayama. Dependence of aerodynamic characteristics and flow pattern on surface structure of a baseball. *Journal of Visualization*, 6(2):185–193, 2003.

[6] 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.

[7] H. Barkla and L. Auchterlonie. The Magnus or Robins effect on rotating spheres. *Journal of Fluid Mechanics*, 47(3):437–447, 1971.

[8] L. Briggs. Effect of spin and speed on the lateral deflection (curve) of a baseball; and the Magnus effect for smooth spheres. *Am. J. Phys.*, 27(8):589–596, 1959.

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

[10] K. Clark. (Dec. 19, 2018). The NFL's analytics revolution has arrived [Online]. Available: www.theringer.com/nfl/2018/12/19/18148153/nfl-analytics-revolution.

[11] R. Cross. *Physics of Baseball & Softball*. Springer, New York, 2011.

[12] R. Cross. Aerodynamics in the classroom and at the ball park. *Am. J. Phys.*, 80(4):289–297, 2012.

[13] G. Healey. The new Moneyball: How ballpark sensors are changing baseball. *Proceedings of the IEEE*, 105(11):1999–2002, 2017.

[14] G. Healey. A Bayesian method for computing intrinsic pitch values using kernel density and nonparametric regression estimates. *Journal of Quantitative Analysis in Sports*, 15(1):59–74, March 2019.

[15] P. Huber and E. Ronchetti. *Robust Statistics*. Wiley, Hoboken, NJ, 2nd edition, 2009.

[16] T. Jinji and S. Sakurai. Direction of spin axis and spin rate of the pitched baseball. *Sports Biomechanics*, 5(2):197–214, 2006.

[17] J. Kensrud and L. Smith. In situ lift measurement of sports balls. *Procedia Engineering*, 13:278–283, 2011.

[18] J. Kensrud and L. Smith. Drag and lift measurements of solid sports balls in still air. *Proceeedings of the Institution of Mechanical Engineers Part P: Journal of Sports Engineering and Technology*, 232(3):255–263, 2018.

[19] J. Kensrud, L. Smith, A. Nathan, and D. Nevins. Relating baseball seam height to carry distance. In *7th Asia-Pacific congress on sports technology (APCST 2015): The impact of technology on sport VI*, pages 406–411, September 2015.

[20] B. Lindbergh and T. Sawchik. *The MVP machine: How baseball's new nonconformists are using data to build better players*. Basic Books, New York, NY, 2019.

[21] 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.

[22] T. Nagami, T. Higuchi, H. Nakata, T. Yanai, and K. Kanosue. Relation between lift force and ball spin for different baseball pitches. *Journal of Applied Biomechanics*, 32:196–204, 2016.

[23] A. Nathan. (March 31, 2015). All spin is not alike [Online]. Available: www.baseballprospectus.com/article.php?articleid=25915.

[24] A. Nathan. (May 25, 2020). Determining the 3D spin axis from TrackMan data [Online]. Available: baseball.physics.illinois.edu/trackman/SpinAxis.pdf.

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

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

[27] A. Nathan and D. Baldwin. An analysis of the gyroball. *Baseball research journal*, 36:77–80, 2007.

[28] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 3rd edition, 1991.

[29] P. Rider. Variance of the median of small samples from several special populations. *Journal of the American Statistical Association*, 55(289):148–150, 1960.

[30] T. Sawchik. *Big data baseball.* Flatiron Books, New York, NY, 2016.

[31] 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.

[32] S. Sheather. Density estimation. *Statistical Science*, 19(4):588–597, 2004.

[33] R. Watts and A.T. Bahill. *Keep your eye on the ball.* W.H. Freeman and Company, 1990.

[34] R. Watts and R. Ferrer. The lateral force on a spinning sphere: aerodynamics of a curveball. *Am. J. Phys.*, 55(1):40–44, 1987.

[35] M. Woo. (Dec. 21, 2018). Artificial intelligence in NBA basketball [Online]. Available: www.insidescience.org/news/artificial-intelligence-nba-basketball.