Modeling and Projecting Offensive Value Using Combined Hit-Tracking and Speed Measurements

Glenn Healey
Electrical Engineering and Computer Science
University of California, Irvine, CA 92617
Email: ghealey@uci.edu
Research Paper for 2020 SABR Analytics Conference

1 Overview and Previous Work

The prediction of a player's future results on batted balls is often cited as one of the most challenging problems in baseball analytics [2]. Traditional outcome-based statistics for representing player skill on batted balls have been shown to have a low degree of repeatability [6] due to the effects of multiple confounding variables such as the defense, weather [1] [11], and ballpark. Sensor systems have created the opportunity to define batted-ball descriptors that are invariant to these variables [9]. MLB's Statcast system [10] measures several parameters of batted balls including the initial speed, vertical launch angle, and horizontal spray angle. The wOBA cube representation uses these measurements to compute intrinsic batted ball values and has been shown to provide more reliable estimates of batter performance than traditional outcome-based statistics [8]. This work also showed [7] that running speed is an important determinant of batter performance that is not captured by hit-tracking data. In this work, we build a model that combines Statcast batted ball and time-to-first physical measurements. The result is offensive statistics that provide a more accurate measure of performance and that support more accurate forecasts. This approach also promises to improve the accuracy of defensive metrics by allowing batter running speed to be included in quantifying the difficulty of a play.

2 Methodology

Radar and optical sensors collect seven terabytes of data during every Major League Baseball (MLB) game [10]. This data can be used to estimate parameters that describe the physical

properties of batted balls and player running speed. We might expect that statistics that are derived from these parameters will provide a more accurate measure of offensive skill than traditional statistics that are derived from outcomes. In order to test this hypothesis, we build a model for a batted ball's value as a function of contact parameters and batter running speed. The model uses a Bayesian framework that employs a kernel method to generate probability density estimates using a large set of sensor data. A cross-validation scheme allows the algorithm to adapt to the data by learning the optimal vector of smoothing bandwidths for each density. The result is a learning algorithm that generates a continuous mapping from batted-ball and running speed measurements to intrinsic values defined using a linear weights representation for run value. Separate mappings are built to accommodate the effects of batter handedness.

2.1 Sensor Data

Technologies such as Trackman's component of the Statcast system [10] use sensors to estimate the initial speed s and direction of batted balls in three dimensions. The direction is specified by two angles. The vertical launch angle v shown in figure 1 is the angle that the batted ball's initial velocity vector makes with the plane of the playing field where a vertical angle of $-90°$ is straight down and a vertical angle of $+90°$ is straight up. The horizontal spray angle h shown in figure 2 specifies the direction of the projection of the batted ball's initial velocity vector onto the plane of the playing field. The three rays in figure 2 intersect at home plate. The horizontal spray angles of $h = -45°, 0°$, and $45°$ define the directions toward third base, second base, and first base respectively. The configuration of the infielders and outfielders causes the expected outcome of a batted ball to have a strong dependence on the (s, v, h) vector. The Statcast system also measures the time from batted ball contact until the batter reaches first base. The success of a batter also depends on his running speed as measured by this time to first data.

Data acquired by Statcast is used for this study and includes measurements from every regular-season MLB game during 2018. The data set includes (s, v, h) data for batted balls and associated time to first running speed measurements. For each batter with at least 20 ground balls, we use the average of his three fastest times to first to represent the batter's

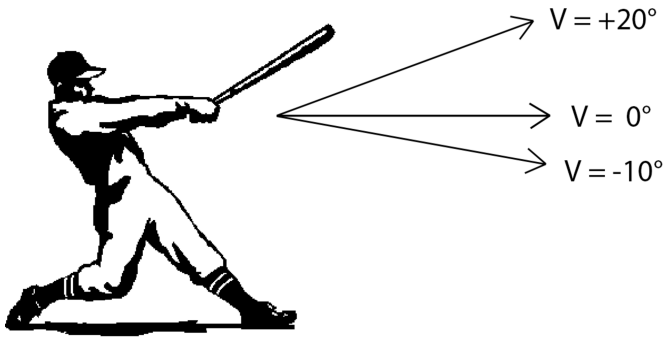

Figure 1: Vertical launch angle v

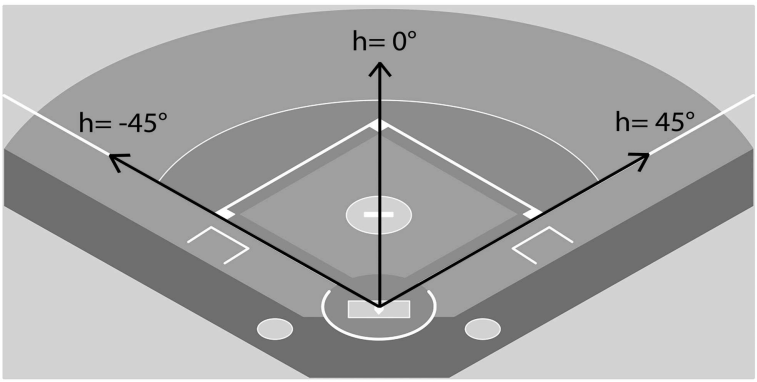

Figure 2: Horizontal spray angle h

time to first speed r. For switch-hitters a separate r value is computed for plate appearances from the right and left side.

2.2 Learning Algorithm

2.2.1 Bayesian Foundation

Given a set of observed batted balls and their outcomes, we develop a method for learning the dependence of a batted ball's value on a measured $d-$dimensional vector x that can include the (s, v, h) contact parameters and the r speed parameter. Using Bayes theorem, the posterior probability of an outcome R_j given the vector x is given by

$$P(R_j|x) = \frac{p(x|R_j)P(R_j)}{p(x)} \tag{1}$$

where $p(x|R_j)$ is the conditional probability density function for x given outcome R_j, $P(R_j)$

is the prior probability of outcome R_j, and $p(x)$ is the probability density function for x. We will show in section 2.2.5 that a weighted sum of the $P(R_j|x)$ values over outcomes provides a measure of the value of a batted ball.

2.2.2 Kernel Density Estimation

The goal of density estimation for our application is to recover the conditional probability density functions $p(x|R_j)$ and $p(x)$ in equation (1) from the x vectors and their outcomes. Given the typical positioning of fielders and the various ways that an outcome can occur, we expect a conditional density $p(x|R_j)$ to have a complicated multimodal structure. Thus, we use a nonparametric technique for density estimation.

Let x_i for $i = 1, 2, \ldots, n$ be a set of n observed x vectors. We first consider the task of estimating $p(x)$ from the vectors x_i. Kernel methods [14] which are also known as Parzen-Rosenblatt [12] [13] window methods are widely used for nonparametric density estimation. A kernel density estimate for $p(x)$ is given by

$$\widehat{p}(x) = \frac{1}{n} \sum_{i=1}^{n} K(x - x_i) \tag{2}$$

where $K(\cdot)$ is a kernel probability density function that is typically unimodal and centered at zero. A standard kernel for approximating a $d-$dimensional density is the zero-mean Gaussian

$$K(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} x^T \Sigma^{-1} x\right] \tag{3}$$

where Σ is the $d \times d$ covariance matrix. For this kernel, $\widehat{p}(x)$ at any x is the average of a sum of Gaussians centered at the sample points x_i and the covariance matrix Σ determines the amount and orientation of the smoothing. Σ is often chosen to be the product of a scalar and an identity matrix which results in equal smoothing in every direction. To recover a more accurate approximation $\widehat{p}(x)$ the covariance matrix should allow different amounts of smoothing in different directions. We enable this goal while also reducing the number of unknown parameters by adopting a diagonal model for Σ. This allows $K(x)$ to be written as a product of one-dimensional Gaussians which depends on the d unknown bandwidth parameters.

2.2.3 Bandwidth Selection

The accuracy of the kernel density estimate $\widehat{p}(x)$ is highly dependent on the choice of the bandwidth parameters [3]. The recovered $\widehat{p}(x)$ will be spiky for small values of the parameters and, in the limit, will tend to a sum of Dirac delta functions centered at the x_i data points as the bandwidths approach zero. Large bandwidths, on the other hand, can induce excessive smoothing which causes the loss of important structure in the estimate of $p(x)$. A number of bandwidth selection techniques have been proposed and a survey of methods and software is given in [5]. Many of these techniques are based on maximum likelihood estimates for $p(x)$ which select the bandwidth vector so that $\widehat{p}(x)$ maximizes the likelihood of the observed x_i data samples. Applying these techniques to the full set of observed data, however, yields a maximum with all bandwidth parameters set to zero which corresponds to the sum of delta functions result. To avoid this difficulty, maximum likelihood methods for bandwidth selection have been developed that are based on leave-one-out cross-validation [14].

The computational demands of leave-one-out cross-validation techniques are excessive for our data set. Therefore, we have adopted a cross-validation method which requires less computation. From the full set of n observed x_i vectors, we generate disjoint subsets S_j of fixed size n_v to be used for validation. For each validation set S_j, we construct the estimate $\widehat{p}(x)$ using the $n - n_v$ vectors that are not in S_j as a function of the bandwidth parameters. The optimal bandwidth vector σ_j^* for S_j is the choice that maximizes the pseudolikelihood [4] [5] according to

$$\sigma_j^* = \arg\max_{\sigma} \prod_{x_i \in S_j} \widehat{p}(x_i) \tag{4}$$

where the product is over the n_v vectors in the validation set S_j. The overall optimized bandwidth vector σ^* is obtained by averaging the σ_j^* vectors.

For this project, two validation sets S_1 and S_2 are used to select the optimized bandwidth vector σ^*. Let O be the number of batted balls in the data set that were hit in games starting on an odd day of the month and let E be the number of batted balls in the data set that were hit in games starting on an even day of the month. Set S_1 contains the first n_v batted

balls from games starting on an odd day and set S_2 contains the first n_v batted balls from games starting on an even day where n_v is the smaller of O and E. For each validation set, a d-dimensional search is conducted to find the optimized σ_j^* vector in equation (4).

2.2.4 Constructing the Estimate for $P(R_j|x)$

An estimate for $P(R_j|x)$ can be derived from estimates of the quantities on the right side of equation (1). The density estimate $\widehat{p}(x)$ for $p(x)$ is obtained using the kernel method defined by equations (2) and (3) with the optimized bandwidth vector σ^* learned using the process described in section 2.2.3. Each conditional probability density function $p(x|R_j)$ is estimated in the same way except that the training set is defined by the subset of the x_i vectors with outcome R_j. We use the σ^* derived for $p(x)$ for each case. This approach has the desirable effect of providing the same smoothing to a x_i vector in the numerator and denominator of (1) which prevents a probability $P(R_j|x)$ from exceeding one. Each prior probability $P(R_j)$ is estimated by the fraction of the n batted balls in the full training set with outcome R_j. The estimate for $P(R_j|x)$ is then constructed by combining the estimates for $p(x|R_j), P(R_j)$, and $p(x)$ according to Bayes theorem.

2.2.5 Intrinsic Value using wOBA

The posterior probabilities $P(R_j|x)$ can be combined into a measure of value. In 2007, Tango and his collaborators [15] defined weighted on base average (wOBA) as a weighted sum of the probability of outcomes where the weights are determined by the average run value of each outcome. The resulting formula for batted balls is

$$\text{wOBA}(x) = \sum_{j=0}^{5} w_j P(R_j|x) \tag{5}$$

where the w_j are the weights for the six outcomes R_0 = out, R_1 = single, R_2 = double, R_3 = triple, R_4 = home run, and R_5 = batter reaches on error (ROE). Thus, wOBA(x) is a measure of run value that depends on the measured x vector but does not depend on a batted ball's outcome. We will refer to wOBA(x) as an intrinsic value. The weights w_j in equation (5) can change from year-to-year. For the 2018 data we use the coefficients

$w_0 = 0.000, w_1 = 0.880, w_2 = 1.247, w_3 = 1.578, w_4 = 2.031$, and $w_5 = 0.920$ which were obtained from [16].

2.3 wOBA(x) Function

If x is the three-dimensional vector $x = (s, v, h)$ of batted ball parameters then the wOBA(x) function in equation (5) is known as the wOBA cube depicted in Figure 3. If x is the four-dimensional vector $x = (s, v, h, r)$ of batted ball and running speed parameters, then the wOBA(x) function in equation (5) is called the wOBA tesseract. A depiction of a tesseract is shown in Figure 4. We will provide examples of the wOBA cube in this section and will analyze the wOBA tesseract in detail in Section 3.

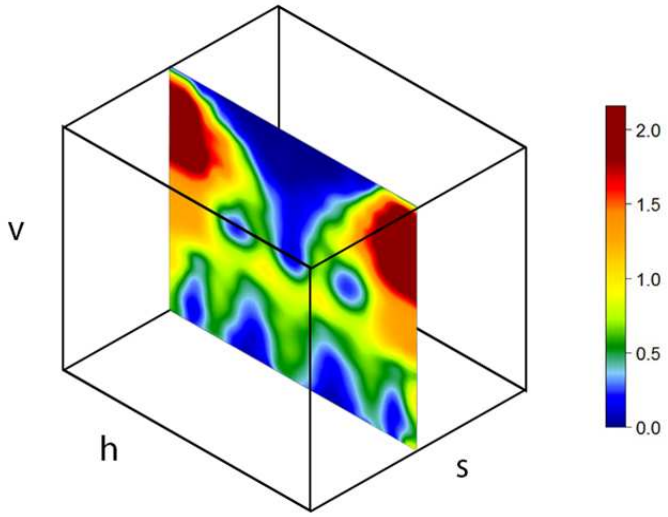

Figure 3: wOBA Cube

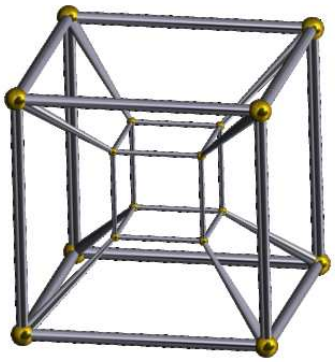

Figure 4: The Tesseract

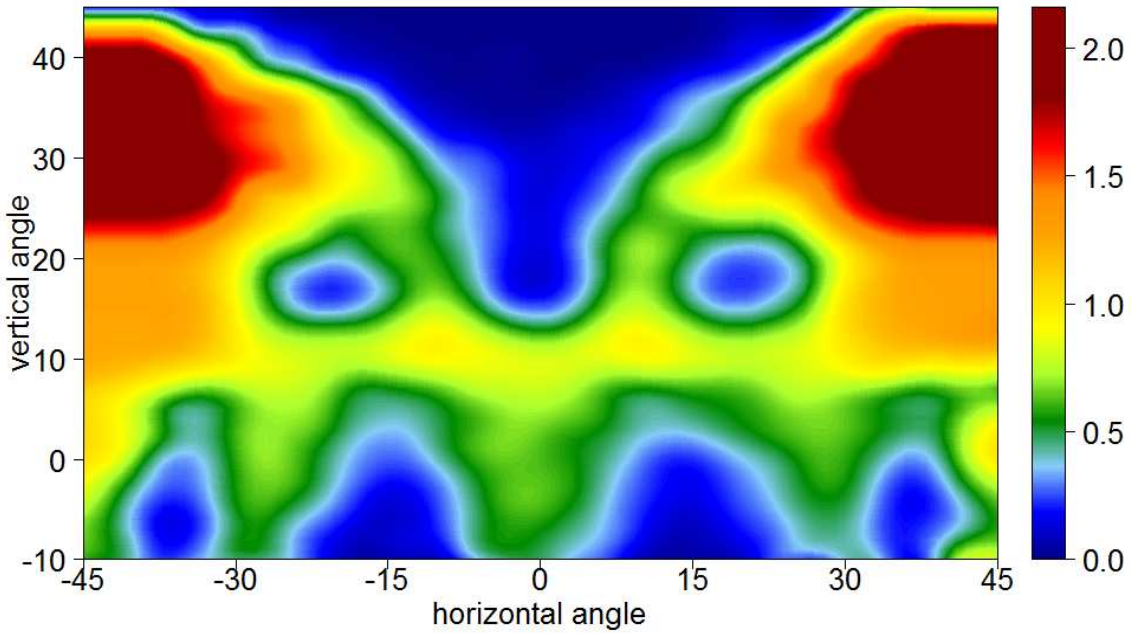

Figure 5: wOBA for an initial speed of 93 mph with angles in degrees

The wOBA cube defines the mapping from (s, v, h) to intrinsic value. As a specific example, Figure 5 displays wOBA(x) on the plane corresponding to a fixed initial speed s of ninety-three miles per hour. For this value of s, the best results for batters occur for balls hit with vertical angles between twenty-five and forty degrees with horizontal angles near the boundaries of fair territory $h \in [-45°, -35°]$ or $h \in [35°, 45°]$ where the field dimensions are typically the shortest. These batted balls often result in home runs. Batted balls hit at the same speed with the same vertical angle are less valuable at horizontal angles near zero degrees which correspond to larger field dimensions. For this initial speed, batted balls with vertical angles near twelve degrees tend to carry over the infielders and land in front of the outfielders and have a high value for all horizontal angles. Typical horizontal angle positions for the three outfielders are evident from the three cold zones for balls hit in the air to the outfield with $v \in [15°, 20°]$. Typical horizontal positions for the four infielders are evident from the four cold zones for balls hit on the ground ($v < 0$) for which infielders are often able to record an out.

Batter handedness affects the positioning of fielders which leads to significant wOBA(x)

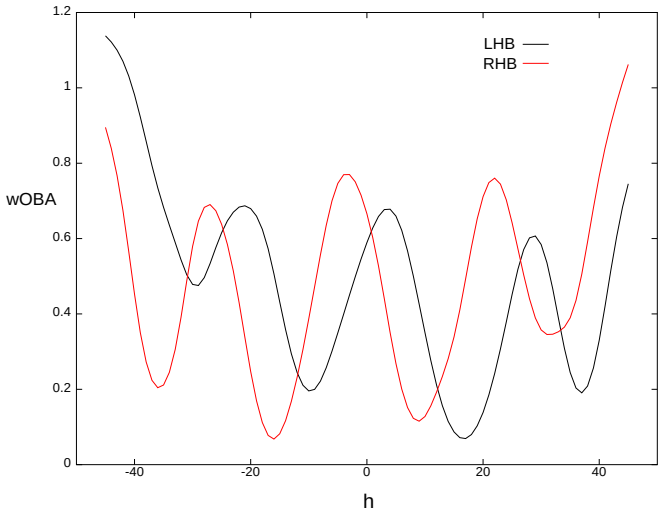

Figure 6: wOBA for speed 93 mph and vertical angle $-2°$

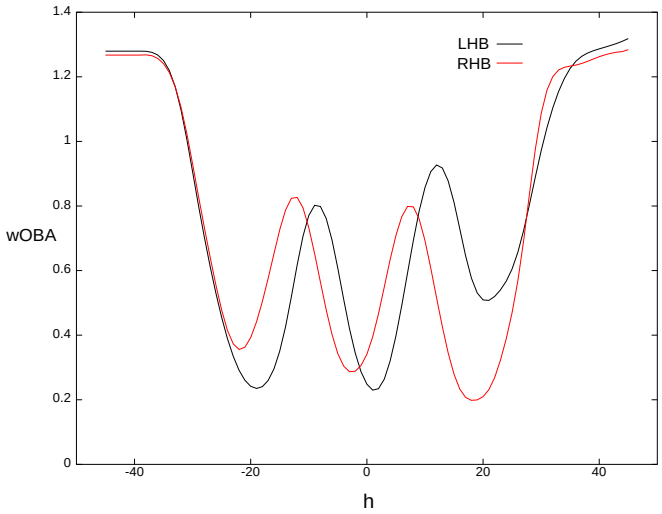

Figure 7: wOBA for speed 93 mph and vertical angle $+15°$

differences between left-handed and right-handed batters. Thus, we generate separate functions wOBA$l(x)$ for left-handed batters (LHB) and wOBA$r(x)$ for right-handed batters (RHB). Figures 6 and 7 illustrate differences between wOBA$l(x)$ and wOBA$r(x)$ for batted balls hit at 93 miles per hour for two vertical angles. Figure 6 considers balls hit on the ground with a vertical angle of $-2°$. We observe four minima in each curve that correspond to the typical position of the four infielders. We see, however, that the minima for

right-handed batters are shifted about seven degrees toward the third base line ($h = -45°$) compared to the corresponding minima for left-handed batters. This shift corresponds to the difference in fielder positioning as a function of batter handedness. Figure 7 examines the impact of batter handedness on balls hit in the air at 93 miles per hour with a vertical angle of $+15°$. The three minima in each curve correspond to the typical positions of the three outfielders. We see that the minima are shifted about three degrees toward the third base line ($h = -45°$) for right-handed batters. We also see that right-handed batters have an advantage for batted balls hit in the direction of the outfielder positioned near $h = -20°$ since this outfielder is typically positioned at a greater distance from home plate for right-handed batters which allows additional batted balls hit at this speed to land safely. We observe the opposite effect for batted balls hit in the direction of the outfielder positioned near $h = 20°$ since this outfielder is typically positioned at a greater distance from home plate for left-handed batters.

2.4 Intrinsic Batted-Ball Statistics

Using the learning algorithm developed in Section 2.2, a vector x can be assigned an intrinsic value given by either wOBA$l(x)$ or wOBA$r(x)$ depending on the handedness of the batter. A batted ball may also be assigned an outcome-based value given by the wOBA coefficient for its result. The outcome-based value depends on several factors that are beyond the control of the batter such as the fielders, the weather, the ballpark, and random luck. Analysts traditionally attempt to quantify the value of an offensive player's batted balls by using the average of his outcome-based values over a period of time. This average, O, is referred to as wOBA on contact or wOBAcon. Since outcome-based values depend on a number of contextual variables that are independent of the batter's quality of contact, the O statistic also depends on these variables. We use the average of a batter's intrinsic values as a more appropriate statistic for representing his offensive skill. We refer to the average of a batter's intrinsic values computed using the three-dimensional vector $x = (s, v, h)$ of batted ball parameters as I_3 and we refer to the average of a batter's intrinsic values using the four-dimensional vector $x = (s, v, h, r)$ that also includes his time to first estimate r as I_4.

3 Results

In previous work [7] we showed that many players who outperform their I_3 wOBAcon estimate tend to be faster runners while many players who underperform their I_3 tend to be slower runners. This motivates augmenting the wOBA cube with batter running speed to generate the wOBA tesseract.

3.1 Running Speed Measurements

The Statcast system generates multiple measurements of running speed. Statcast measures **sprint speed** which is derived from a runner's fastest one second window on individual plays and **time to first** which measures the time from batted ball contact to when the batter touches first base. For our application we use **time to first** which includes factors such as a batter's time to recover from the swing and start initial acceleration which affects his ability to beat out a hit.

Table 1: Fastest Time to First for LHB, 2018

LHB	r
Dee Gordon	3.807
Billy Hamilton	3.814
Roman Quinn	3.824
Magneuris Sierra	3.836
Cody Bellinger	3.879
J.B. Shuck	3.882
Brett Gardener	3.909
Mallex Smith	3.929

As described in Section 2.1, we define the running speed parameter r for batters with at least 20 ground balls as the average of the player's three fastest measured times to first. For switch-hitters a separate r value is computed for plate appearances as a right-handed and as a left-handed batter. For the 2018 season, the average r value over 207 qualifying left-handed batters was 4.245 seconds and the average r value over 319 qualifying right-handed batters was 4.305 seconds. Tables 1 and 2 present the left-handed and right-handed batters with the fastest r values for 2018. Figure 8 plots wOBA as a function of r for right-handed

Table 2: Fastest Time to First for RHB, 2018

RHB	r
Delino DeShields	3.855
Dansby Swanson	3.884
Trea Turner	3.896
Jose Altuve	3.896
Harrison Bader	3.899
Starling Marte	3.904
Scott Kingery	3.923
Adam Engel	3.929

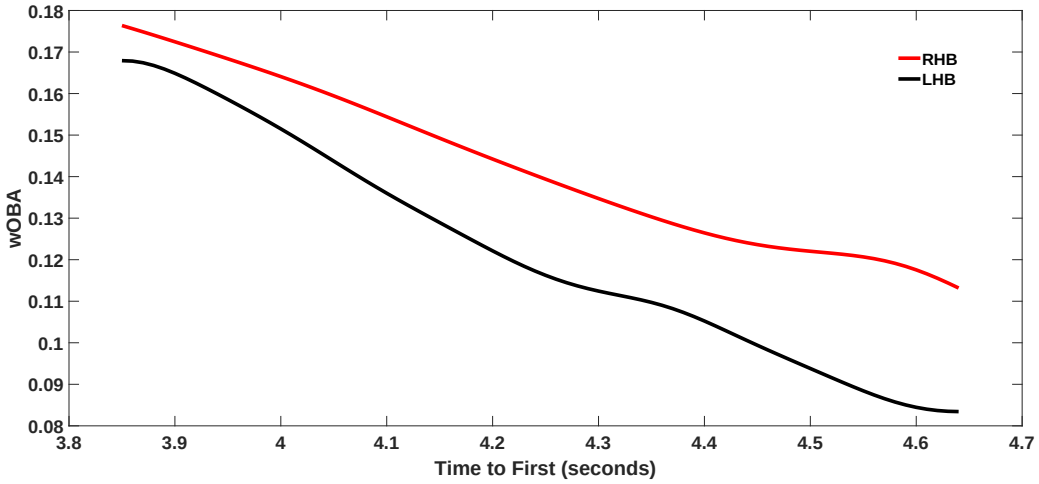

Figure 8: wOBA with r over all batted balls with $v < 10°$ in 2018

and left-handed batters for all batted balls with a vertical angle of less than 10 degrees in 2018. We see that there is a strong dependence of batted ball value on running speed as wOBA decreases as r increases. We also see that right-handed batters have a higher wOBA for a given r since a higher fraction of ground balls from RHB are hit to the left side of the infield which requires a longer throw to first base .

3.2 wOBA Tesseract

The wOBA tesseract defines the mapping from (s, v, h, r) to intrinsic value. A separate wOBA tesseract is generated for right-handed and left-handed batters using the process described in Section 2. Figures 9 and 10 provide examples of slices through the tesseract.

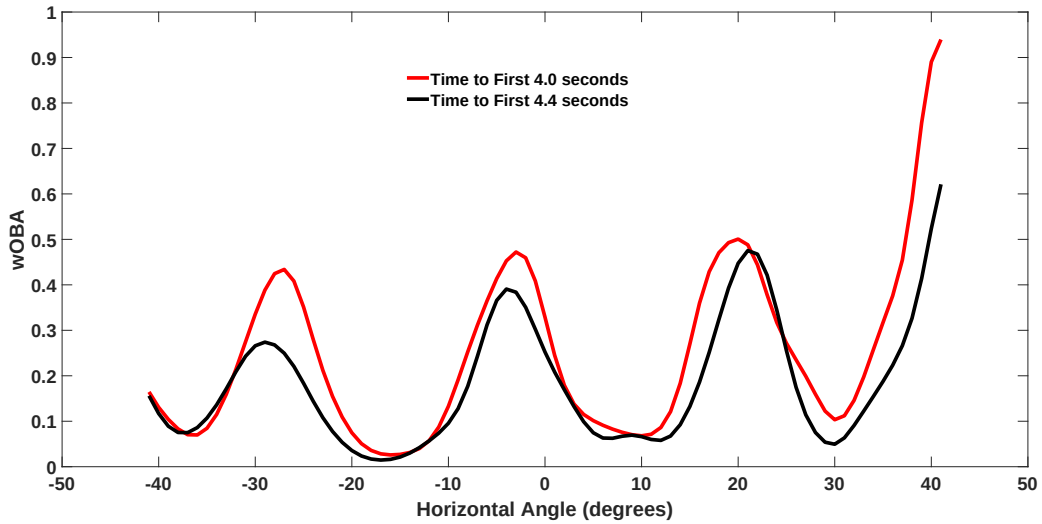

Figure 9: wOBA for RHB batted balls with $s = 87, v = -9°$ for two r values

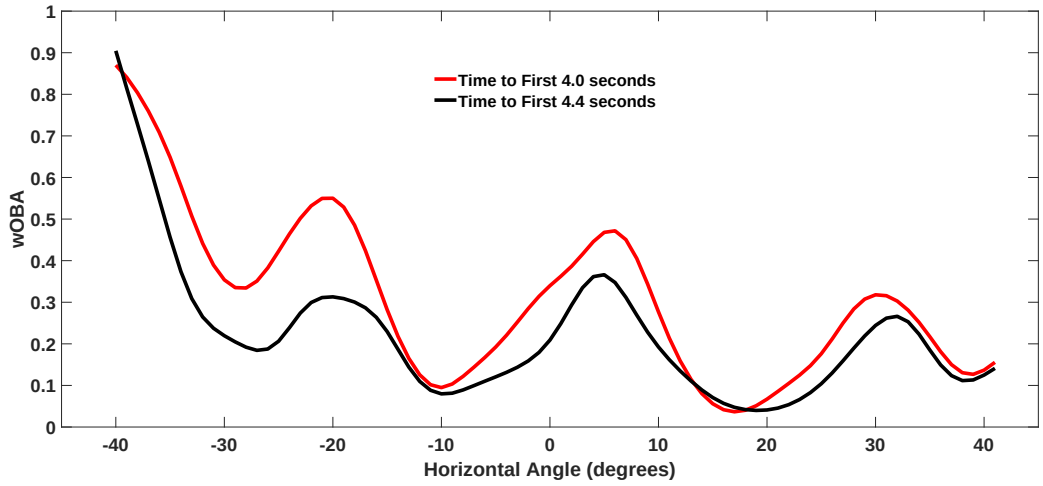

Figure 10: wOBA for LHB batted balls with $s = 97, v = -12°$ for two r values

Figure 9 plots wOBA(x) for right-handed batters for two different values of r as a function of the horizontal spray angle h with the initial batted ball speed and vertical launch angle fixed at $s = 87$ mph and $v = -9°$. The red curve corresponds to a faster than average time of $r = 4.0$ seconds and the black curve corresponds a slower than average time of $r = 4.4$ seconds. The four minima in the curves correspond to the typical position of the four infielders against right-handed batters. Near these minima we have a ground ball hit directly at an infielder and the wOBA values are similar for the different values of r. As we move away from the minima we see that a faster runner (red curve) tends to produce a

higher wOBA. We see that the largest wOBA values are observed for ground balls hit near the first base line as this horizontal angle is often undefended against right-handed batters and balls down the line may go for extra bases.

Figure 10 plots wOBA(x) for left-handed batters for two different values of r as a function of the horizontal spray angle h with the initial batted ball speed and vertical launch angle fixed at $s = 97$ mph and $v = -12°$. The red curve corresponds to a faster than average time of $r = 4.0$ seconds and the black curve corresponds a slower than average time of $r = 4.4$ seconds. The four minima in the curves correspond to the typical position of the four infielders against left-handed batters. We see that the minima are shifted to the right compared to the minima for right-handed batters shown in Figure 9. Near three of these minima the wOBA values are similar for the different values of r. For a ground ball hit directly at the third baseman near $h = -28°$, a faster runner enjoys an advantage since the third baseman will often be playing shallower to defend against a bunt for the faster runner and a 97 mph ground ball has a better chance of resulting in a hit. As we move away from the minima we see that a faster runner (red curve) tends to produce a higher wOBA. We see that the largest wOBA values are observed for ground balls hit near the third base line as this horizontal angle is often undefended against left-handed batters and balls down the line may go for extra bases.

3.3 Comparing I_3 and I_4

We computed the I_3 (wOBA cube) and I_4 (wOBA tesseract) estimates of wOBAcon for all batters in 2018 with at least 250 balls in play. Table 3 is a list of the I_3 leaders. These batters are known for their high quality of contact. Table 4 is a list of the I_4 leaders which factors running speed in addition to quality of contact into the value of each batted ball. We see that several of the slower runners (Gallo, Martinez, Judge, Goldschmidt) have a lower I_4 than I_3 while several of the faster runners (Trout, Story, Yelich, Betts) have a higher I_4 than I_3. The value of $I_4 - I_3$ depends on both the batter's running speed parameter r and his particular collection of batted balls.

Table 5 is a list of the batters with the highest $I_4 - I_3$ for 2018. These are the batters that would be expected to have the largest gain in wOBAcon due to their running speed given

Table 3: I_3 leaders for 2018

Batter	I_3
Joey Gallo	.597
Aaron Judge	.544
J.D. Martinez	.544
Mike Trout	.541
Paul Goldschmidt	.531
Matt Carpenter	.527
Giancarlo Stanton	.524
Christian Yelich	.522

Table 4: I_4 leaders for 2018

Batter	I_4	$I_4 - I_3$	r
Joey Gallo	.589	-.008	4.319
Mike Trout	.542	+.001	4.062
J.D. Martinez	.535	-.009	4.340
Aaron Judge	.534	-.010	4.487
Trevor Story	.529	+.015	3.955
Christian Yelich	.527	+.005	4.080
Mookie Betts	.526	+.007	4.055
Paul Goldschmidt	.522	-.009	4.309

Table 5: Highest $I_4 - I_3$ for 2018

Batter	$I_4 - I_3$	r
Cody Bellinger	.025	3.879
Ozzie Albies	.022	3.936/3.942
Niko Goodrum	.019	4.08/4.022
Rougned Odor	.018	3.984
Dansby Swanson	.018	3.884
Odubel Herrera	.017	3.969
Scott Kingery	.017	3.923
Brandon Nimmo	.017	4.113

Table 6: Lowest $I_4 - I_3$ for 2018

Batter	$I_4 - I_3$	r
Yasmani Grandal	-.035	4.663/4.966
Victor Martinez	-.034	4.634/4.965
Kendrys Morales	-.031	4.788/4.816
Justin Bour	-.029	4.498
Chris Davis	-.027	4.491
Albert Pujols	-.025	4.839
Yangervis Solarte	-.022	4.556/4.649
Joey Votto	-.022	4.575

their collection of batted balls. We see that all of these players have better than average values of the running speed parameter r. Note that for switch hitters two values (L/R) of r are used.

Table 6 is a list of the batters with the lowest $I_4 - I_3$ for 2018. These are the batters that would be expected to have the largest loss in wOBAcon due to their running speed parameter r given their collection of batted balls. We see that all of these players have worse than average values of r.

3.4 Variance Reduction

Differences between a batter's observed wOBAcon O and his I_3 are due to several factors including running speed, susceptibility to shifts, the ballpark, the weather, and random noise. By developing the I_4 statistic we improve the accuracy of the estimate by explicitly modeling the dependence of each batted ball on the running speed parameter r.

Table 7 is a list of the batters with at least 250 batted balls with the highest $O - I_3$. We see that each of these batters had a faster than average running speed r. In addition, several of these batters such as Carlos Gonzalez and Trevor Story in Colorado benefited from their home ballpark. We see that in each case the use of the wOBA tesseract to generate I_4 improved the accuracy of the model as $O - I_4$ is less than $O - I_3$.

Table 8 is a list of the batters with at least 250 batted balls with the lowest $O - I_3$. We see that each of these batters had a slower than average running speed r except Joe Panik who was slightly better than average. Several of these players (Morales, Moreland, Calhoun,

Martinez, Carpenter) were shifted on during a large fraction of their plate appearances. We see that in each case the use of the wOBA tesseract to generate I_4 improved the accuracy of the model as $|O - I_4|$ is less than $|O - I_3|$.

Table 7: Highest $O - I_3$ for 2018

Batter	$O - I_3$	$O - I_4$	r
Carlos Gonzalez	.063	.054	4.150
Ronald Acuna	.051	.039	3.945
Mallex Smith	.050	.039	3.929
Brandon Nimmo	.049	.033	4.113
Chris Taylor	.048	.039	4.017
Trevor Story	.045	.030	3.955
Eddie Rosario	.045	.029	3.969
Yoan Moncada	.045	.029	4.094/4.175

Table 8: Lowest $O - I_3$ for 2018

Batter	$O - I_3$	$O - I_4$	r
Kendrys Morales	-.064	-.033	4.788/4.816
Mitch Moreland	-.063	-.052	4.262
Kole Calhoun	-.058	-.045	4.315
Nelson Cruz	-.055	-.049	4.395
Albert Pujols	-.054	-.029	4.839
Victor Martinez	-.052	-.018	4.634/4.965
Matt Carpenter	-.048	-.037	4.281
Joe Panik	-.047	-.046	4.241

If we consider all of the players with at least 250 batted balls in 2018, the R-squared for the set of points (O, I_3) is 0.79 and the R-squared for the set of points (O, I_4) is 0.85. Therefore, the model that includes running speed using the r parameter has increased accuracy for representing a batter's wOBAcon. We therefore expect that I_4 is a better estimate of true talent wOBAcon and provides more value for projection [8].

4 Contributions to Baseball Analytics

This work makes several important contributions to baseball analytics. We have generalized the 3-D wOBA cube to the 4-D wOBA tesseract to include the impact of batter running speed. This enables the computation of offensive statistics that provide a more accurate assessment of talent level on batted balls and support more accurate projections. This approach also allows separation of the impact of batted ball skill and running speed on offensive value. An important advantage of this separation is that each skill can be regressed and projected using individual reliability and aging curves before conversion to projected offensive value during forecasting. The wOBA tesseract also has the potential to improve defensive metrics by quantifying the relationship between the batter's running speed and the difficulty of a play. The model also allows quantification of the loss of offensive value due to susceptibility to defensive shifts. The wOBA tesseract representation enables visualizations that provide insight into the mapping between batted-ball and running speed parameters and intrinsic value. The new method can also be used to monitor batters over time and to improve our understanding of how offensive value varies with age. The overall process of combining sensor data and machine learning techniques to generate new statistics can be readily adapted to support other areas of baseball analytics.

Acknowledgment

I thank Travis Petersen at MLB Advanced Media for providing Statcast data that was used in this study.

References

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

[2] B. Baumer and A. Zimbalist. *The Sabermetric Revolution: Assessing the growth of analytics in baseball*. University of Pennsylvania Press, Philadelphia, 2014.

[3] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2001.

[4] R. Duin. On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Transactions on Computers*, C-25(11):1175–1179, 1976.

[5] A.C. Guidoum. Kernel estimator and bandwidth selection for density and its derivatives. The kedd package, version 1.03, October 2015.

[6] G. Healey. (Aug. 2, 2016). The reliability of intrinsic batted ball statistics [Online]. Available: www.hardballtimes.com/the-reliability-of-intrinsic-batted-ball-statistics.

[7] G. Healey. (Mar. 17, 2016). The intrinsic value of a batted ball [Online]. Available: tht.fangraphs.com/the-intrinsic-value-of-a-batted-ball.

[8] G. Healey. Learning, visualizing, and assessing a model for the intrinsic value of a batted ball. *IEEE Access*, 5:13811–13822, 2017.

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

[10] J. Keri. (Mar. 4, 2014). Q&A: MLB Advanced Media's Bob Bowman discusses revolutionary new play-tracking system [Online]. Available: grantland.com/the-triangle/mlb-advanced-media-play-tracking-bob-bowman-interview.

[11] A. Nathan. Baseball at high altitude [Online]. Available: baseball.physics.illinois.edu/Denver.html.

[12] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[13] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3):832–837, 1956.

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

[15] T. Tango, M. Lichtman, and A. Dolphin. *The Book: Playing the Percentages in Baseball*. Potomac Books, Dulles, Virgina, 2007.

[16] wOBA and FIP constants [Online]. Available: www.fangraphs.com/guts.aspx?type=cn.