

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

GLENN HEALEY<sup>ID</sup>, (Fellow, IEEE), AND LEQUAN WANG

Department of Electrical Engineering and Computer Science, University of California at Irvine, Irvine, CA 92617, USA

Corresponding author: Glenn Healey (ghealey@uci.edu)

**ABSTRACT** We develop a new method for characterizing the lift force on a baseball. The methodology addresses this task from the novel perspective of considering a large set of radar measurements acquired outside of a laboratory setting. The reduced degree of standardization in the measurements is countered by several elements of the approach. A new optimization method is developed that incorporates domain knowledge and constraints derived from optical measurements. The optimization accounts for the uncertainty in the 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 laboratory experiments. Fine-grained weather data is associated with each radar measurement to enable compensation for the local air density. By applying this methodology to a set of over two million trajectory measurements, we achieve unprecedented accuracy in the characterization of the lift force. We show that the lift coefficient is more than six percent greater than measured by previous laboratory experiments. We also demonstrate the ability to predict increases in the lift coefficient in response to changes in seam height on the order of a thousandth of an inch. Previous methods based on smaller sets of laboratory measurements have been unable to discern changes in the lift coefficient in response to changes in seam height of 0.02 inches. We demonstrate the statistical significance of the results. This work benefits several important application areas including the monitoring of sensor calibration systems and the definition of ball specifications that constrain trajectories to acceptable ranges.

**INDEX TERMS** Radar, air density, optical, sensor data, lift force, spin vector, surface roughness, optimization.

## I. 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], [2]. Of particular interest is the flight of a pitch which is a complicated function of the forces on the ball after it leaves a pitcher's hand. The force that the pitcher influences the most, the lift force, determines how much a pitch trajectory will change due to spin. A typical pitch is airborne for about 400 milliseconds and the batter must predict its path and start his swing within the first 200 milliseconds [3]. Small errors in prediction impair the batter's ability to make contact

and, as a result, pitchers benefit from using spin to alter pitch trajectories [4].

The lift force acts perpendicular to the ball's direction of motion and is caused by a pressure differential between sides of the spinning ball due to its interaction with surrounding air molecules. This force is also known as the Magnus force [5] and, contrary to the usual definition of the word lift, does not necessarily act in the vertical direction. The magnitude of the lift force depends on the air density, the ball speed and cross-sectional area, and the lift coefficient. The lift coefficient itself depends on physical quantities that the pitcher controls including the velocity and spin vectors as well as properties of the ball such as the surface roughness.

An accurate characterization of the lift force is important in a variety of contexts. The ability to describe the dependence of the lift force on quantities that the pitcher controls can be

The associate editor coordinating the review of this manuscript and approving it for publication was Brian Ng<sup>ID</sup>.

used to streamline the pitcher development process [6], [7]. Since aerodynamic properties of the ball have a significant impact on player valuation and the competitive balance of the sport [8], models that relate the lift force to characteristics of the ball can be used to define specifications that ensure that ball trajectories remain stable over time. This issue is of such paramount importance that Major League Baseball (MLB) has commissioned scientific committees [9], [10] to study how the ball's properties affect the game on the field. In addition, models that relate the lift force to contextual factors such as altitude can be used to inform decisions about how to achieve success in different environments [11].

Key to quantifying the lift force is an accurate characterization of the dependence of the lift coefficient on the ball's velocity vector, spin vector, and surface roughness [12]. 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. Controlled experiments utilizing wind tunnels [13]–[15], light gates [16], [17], or high-speed video systems [18]–[22] have been used for these measurements. 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 [23] 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. This data presents the opportunity to characterize the lift force with improved accuracy due to the reduction in the variance of estimators associated with larger samples [24]. But this data also presents specific challenges. Most MLB games are played outdoors where the weather conditions are uncontrolled. The radar 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 controlled experiments.

In this work, we develop a new methodology that exploits large sets of TM sensor data acquired under uncontrolled conditions to characterize the lift force. The characterization considers all of the more than two million pitches that have been measured in Major League games since the introduction of the TM sensor in 2017. At the heart of the methodology is an optimization technique that utilizes the knowledge generated by previous laboratory experiments, accounts for the uncertainty in different data 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, the TM data is augmented with measurements from weather sensors near the time and location of each pitch. Systematic biases due to pitchers, pitch types, and site calibration differences are accounted for by partitioning the TM data into pitch groups. Outliers

are a common problem for sensor systems and we use a robust estimation process to mitigate their effects. The new optimization method also enhances the generality of the characterization by enabling sensor fusion of the TM data with optical measurements acquired for a wide range of pitch parameters. The new approach can be adopted for a range of tasks that utilize large sets of sensor data that are frequently becoming available.

The new approach allows the lift force to be modeled with unprecedented accuracy. We show that a model derived using the new methodology provides a significantly better fit to a large set of sensor data than a previous model [2] that was derived from several sets of optical measurements [18], [19], [22]. We also show that the new model can account for small changes in surface roughness due to variation in seam height. This effect is important in the quantification of pitches [4], but has not been detectable by previous models developed using sophisticated experimental setups [17], [20].

We demonstrate that an accurate characterization of the lift force can be used for several applications. We show that a measured upper bound on the value of the lift coefficient as a function of pitch parameters can be used to monitor sensor calibration systems. We also show that the new model can be used to derive pitch descriptors including the spin efficiency [25] and spin vector that are useful for pitcher evaluation and development [6], [7]. This work continues the recent trend of exploiting sensor data to develop improved models for the mechanics of sports [23], [26]–[28].

## II. METHODOLOGY

In this section we provide an overview of the process developed in this paper and summarized in Figure 1 for modeling the lift force on a baseball. Sec. III presents the physical model for the lift force and introduces the key parameters which include the dimensionless lift coefficient and spin parameter which are typically measured using controlled laboratory experiments. Sec. IV describes the Trackman radar which has been used to measure the trajectories of millions of pitches during games and we show that these measurements can be combined with weather data to estimate the model parameters. In Sec. V we explain the challenges in using the radar data that include biases, outliers, and the lack of information about the geometric relationship between the spin and velocity vectors. We show that an important advantage of the new approach as compared to the use of laboratory measurements is that the availability of millions of pitch trajectory measurements allows the use of groupings and robust estimates to overcome these challenges. We also show that small sets of optical measurements can be used to further constrain the model. The physical properties and statistical uncertainty associated with each of the data sources are accounted for by an optimization method that generates the new model presented in Sec. V. We use the Akaike information criterion [29] to show that the new model is significantly more accurate than a previous model derived from small sets of laboratory measurements. We also show in

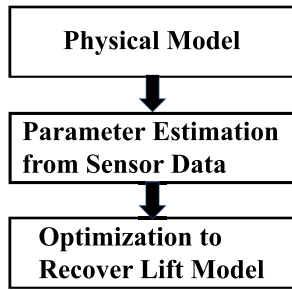


FIGURE 1. Summary of model generation methodology.

Sec. V that the new approach recovers a statistically significant relationship between the lift coefficient and seam height which could not be discerned by previous methods [17], [20] using sophisticated experimental setups.

### III. BASEBALL AERODYNAMICS

#### A. THE FLIGHT OF A BASEBALL

A baseball traveling through the air with a translational velocity vector  $\vec{v}$  is acted on by three forces as shown in Figure 2. Gravity pulls the ball down, drag acts opposite the velocity direction, and the lift force causes the ball to change direction due to spin. The lift force depends on the spin vector  $\vec{\omega}$  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 3. The magnitude of the lift force [22] is given by

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

where  $\rho$  is the air density,  $A$  is the ball cross-sectional area, and  $C_L$  is the dimensionless lift coefficient. Increases in  $C_L$  increase  $|\vec{F}_L|$  and cause larger spin-induced changes in pitch trajectory which typically lead to improved pitch quality [4]. If we define the velocity and spin vector directions by the unit vectors  $\hat{v} = \vec{v}/|\vec{v}|$  and  $\hat{\omega} = \vec{\omega}/|\vec{\omega}|$ , then the lift force is in the direction of  $\hat{\omega} \times \hat{v}$ .

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

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

where  $\vec{\omega}_{\parallel}$  is parallel to  $\hat{v}$  and  $\vec{\omega}_{\perp}$  is perpendicular to  $\hat{v}$ .  $\vec{\omega}_{\parallel}$  is known as the gyro component of the spin and does not contribute to the lift force [30]. The magnitude of  $\vec{\omega}_{\perp}$  is given by

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

The dimensionless spin parameter  $S$  [12] 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

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

where  $R$  is the ball radius.

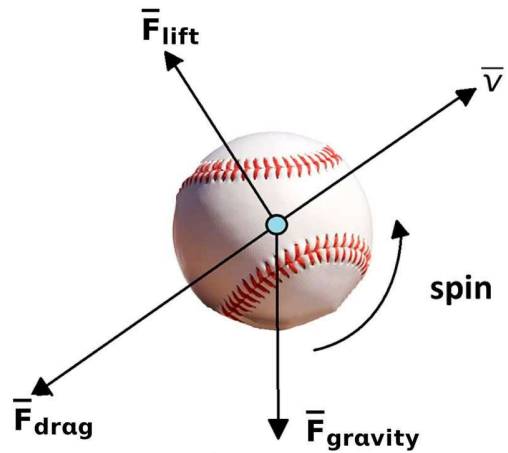


FIGURE 2. Forces on a spinning baseball in flight.

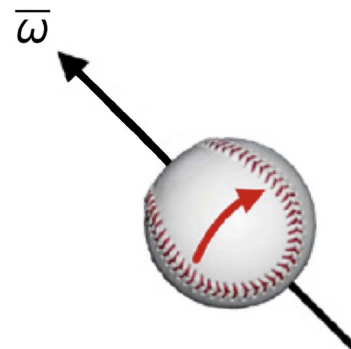


FIGURE 3. Spin vector  $\vec{\omega}$ .

#### B. THE RELATIONSHIP BETWEEN $C_L$ AND $S$

Watts and Bahill [31] speculated in 1990 that the lift coefficient  $C_L$  depends on the ratio  $|\hat{\omega} \times \hat{v}|$  of  $|\vec{\omega}_{\perp}|$  to  $|\vec{\omega}|$  and Jinji and Sakurai [19] later confirmed this using an experiment with measurements for 168 pitches using a set of synchronized video cameras. Nagami *et al.* [21] used a similar setup to make measurements for 75 pitches and to show experimentally that

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

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 Alaways and Hubbard [18] (17 pitches) and Nathan [22] (22 pitches) for the special case where  $|\hat{\omega} \times \hat{v}| = 1$ . The studies reported in [18] and [22] assigned uncertainties to the measurements using methods described in the articles.

A frequently used approximation to  $f(S)$  was presented in [2] that is based on a fit of experimental data from several sources including [18], [19], [22]. These data sets, however, include 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.

#### IV. PARAMETER ESTIMATION

##### A. TM 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  $\bar{a} = (a_x, a_y, a_z)$  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 [32]. The system also estimates the magnitude of the spin vector  $|\bar{\omega}|$  from the distribution of Doppler shifts.

##### B. 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  $|\bar{v}|$  which is not constant, we use the mean velocity magnitude  $|\bar{v}_\mu|$  over each pitch trajectory to construct the estimates. A similar approach has been used in previous studies [33]. The acceleration vector recovered by the TM system can be represented by

$$\bar{a} = \bar{a}_D + \bar{a}_L + \bar{a}_G \quad (6)$$

where  $\bar{a}_D$ ,  $\bar{a}_L$ , and  $\bar{a}_G$  are the accelerations corresponding to the drag, lift, and gravitational forces depicted in Figure 2. Since the drag force is parallel and opposite to the velocity direction and the lift force is perpendicular to the velocity direction, we can compute the magnitude of  $\bar{a}_D$  as the projection of  $\bar{a} - \bar{a}_G$  onto the velocity direction so that

$$\bar{a}_D = [(\bar{a} - \bar{a}_G) \cdot \hat{v}_\mu] \hat{v}_\mu \quad (7)$$

where  $\hat{v}_\mu = \bar{v}_\mu / |\bar{v}_\mu|$ . Therefore the lift acceleration is given by

$$\bar{a}_L = \bar{a} - \bar{a}_D - \bar{a}_G = \bar{a} + [(\bar{a} - \bar{a}_G) \cdot \hat{v}_\mu] \hat{v}_\mu - \bar{a}_G \quad (8)$$

and using equation (1) the lift coefficient for the pitch trajectory can be estimated by

$$C_L = \frac{2m|\bar{a}_L|}{\rho A |\bar{v}_\mu|^2} \quad (9)$$

where  $m$  is the mass of the ball and Newton's second law is used to relate the lift force  $\bar{F}_L$  and the lift acceleration  $\bar{a}_L$ . Equation (4) can be used to compute the spin parameter for the trajectory using

$$S = \frac{2\pi R |\bar{\omega}|}{|\bar{v}_\mu|} \quad (10)$$

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  $\rho$ .

##### C. ESTIMATING AIR DENSITY

The air density  $\rho$  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  $\rho$  associated with each pitch was computed in units of  $\text{kg/m}^3$  using the model from [34] given by

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

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

$$P = b * \exp [(-gME)/(RT + 273.15)] \quad (12)$$

where  $b$  is the barometric pressure in millimeters of mercury,  $g$  is the earth's gravitational acceleration in  $\text{m/sec}^2$ ,  $M$  is the molecular mass of air in  $\text{kg/mole}$ ,  $E$  is the elevation in meters, and  $R$  is the universal gas constant in  $\text{joules}/(\text{K mole})$ .  $V$  is the saturation vapor pressure in millimeters of mercury which is computed using the model in [35] given by

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

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

We examined the sensitivity of the air density estimate to small changes in location and time using the 2019 pitch data. For each pitch we considered the three closest weather stations  $W_1$ ,  $W_2$ , and  $W_3$  and the measurement for each station that was nearest to the time of the pitch. This yields three pairs of stations  $(W_1, W_2)$ ,  $(W_1, W_3)$ , and  $(W_2, W_3)$  for each pitch

with the associated absolute differences  $\Delta\text{Time}$ ,  $\Delta\text{Space}$ , and  $\Delta\rho$  for each pair. The average  $\Delta\text{Time}$  and  $\Delta\text{Space}$  differences over the pairs were 7.18 minutes and 5.53 km. Using the more than 2.032 million resulting  $(\Delta\text{Time}, \Delta\text{Space}, \Delta\rho)$  vectors we fit a model of the form

$$\Delta\rho = a \cdot \Delta\text{Time} + b \cdot \Delta\text{Space} \quad (14)$$

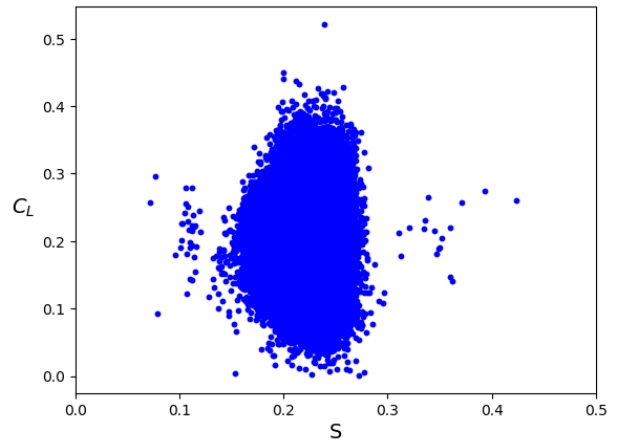
and found  $a = 3.27 \times 10^{-4}$  and  $b = 1.44 \times 10^{-3}$  using units of minutes, kilometers, and  $\text{kg/m}^3$  for  $\rho$ . For the average time (14.06 minutes) and space (1.93 km) differences associated with the estimate for each pitch, the model predicts a  $\Delta\rho$  of  $0.0078 \text{ kg/m}^3$  which is less than one percent of the average air density of  $1.149 \text{ kg/m}^3$  over the pitches. We will show in Sec. V-C that this uncertainty in air density has a negligible effect on the new model that we develop for the lift coefficient.

**V. OPTIMIZATION TO RECOVER  $f(S)$**

**A. SOURCES OF SCATTER IN  $(S, C_L)$  DATA**

The goal of this work is to find the function  $f(S)$  that defines the mapping  $C_L = f(S)|\hat{\omega} \times \hat{v}|$ . We showed in Sec. IV 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 Sec. III-B, however, the TM system does not allow direct measurement of  $|\hat{\omega} \times \hat{v}|$ . To alleviate this difficulty, we can consider using the domain knowledge that  $|\hat{\omega} \times \hat{v}|$  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 4 is a scatterplot of  $C_L$  versus  $S$  for pitches classified as four-seam fastballs in 2017. We see that there is significant variation in the values of  $C_L$  for a given value of  $S$  which prevents the direct use of these points for determining  $f(S)$ . Thus, we will develop a new methodology for finding  $f(S)$  that overcomes this variation.

There are several sources that contribute to the scatter in Figure 4. We examine a few of these sources in more detail in Figure 5 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 [36]. If we restrict the analysis to either the Target Field games (blue points) or the away games (red points), we see that the points are clustered around a central value with scatter that includes multiple outliers. Scatter is due to factors that include natural variation in pitches, variation in the physical properties of the baseball [9], [10], sensor noise, and pitch classification errors. We can further examine the scatter within the Away Games cluster in Figure 5 by partitioning the data by day. Figure 6 plots the average  $S$  value for points in this cluster for each of the twelve days that included at least twenty pitches. Figure 7 is the corresponding plot for the average  $C_L$  values. We see



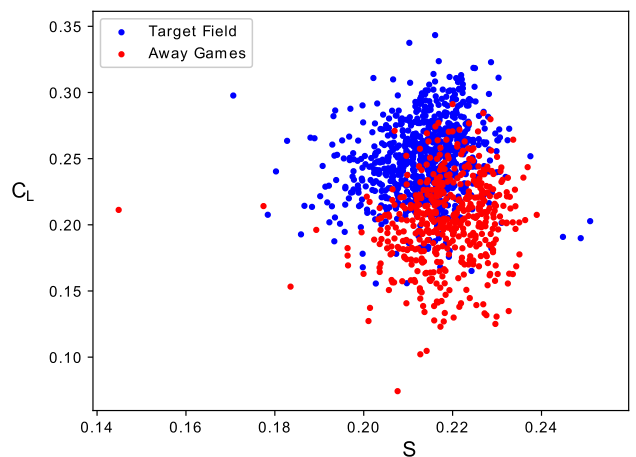
**FIGURE 4.  $C_L$  versus  $S$  for four-seam fastballs, 2017.**

that the day-to-day variations appear random and are smaller than the effects of park bias shown in Figure 5.

**B. 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



**FIGURE 5.  $C_L$  versus  $S$  for Ervin Santana, Four-Seam Fastball, 2017.**

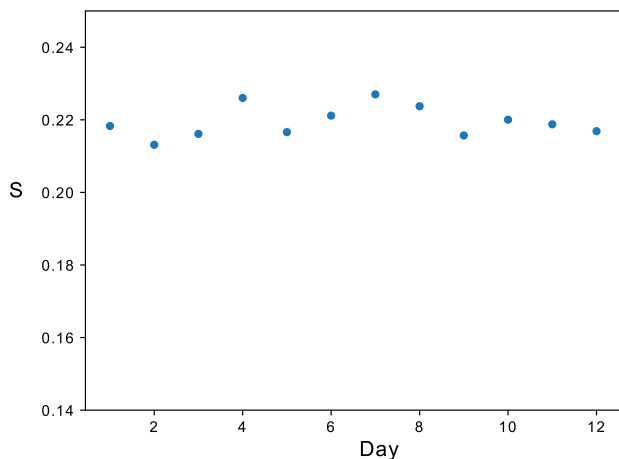


FIGURE 6. Average S by day for Ervin Santana, Four-Seam Fastball, away games, 2017.

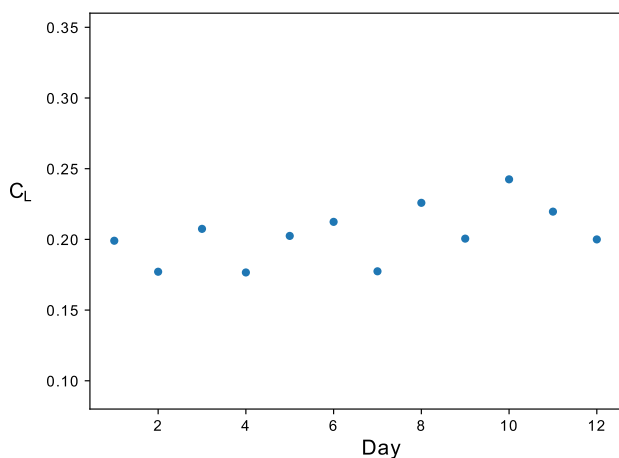


FIGURE 7. Average CL by day for Ervin Santana, Four-Seam Fastball, away games, 2017.

sample median [37]. Figure 8 demonstrates the action of the sample median as compared to the standard sample mean. This figure plots the histogram of the CL 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, CL) estimate for each pitch group is given by (S-hat, CL-hat) where S-hat is the sample median of the group S values and CL-hat is the sample median of the group CL values.

For a sample of size n derived from a distribution with probability density p(x) and median m, the uncertainty in the sample median m-hat can be approximated by the asymptotic variance of the estimator [38] which is given by

$$Var(\hat{m}) = \frac{1}{4np^2(m)} \tag{15}$$

For a computed m-hat, we can approximate the right-hand side of equation (15) by evaluating a kernel density estimate [39] for p(x) at the sample median m-hat. Figure 9 illustrates this process for the CL-hat estimate for two pitch groups. The first group is (Matt Boyd, Changeup, 2017) with a size of 244 pitches and a sample median of CL-hat1 = 0.191. The second group is (Marco Estrada, Four-seam fastball, 2017) with a size of 908 pitches and a sample median of CL-hat2 = 0.264. The kernel density estimates p-hat1(CL) and p-hat2(CL) for these two groups are plotted in Figure 9 and yield values of p-hat1(CL-hat1) = 5.793 and p-hat2(CL-hat2) = 17.975. Equation (15) then yields a standard deviation for CL-hat1 of 1/(2 \* sqrt(244 \* 5.793)) = 0.00553 and for CL-hat2 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. The average standard deviation for the CL-hat estimate over the 1678 pitch groups is 0.00201.

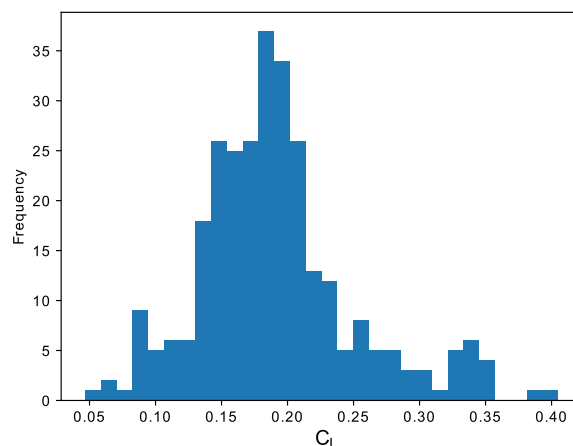


FIGURE 8. Distribution of CL for CC Sabathia, sinker, 2017, away games.

C. COMBINING RADAR AND OPTICAL MEASUREMENTS

Figure 10 is a scatterplot of the 1678 (S-hat, CL-hat) points generated by applying the method described in the previous section to TM radar data. We see that there is still significant variation in CL for a given S. Since CL depends on both S and |w-hat x v-hat|, this variation is due largely to differences in |w-hat x v-hat| for different pitch groups. Based on the results of multiple previous experiments [2], we can reasonably assume that there are (S-hat, CL-hat) points for which |w-hat x v-hat| = 1 which allows estimation of f(S) by finding a curve that is an upper bound to the points. To improve the accuracy of the estimate, we can also consider the use of the optical video measurements with assigned uncertainties [18], [22] that were described in Sec. III-B. In addition to having known values for |w-hat x v-hat|, the optical data also includes measurements over a wider range of S values.

Let (So(i), CLo(i)) for 1 ≤ i ≤ NO denote the set of (S, CL) points estimated using the optical video-based techniques and let σo(i) be the standard deviation of each CLo(i). Let (St(i), CLT(i)) for 1 ≤ i ≤ NT denote the set of

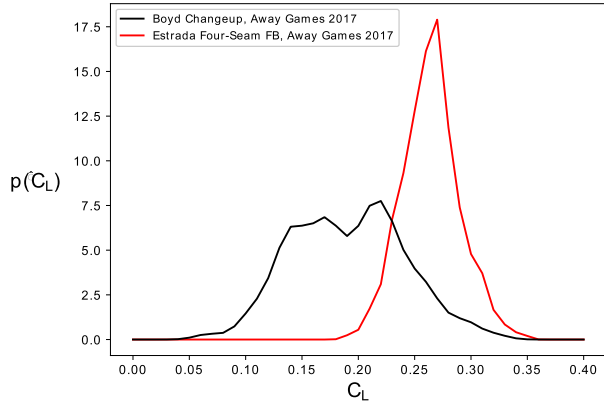


FIGURE 9. Kernel density estimates for  $C_L$ , Boyd and Estrada pitch groups, 2017.

$(\hat{S}, \hat{C}_L)$  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 (15) described in Sec. V-B. 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

$$E = \sum_{i=1}^{N_O} \left( \frac{E_O(i)}{\sigma_O(i)} \right) + \sum_{i=1}^{N_T} \left( \frac{E_T(i)}{\sigma_T(i)} \right) \quad (16)$$

where

$$E_O(i) = |C_{LO}(i) - f(S_O(i))| \quad (17)$$

and

$$E_T(i) = \begin{cases} C_{LT}(i) - f(S_T(i)) & \text{if } C_{LT}(i) \geq f(S_T(i)) \\ 0 & \text{if } C_{LT}(i) < f(S_T(i)) \end{cases} \quad (18)$$

Since the optical measurements were generated for the  $|\hat{\omega} \times \hat{v}| = 1$  configuration, the error  $E_O(i)$  for each point  $(S_O(i), C_{LO}(i))$  is considered in  $E$ . Since  $|\hat{\omega} \times \hat{v}|$  may be less than one for the radar measurements, the error  $E_T(i)$  in

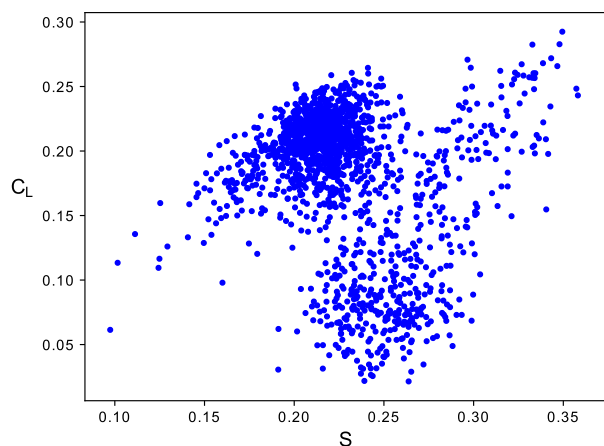


FIGURE 10.  $\hat{C}_L$  versus  $\hat{S}$  for 1678 TM pitch groups over 2017 to 2019.

equation (18) 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 [18] and Nathan [22]. After considering a range of increasing parametric functions, we found that a Hill function of the form

$$C_L(S) = \frac{AS^n}{a^n + S^n} \quad (19)$$

with parameters  $A = 0.370$ ,  $n = 1.651$ , and  $a = 0.137$  gave the best fit using the error measure in equation (16). The computation to generate the model can be performed by a search over the parameter space which requires less than a minute on a standard PC. We recomputed the fit after perturbing the air density estimate by  $\Delta\rho$  according to equation (14) using the  $\Delta$ Time and  $\Delta$ Space separation of the weather measurement from the pitch time and location and found similar parameters  $A = 0.370$ ,  $n = 1.658$ , and  $a = 0.138$ . The maximum difference between the two estimated  $C_L(S)$  functions was 0.0012 which corresponds to a difference of less than one percent so we conclude that uncertainty in the air density estimate has little effect on the recovered model.

Figure 11 plots the new model  $\hat{f}(S)$  along with the TM data and the optical video data. The Previous Model curve in the figure was presented in [2] as a representation for several sets of optical measurements [22] and is given by

$$C_L(S) = \frac{S}{2.32S + 0.4} \quad (20)$$

We see that there are significant differences between the models represented by equations (19) and (20) and the New Model curve is more than 6 percent greater than the Previous Model curve in the  $0.20 \leq S \leq 0.35$  range which is important for MLB pitches. In particular, the Previous Model which is based on a small number of optical samples is unable to account for a large number of TM pitch groups that are above the Previous Model curve in Figure 11.

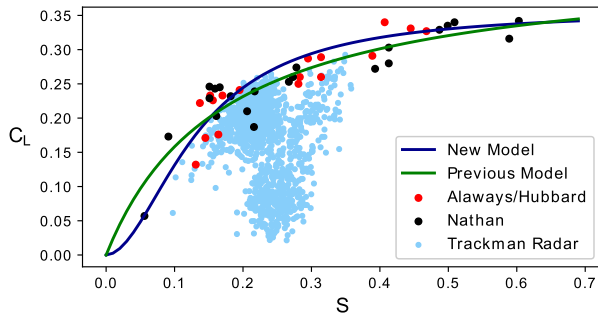
We can compare the new model  $M_1$  in equation (19) and the previous model  $M_0$  in equation (20) using the Akaike information criterion (AIC) [29] which is defined for a model  $M$  by

$$AIC(M) = 2q(M) - 2\ln(L(M)) \quad (21)$$

where  $q(M)$  is the number of estimated parameters in model  $M$  and  $\ln(L(M))$  is the log-likelihood [40] of the model. For the data analyzed in this section the log-likelihood is given by

$$\ln(L(M)) = -\frac{1}{2} \sum_{i=1}^{N_O} \left[ \ln \left( 2\pi\sigma_O^2(i) \right) + \left( \frac{E_O(i, M)}{\sigma_O(i)} \right)^2 \right] - \frac{1}{2} \sum_{i=1}^{N_T} \left[ \ln \left( 2\pi\sigma_T^2(i) \right) + \left( \frac{E_T(i, M)}{\sigma_T(i)} \right)^2 \right] \quad (22)$$

where  $E_o(i, M)$  and  $E_T(i, M)$  are the error values defined by equations (17) and (18) for model  $M$ . The difference  $\ln(L(M_1)) - \ln(L(M_0))$  for this data is 427.625 and the difference  $q(M_1) - q(M_0)$  is 1. Thus, the AIC difference is  $AIC(M_1) - AIC(M_0) = 2 - 2 * 427.625 = -853.25$ . Since  $AIC(M_1)$  is much smaller than  $AIC(M_0)$ , the new model  $M_1$  is preferred. If we use the AIC difference to compute the relative likelihood [41] we find that model  $M_1$  has a relative likelihood that exceeds 0.99 and we conclude that the new model  $M_1$  provides a significantly better fit to the data.



**FIGURE 11.** New model  $\hat{f}(S)$  with points recovered from TM and optical systems.

We can also ask whether the increased accuracy of the new model corresponds to differences in ball trajectories that are important for quantifying pitcher skill. Pitch movement [42] is defined as the displacement of a pitch over a distance of 40 feet due to the lift force. From the equations of motion [32], movement is proportional to the magnitude of the lift acceleration which by equation (9) is proportional to the lift coefficient  $C_L$ . The average movement of a four-seam fastball over the three years of our study was 10.8 inches and the six percent difference between the new model and previous model corresponds to a difference in movement of about 0.65 inches. Since the radius of a baseball is about 1.5 inches this difference in movement can significantly impact the ability of the batter to make solid contact with a pitch and, not surprisingly, an analysis over a large set of pitches [4] has shown that a change in movement of 0.65 inches has a meaningful impact on pitch value.

## D. SURFACE ROUGHNESS

### 1) PREVIOUS WORK

We have seen that the lift coefficient  $C_L$  has a strong dependence on the spin parameter  $S$ . Watts and Ferrer [15] 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 [16] reported in 2011 confirmed the Watts/Ferrer hypothesis by demonstrating that the high-seamed collegiate ball had a lift coefficient that was measurably greater than for the lower-seamed MLB ball.

More recently [20] accurate techniques have been developed to measure the seam height which has enabled

sophisticated experiments to be devised to analyze its relationship to  $C_L$ . In 2015 an experiment was reported that used balls projected from a custom machine utilizing a piston-based pneumatic cannon with wheels to impart spin [20]. Radar and high-speed video sensors were used to measure the flight path and spin rate of the projected baseballs in a controlled indoor environment. The flight time and carry distance were used to estimate  $C_L$  for each projected ball. A range of seam heights was considered that varied from 0.035 inches to 0.055 inches as measured by a custom surface profiler. A total of 52 trajectories were analyzed and the study concluded that there was no discernible difference in  $C_L$  over this range of seam heights.

Another experiment [17] was reported in 2018 that used a similar custom machine to project baseballs but used light gates to generate multiple speed and position measurements along a trajectory to estimate  $C_L$ . A high-speed video camera was used to verify the measurements. The experiment considered a low-seam ball with a seam height of 0.034 inches and a high-seam ball with a seam-height of 0.046 inches as measured by a custom surface profiler. The study analyzed fewer than 100 trajectories and concluded that  $C_L$  was not sensitive to these differences in seam height.

### 2) USING $\hat{f}(S)$ TO RELATE $C_L$ AND 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 [10]. As described in Sec. V-D1 measurements made with various custom experimental setups were unable to discern differences in  $C_L$  over significantly larger changes in seam height. We can ask whether the new method for modeling  $C_L$  that uses large sets of TM data partitioned into pitch groups is sensitive to these small changes in seam height.

The large majority of the  $(\hat{S}, \hat{C}_L)$  points in Figure 11 lie well below the  $\hat{f}(S)$  curve and contribute an error of zero to  $E_T(i)$  in equation (18). These points do not affect the recovered  $\hat{f}(S)$  function. The points that are most influential in determining the function are points that lie above or near  $\hat{f}(S)$ . Some of these points are shown in Figure 12 which plots the twenty  $(\hat{S}, \hat{C}_L)$  points in Figure 11 for each of the three years with the largest ratios  $\hat{C}_L/\hat{f}(\hat{S})$ . For a fixed  $S$ , we observe that the value of  $\hat{C}_L$  depends on the year with the largest  $\hat{C}_L$  values occurring for 2017. If we separately find a model of the form  $K\hat{f}(S)$  for each year that minimizes  $E$  in equation (16) 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  $K_{2019} = 0.968$ . This provides a separate estimated  $f(S)$  function for each of the three years using the measured trajectory and weather data. We can examine if these estimated functions for predicting  $C_L$  are sensitive to the small changes in average seam height from year to year as measured in the laboratory [10]. Figure 13 is a plot of  $K$  as determined using the methodology developed in this paper versus average seam height as measured independently in the



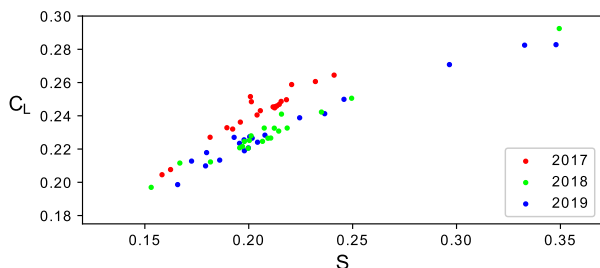
laboratory for each year. We see that  $K$  and seam height have a strong dependence as expressed by a sample correlation coefficient of  $r = 0.995$ . The statistical significance of this relationship can be assessed by testing the hypothesis  $r > 0$  [43]. This gives a  $t$ -statistic of 9.96 and a corresponding  $p$ -value of 0.03. Thus, the relationship is significant and, in contrast to methods [17], [20] described in the previous subsection, the new approach can be used to model increases in  $C_L$  due to small increases in seam height.

**VI. APPLICATIONS**

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

**A. MONITORING SENSOR CALIBRATION**

The function  $\hat{f}(S)$  represents the maximum value of the lift coefficient  $C_L$  for spin parameter  $S$  as described in Sec. V-C. Since  $C_L$  and  $S$  can be estimated from the TM data for each pitch using the method described in Sec. IV-B, we can use this function to monitor the accuracy of the TM calibration. For this purpose, we consider the distribution of the ratio  $C_L/\hat{f}(S)$  where  $C_L$  and  $S$  are estimated for each pitch in a game. As an example, Figure 14 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 some of the ratios exceeding one due to the sources of scatter described in Sec. V-A.

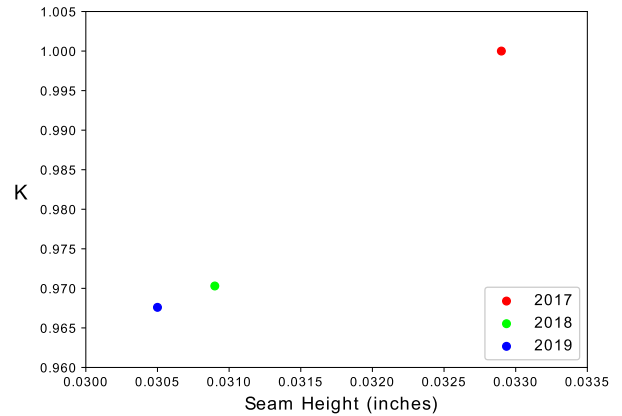


**FIGURE 12.**  $(\hat{S}, \hat{C}_L)$  points with the largest  $\hat{C}_L/\hat{f}(\hat{S})$  for each year.

By accounting for variation in  $C_L$  measurements we can determine typical ranges for the  $C_L/\hat{f}(S)$  ratio. For 2017 the average standard deviation of  $C_L$  within a (pitcher, pitch type) group for away games is  $\bar{\sigma}_C = 0.036$ . This standard deviation captures sources of variability that include natural variation in pitches and sensor noise. For a pitch with spin parameter  $S$  and the maximum  $|\hat{\omega} \times \hat{v}|$  of 1, the expected value of  $C_L$  is  $\hat{f}(S)$ . If the observed value of  $C_L$  is two standard deviations above this expected value, then the ratio  $C_L/\hat{f}(S)$  is given by

$$\frac{\hat{f}(S) + 2\bar{\sigma}_C}{\hat{f}(S)} \tag{23}$$

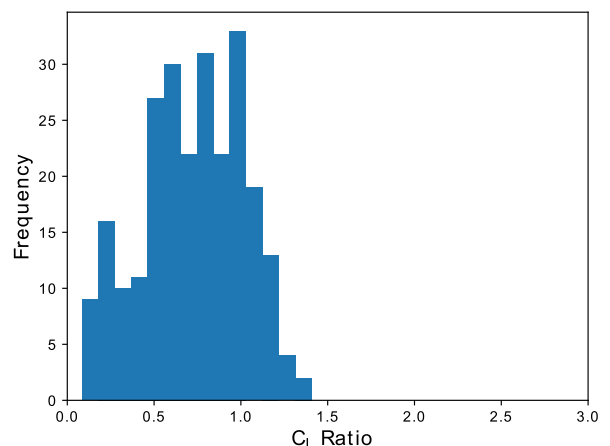
To approximate an upper bound for (23), we take a value for  $\hat{f}(S)$  of 0.15 which is smaller than the  $\hat{f}(S)$  for nearly



**FIGURE 13.** Lift coefficient dependence on average seam height.

all of the measured  $S$  values in Figure 11. This gives a ratio of  $(0.15 + 2*0.036)/0.15=1.48$ . Thus, we expect that a large majority of the ratios should fall below 1.5 even after allowing for a large  $|\hat{\omega} \times \hat{v}|$ , a small  $\hat{f}(S)$ , natural variation in pitches, and variation due to sensor noise. We see that this expectation is consistent with the distribution shown in Figure 14.

Figure 15 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 shifted to the right with a significant fraction of ratios exceeding two. Based on the preceding analysis, this distribution is extremely unlikely to result from variation in pitch parameters or sensor noise and strongly suggests an issue with the sensor calibration system. Thus, by estimating the likelihood of the  $C_L/\hat{f}(S)$  distribution we can identify potential calibration issues in real-time.



**FIGURE 14.** Distribution of  $C_L/\hat{f}(S)$  for SunTrust Park, 11 June 2017.

**B. RECOVERING THE SPIN EFFICIENCY AND SPIN VECTOR**

The function  $\hat{f}(S)$  is the link that enables the spin vector  $\hat{\omega}$  to be derived from TM data [33]. This vector and the associated spin efficiency [25], [44] are important tools for pitcher development and evaluation [6]. The spin efficiency

$|\bar{\omega}_\perp|/|\bar{\omega}|$  measures the proportion of the spin vector magnitude that is transferred to the lift force  $\bar{F}_L$ . The spin vector  $\bar{\omega}$ , more generally, determines the magnitude and direction of  $\bar{F}_L$  as detailed in Sec. III. The lift force causes a change in a pitcher's location in a predefined plane which is described by a movement vector [42]. The direction and magnitude of the movement vector have been shown to be key determinants of a pitch's effectiveness [4]. 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.

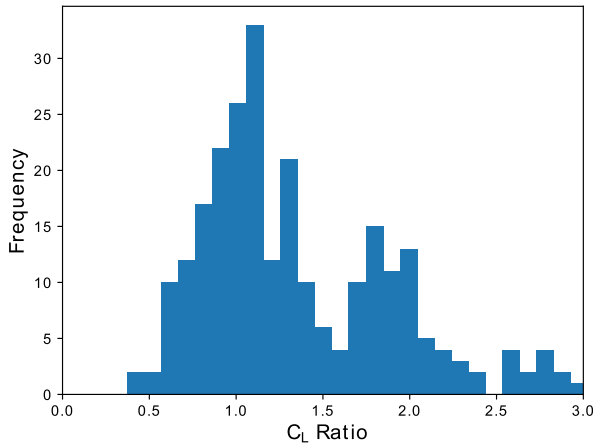


FIGURE 15. Distribution of  $C_L/\hat{f}(S)$  for SunTrust Park, 16 June 2017.

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

$$E = \frac{|\bar{\omega}_\perp|}{|\bar{\omega}|} \quad (24)$$

which simplifies to  $E = |\hat{\omega} \times \hat{v}|$  using equation (3). Since  $C_L$  and  $S$  can be estimated from TM data for a pitch as described in Sec. IV-B,  $E$  can be estimated using equation (5) by

$$\hat{E} = \frac{C_L}{\hat{f}(S)}. \quad (25)$$

In a similar way,  $\hat{f}(S)$  can be used to estimate the spin vector  $\bar{\omega}$ . As described in Sec. III-A, the  $\bar{\omega}_\parallel$  component of the spin vector  $\bar{\omega}$  is parallel to  $\hat{v}$  and can be expressed as

$$\bar{\omega}_\parallel = |\bar{\omega}_\parallel| \hat{v} \quad (26)$$

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

$$\bar{\omega}_\parallel = \sqrt{|\bar{\omega}|^2 - |\bar{\omega}_\perp|^2} \hat{v}. \quad (27)$$

Since the  $\bar{\omega}_\perp$  component of the spin vector is perpendicular to both the velocity vector  $\bar{v}$  and the lift acceleration  $\bar{a}_L$ , the direction of  $\bar{\omega}_\perp$  is specified by the unit vector  $\hat{v} \times \hat{a}_L$  where  $\hat{a}_L = \bar{a}_L/|\bar{a}_L|$ . Using equation (3) we can write

$$\bar{\omega}_\perp = |\bar{\omega}| |\hat{\omega} \times \hat{v}| (\hat{v} \times \hat{a}_L) \quad (28)$$

where the estimate  $\hat{E}$  from (25) can be used for  $|\hat{\omega} \times \hat{v}|$  and the remaining quantities in equations (27) and (28) are either directly measured by the TM radar or can be derived as described in Sec. IV. This allows an estimate of  $\bar{\omega}$  to be generated by combining the right-hand sides of equations (27) and (28) according to

$$\bar{\omega} = |\bar{\omega}| |\hat{\omega} \times \hat{v}| (\hat{v} \times \hat{a}_L) \pm \sqrt{|\bar{\omega}|^2 - |\bar{\omega}_\perp|^2} \hat{v} \quad (29)$$

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 Sec. V-A. This suggests the use of robust estimates over pitch groups and a careful consideration of uncertainty as described in Sec. V-B when using  $\hat{f}(S)$  to estimate  $E$  and  $\bar{\omega}$ .

## VII. CONCLUSION

The use of sensor systems to acquire data at sporting events has enabled a range of new applications [23], [26]–[28]. In this work, we have developed a new method for using sensor measurements to characterize the lift force on a baseball. The approach combines a large set of TM radar measurements made under uncontrolled game conditions with smaller sets of optical measurements made under controlled laboratory conditions. We have shown that the new approach provides a significantly more accurate characterization of the lift force than previous methods. In Sec. V-C we demonstrate that the new model is more consistent with a set of more than two million pitch measurements than an alternative model [2] that was derived using small sets of optical measurements [18], [19], [22] and we use the AIC difference to demonstrate the statistical significance of this result. In Sec. V-D we show that the new model captures dependence of the lift coefficient on small changes in surface roughness that could not be discerned by multiple previous efforts [17], [20] that employed elaborate experimental setups and we compute a  $t$ -statistic to demonstrate the statistical significance of this result. Each of the previous experiments considered the analysis of fewer than 100 trajectories. The results presented in this paper demonstrate that by applying the new methodology to a large set of in-game sensor measurements we can improve on the accuracy of models derived from smaller sets of measurements acquired under carefully controlled conditions.

The new model for the lift force can be used for a diverse set of applications. We showed in Sec. VI-A that constraints on the lift coefficient can be used to monitor sensor calibration systems. The relationship between the lift force and the velocity and spin vectors can be used to support pitcher development [6] and evaluation [45]. In particular, we showed in Sec. VI-B that the new characterization allows recovery of the spin efficiency and spin vector which are critical determinants of the effectiveness of pitches [25]. The new model also relates pitch trajectories to physical characteristics of the ball. This allows the definition of ball specifications that

constrain trajectories to a suitable range which is a topic of considerable interest [9], [10]. As an example, we showed in Sec. V-D that the new model can quantify changes in the lift coefficient in response to small changes in surface roughness. These changes are masked by the uncertainty in less precise models [17], [20]. The new model may also be helpful in quantifying other 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 [31], [46]. In summary, this work enhances understanding of the lift force and will improve the utility of the TM radar system which is used at many professional, college, and high school baseball facilities.

## ACKNOWLEDGMENT

The authors 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](http://baseballsavant.mlb.com) and weather data was obtained from [wunderground.com](http://wunderground.com).

## REFERENCES

- [1] R. Adair, *The Physics of Baseball*, 3rd ed. New York, NY, USA: Perennial, 2002.
- [2] R. Cross, *Physics of Baseball & Softball*. New York, NY, USA: Springer, 2011.
- [3] R. Gray, "Behavior of college baseball players in a virtual batting task," *J. Exp. Psychol., Hum. Perception Perform.*, vol. 28, no. 5, pp. 1131–1148, 2002.
- [4] G. Healey, "A Bayesian method for computing intrinsic pitch values using kernel density and nonparametric regression estimates," *J. Quant. Anal. Sports*, vol. 15, no. 1, pp. 59–74, Feb. 2019.
- [5] H. M. Barkla and L. J. Auchterlonie, "The Magnus or Robins effect on rotating spheres," *J. Fluid Mech.*, vol. 47, no. 3, pp. 437–447, Jun. 1971.
- [6] B. Lindbergh and T. Sawchik, *The MVP Machine: How Baseball's New Nonconformists are Using Data to Build Better Players*. New York, NY, USA: Basic Books, 2019.
- [7] T. Sawchik. (May 23, 2019). *MLB Curves and Sliders Have Gotten Alarmingly Nasty*. [Online]. Available: <https://www.fivethirtyeight.com/features/mlb-curves-and-sliders-have-gotten-alarmingly-nasty>
- [8] D. Cameron. (Nov. 7, 2017). *The Big Question Every Team Will Have to Answer this Winter*. [Online]. Available: <https://www.blogs.fangraphs.com/the-big-question-every-team-will-have-to-answer-this-winter>
- [9] 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*. [Online]. Available: [https://www.mlb.com/documents/7/9/4/278128794/Full\\_Report\\_of\\_the\\_Committee\\_Studying\\_Home\\_Run\\_Rates\\_in\\_Major\\_League\\_Baseball\\_052418.pdf](https://www.mlb.com/documents/7/9/4/278128794/Full_Report_of_the_Committee_Studying_Home_Run_Rates_in_Major_League_Baseball_052418.pdf)
- [10] J. Albert, A. Hosoi, A. Nathan, and L. Smith. (Dec. 10, 2019). *Preliminary Report of the Committee Studying Home Run Rates in MLB*. [Online]. Available: [https://www.mlb.com/documents/0/5/6/312149056/Report\\_of\\_the\\_Committee\\_Studying\\_Home\\_Run\\_Rates\\_in\\_MLB\\_121119.pdf](https://www.mlb.com/documents/0/5/6/312149056/Report_of_the_Committee_Studying_Home_Run_Rates_in_MLB_121119.pdf)
- [11] A. Nathan. *Baseball at High Altitude*. Accessed: Sep. 1, 2021. [Online]. Available: <https://www.baseball.physics.illinois.edu/Denver.html>
- [12] R. Mehta and J. Pallis, "Sports ball aerodynamics: Effects of velocity, spin, and surface roughness," in *Materials and Science in Sports*, F. Froes and S. Haake, Eds. Warrendale, PA, USA: TMS, 2001, pp. 185–197.
- [13] K. Aoki, Y. Kinoshita, J. Nagase, and Y. Nakayama, "Dependence of aerodynamic characteristics and flow pattern on surface structure of a baseball," *J. Vis.*, vol. 6, pp. 185–193, Jun. 2003.
- [14] L. J. Briggs, "Effect of spin and speed on the lateral deflection (curve) of a baseball, and the Magnus effect for smooth spheres," *Amer. J. Phys.*, vol. 27, pp. 589–596, Nov. 1959.
- [15] R. G. Watts and R. Ferrer, "The lateral force on a spinning sphere: Aerodynamics of a curveball," *Amer. J. Phys.*, vol. 55, pp. 40–44, Jan. 1987.
- [16] J. R. Kensrud and L. V. Smith, "In situ lift measurement of sports balls," *Proc. Eng.*, vol. 13, pp. 278–283, Jan. 2011.
- [17] J. R. Kensrud and L. V. Smith, "Drag and lift measurements of solid sports balls in still air," *Proc. Inst. Mech. Eng. P, J. Sports Eng. Technol.*, vol. 232, no. 3, pp. 255–263, 2018.
- [18] L. W. Alaways and M. Hubbard, "Experimental determination of baseball spin and lift," *J. Sports Sci.*, vol. 19, no. 5, pp. 349–358, Jan. 2001.
- [19] T. Jinji and S. Sakurai, "Direction of spin axis and spin rate of the pitched baseball," *Sports Biomech.*, vol. 5, no. 2, pp. 197–214, 2006.
- [20] J. R. Kensrud, L. V. Smith, A. Nathan, and D. Nevins, "Relating baseball seam height to carry distance," *Proc. Eng.*, vol. 112, pp. 406–411, 2015.
- [21] T. Nagami, T. Higuchi, H. Nakata, T. Yanai, and K. Kanosue, "Relation between lift force and ball spin for different baseball pitches," *J. Appl. Biomech.*, vol. 32, no. 2, pp. 196–204, Apr. 2016.
- [22] A. M. Nathan, "The effect of spin on the flight of a baseball," *Amer. J. Phys.*, vol. 76, no. 2, pp. 119–124, Feb. 2008.
- [23] G. Healey, "The new moneyball: How ballpark sensors are changing baseball," *Proc. IEEE*, vol. 105, no. 11, pp. 1999–2002, Nov. 2017.
- [24] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York, NY, USA: McGraw-Hill, 1991.
- [25] E. Sarris. (Aug. 18, 2020). *What is Spin Efficiency and why do the Rays Have it in Spades*. [Online]. Available: <https://www.theathletic.com/2003319/2020/08/18/sarris-what-it-spin-efficiency-and-why-do-the-rays-have-it-in-spades>
- [26] K. Clark. (Dec. 19, 2018). *The NFL's Analytics Revolution has Arrived*. [Online]. Available: <https://www.theringer.com/nfl/2018/12/19/18148153/nfl-analytics-revolution>
- [27] M. Woo. (Dec. 21, 2018). *Artificial Intelligence in NBA Basketball*. [Online]. Available: <https://www.insidescience.org/news/artificial-intelligence-nba-basketball>
- [28] S. Wang, Y. Xu, Y. Zheng, M. Zhu, H. Yao, and Z. Xiao, "Tracking a golf ball with high-speed stereo vision system," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2742–2754, Aug. 2019.
- [29] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*, E. Parzen, K. Tanabe, and G. Kitagawa, Eds. Cham, Switzerland: Springer, 1998, pp. 199–213.
- [30] A. Nathan and D. Baldwin, "An analysis of the gyroball," *Baseball Res. J.*, vol. 36, no. 1, pp. 77–80, 2007.
- [31] R. Watts and A. Bahill, *Keep Your Eye Ball*. San Francisco, CA, USA: Freeman, 1990.
- [32] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics*, 7th ed. Hoboken, NJ, USA: Wiley, 2005.
- [33] A. Nathan. (May 25, 2020). *Determining the 3D Spin Axis From TrackMan Data*. [Online]. Available: <https://www.baseball.physics.illinois.edu/trackman/SpinAxis.pdf>
- [34] A. Bahill, D. Baldwin, and J. Ramberg, "Effects of altitude and atmospheric conditions on the flight of a baseball," *Int. J. Sports Sci. Eng.*, vol. 3, no. 2, pp. 109–128, 2009.
- [35] A. L. Buck, "New equations for computing vapor pressure and enhancement factor," *J. Appl. Meteor.*, vol. 20, no. 12, pp. 1527–1532, Dec. 1981.
- [36] G. Schifman. (Mar. 27, 2018). *The Lurking Error in Statcast Pitch Data*. [Online]. Available: <https://www.fangraphs.com/tht/the-lurking-error-in-statcast-pitch-data>
- [37] P. Huber and E. Ronchetti, *Robust Statistics*, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.
- [38] P. R. Rider, "Variance of the median of small samples from several special populations," *J. Amer. Stat. Assoc.*, vol. 55, no. 289, pp. 148–150, Mar. 1960.
- [39] S. J. Sheather, "Density estimation," *Statist. Sci.*, vol. 19, no. 4, pp. 588–597, 2004.
- [40] A. Papoulis, *Probability & Statistics*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1990.
- [41] E.-J. Wagenmakers and S. Farrell, "AIC model selection using Akaike weights," *Psychonomic Bull. Rev.*, vol. 11, no. 1, pp. 192–196, Feb. 2004.
- [42] A. Nathan. (Oct. 21, 2012). *Determining Pitch Movement From PITCHfx Data*. [Online]. Available: <https://www.baseball.physics.illinois.edu/Movement.pdf>
- [43] C. Judd, G. McClelland, and C. Ryan, *Data Analysis: A Model Comparison Approach to Regression, ANOVA, and Beyond*, 3rd ed. New York, NY, USA: Routledge, 2017.

- [44] A. Nathan. (Mar. 31, 2015). *All Spin is Not Alike*. [Online]. Available: <https://www.baseballprospectus.com/article.php?articleid=25915>
- [45] T. Sawchik, *Big Data Baseball*. New York, NY, USA: Flatiron Books, 2016.
- [46] R. Cross, "Aerodynamics in the classroom and at the ball park," *Amer. J. Phys.*, vol. 80, no. 4, pp. 289–297, Apr. 2012.



**LEQUAN WANG** received the B.S.E. degree in electronic engineering from Northwestern Polytechnic University and the M.S. degree in electrical engineering from the University of California, Irvine, where he is currently pursuing the Ph.D. degree in electrical engineering and computer science.

...



**GLENN HEALEY** (Fellow, IEEE) received the B.S.E. degree in computer engineering from the University of Michigan and the M.S. degree in computer science, the M.S. degree in mathematics, and the Ph.D. degree in computer science from Stanford University. He is currently a Professor of electrical engineering and computer science with the University of California, Irvine. Before joining UC Irvine, he worked at IBM Research. He is the Director of the Computer Vision Laboratory,

UC Irvine, where his work focuses on using physical models to develop algorithms that extract information from large sets of data. He has been elected as a fellow of SPIE. He is a member of SABR. He has served on the editorial boards for *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, and the *Journal of the Optical Society of America A: Optics and Image Science, and Vision*.