The Annals of Applied Statistics
2019, Vol. 13, No. 4, 2416–2451
https://doi.org/10.1214/19-AOAS1285

OUTLINE ANALYSES OF THE CALLED STRIKE ZONE IN MAJOR LEAGUE BASEBALL

BY DALE L. ZIMMERMAN, JUN TANG AND RUI HUANG

University of Iowa

We extend statistical shape analytic methods known as outline analysis for application to the strike zone, a central feature of the game of baseball. Although the strike zone is rigorously defined by Major League Baseball's official rules, umpires make mistakes in calling pitches as strikes (and balls) and may even adhere to a strike zone somewhat different than that prescribed by the rule book. Our methods yield inference on geometric attributes (centroid, dimensions, orientation and shape) of this "called strike zone" (CSZ) and on the effects that years, umpires, player attributes, game situation factors and their interactions have on those attributes. The methodology consists of first using kernel discriminant analysis to determine a noisy outline representing the CSZ corresponding to each factor combination, then fitting existing elliptic Fourier and new generalized superelliptic models for closed curves to that outline and finally analyzing the fitted model coefficients using standard methods of regression analysis, factorial analysis of variance and variance component estimation. We apply these methods to PITCHf/x data comprising more than three million called pitches from the 2008–2016 Major League Baseball seasons to address numerous questions about the CSZ. We find that all geometric attributes of the CSZ, except its size, became significantly more like those of the rule-book strike zone from 2008–2016 and that several player attribute/game situation factors had statistically and practically significant effects on many of them. We also establish that the variation in the horizontal center, width and area of an individual umpire's CSZ from pitch to pitch is smaller than their variation among CSZs from different umpires.

1. Introduction. The statistical analysis of shape has a relatively short history but a rich literature. Most of its early development occurred within systematic biology, where it was coined "morphometrics" and used primarily to infer phylogeny from shapes of anatomical objects (e.g., dinosaur bones or insect wings). Morphometrics continues to be an important part of biologists' statistical toolkit; book-length treatments of the subject include Bookstein (1997), Claude (2008), and Zelditch, Swiderski and Sheets (2012). From its original roots in systematic biology, statistical shape analysis has been extended to many other application areas, including biometric identification (Vacca (2007)), precision manufacturing (del Castillo and Colosimo (2011)), medical imaging (Kurtek et al. (2012)), computer vision, and other automatic object recognition technologies (Belongie, Ma-

Received January 2018; revised May 2019.

Key words and phrases. Elliptic Fourier model, kernel discriminant analysis, morphometrics, orthogonal distance fitting, shape analysis, superellipse.

lik and Puzicha (2002), Teutsch et al. (2013)); see Dryden and Mardia (1998) and Srivastava and Klassen (2016) for additional applications. Recently, it has even been applied to team sports (Jäger and Schöllhorn (2012)). In this article we consider an application of statistical shape analysis to an object that plays a central role in the sport of baseball, the strike zone. According to the Official Rules of Baseball,

> "The STRIKE ZONE is that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants, and the lower level is a line at the hollow beneath the kneecap. The Strike Zone shall be determined from the batter's stance as the batter is prepared to swing at a pitched ball."

We refer to the region so defined, or alternatively the rectangle comprising its side that lies over the front edge of home plate, as the rule-book strike zone (RBSZ). Baseball pitchers and batters use the RBSZ, together with a myriad of other information (strengths, weaknesses and past tendencies of each other, whether there are runners on base, current ball-strike count and on and on), to decide where to target the pitch and whether to swing at it. Furthermore, if the batter does not swing at the pitch, the home-plate umpire uses the RBSZ to immediately classify, or "call," the pitch as a ball or a strike in accordance with the following additional sentence from the Official Rules of Baseball:

> "A strike is a legal pitch when so called by the umpire, which (b) is not struck at, if any part of the ball passes through any part of the strike zone."

Ostensibly, umpires' calls are to be made in strict accordance with the RBSZ, but it is widely recognized that umpires make mistakes (sometimes calling a pitch a ball when it should have been called a strike and vice versa) and may even purposely adhere to a strike zone somewhat different than that defined by the rule book. We refer to this alternative strike zone as the called strike zone (henceforth CSZ) and define it loosely for now as a smooth-outlined region consisting of all those points for which a called pitch is more likely to be called a strike than a ball; a more precise definition will be given subsequently. Furthermore, we regard the CSZ as a "fluid" object that may vary across time, pitchers, batters, umpires, game situations and possibly other factors. To illustrate, Figure 1 displays CSZs for each of the years 2008–2016, obtained using methods to be described subsequently; an animation of this figure may be found in Section A of the Supplementary Material (Zimmerman, Tang and Huang (2019)). It appears that the CSZ gradually morphed over this time period, becoming taller and narrower, and that other geometric attributes (e.g., its centroid and area) also may have changed. But are the perceived changes real, and are they large enough to be practically relevant? Do other factors not considered in Figure 1, such as handedness of the batter and the ball-strike count, significantly affect the geometry of the CSZ? For those that do, which geometric attributes are affected and by how much? And finally, by how much does the geometry of the CSZ differ among umpires and from that prescribed by the rule book? In this article we develop shape-analytic methods to answer these questions.

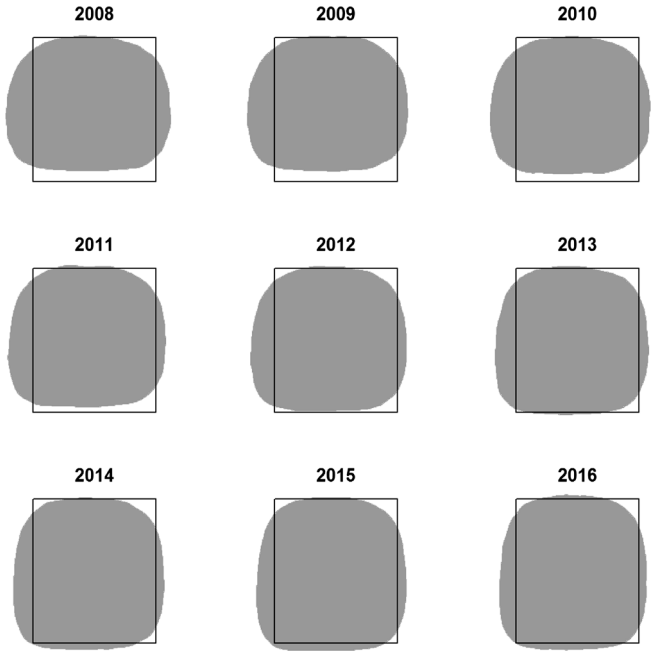

FIG. 1. *Called strike zones by year, superimposed upon the average RBSZ. Corners of the average RBSZ have coordinates* $(-0.829, 1.601)$, $(-0.829, 3.426)$, $(0.829, 1.601)$ *and* $(0.829, 3.426)$ *in units of feet in a coordinate system whose origin lies at the middle of the top front edge of home plate.*

While fans of baseball have pondered such questions for a long time, the opportunity to address them became much more realistic after 2008 when Major League Baseball began using the automated pitch tracking system known as PITCHf/x. The PITCHf/x system uses three cameras in every major league ballpark to record the speed and location of every pitched ball in every game from when it leaves the pitcher's hand until it crosses the vertical plane aligned with the front of home plate. The system also includes an on-site human operator, who monitors the calibration of the system and records the top and bottom of the RBSZ for each batter from the centerfield camera video, and a "stringer," another person who records other pertinent data such as the identity of the pitcher, batter and umpire, the game situation (e.g., inning, number of outs and ball-strike count) and the result of the pitch including, if it was not swung at, whether it was called a ball or a strike. The speed and location of the pitch are shown in near real-time on Major League Baseball's online Gameday webcast, as well as on many television broadcasts. Furthermore, the data are made freely available to the public on a website hosted by Major League Baseball Advanced Media (http://gd2.mlb.com/components/game/mlb/).

The recent availability of PITCHf/x data, combined with perennial interest in the strike zone among baseball fans and analysts, have prompted many studies of

the CSZ. Most nonacademic studies that have described the CSZ's geometry (e.g., Walsh (2010), Fast (2011a, 2011b), Roegele (2013a, 2013b, 2013c, 2014a, 2014b, 2015, 2016)) define it for a given factor combination as the union of cells of a rectangular grid, laid over the vertical plane containing the front edge of home plate and within which the the number of called strikes exceeds the number of called balls for that combination. This CSZ can be visualized easily, and its overall area computed by adding the areas of individual cells. However, it does not lend itself easily to performing formal inference (e.g., estimation of standard errors of the area estimate or formal comparison of the CSZ for one factor combination to the CSZ for another combination), and there is considerable arbitrariness in the origin, size and relative dimensions of the grid cells which could affect conclusions. Many academic studies model the probability that a pitch is called correctly as a function of various factors. Among the factors considered—some of which were found to be statistically significant—are year (Mills (2017)), batter handedness, race (Parsons et al. (2011)) and status (e.g., All-Star or not) of the batter and the pitcher (Kim and King (2014)), ball-strike count (Green and Daniels (2014), Marchi and Albert (2014), Walsh (2010)), type of pitch (fastball, curve, slider, etc.), league (National or American), inning and pitcher's home field advantage. Most such studies do not display CSZs or consider their geometry. Other studies model the probability that a called pitch is called a strike as a function of its location. The most successful of these latter studies, for example, Mills (2014, 2016a, 2016b), Tainsky, Mills and Winfree (2015), and Deshpande and Wyner (2017), are based on a semiparametric generalized additive model (GAM) of the log odds that the ith pitch is called a strike of form

$$\log\left(\frac{P(z_i=1)}{P(z_i=0)}\right) = f_k(x_i, y_i) + a_i'\beta + \epsilon_i,$$

where z_i equals 1 if pitch i is called a strike and equals 0 otherwise; $f_k(\cdot,\cdot)$ is a smooth function of the coordinates (x_i, y_i) of pitch i and k indexes the levels of one or more factors of interest (e.g., batter handedness); and a_i and β are column vectors containing, respectively, the values of (or dummy variables for) other factors of interest for pitch i and regression coefficients associated with those factors. The outline of the CSZ for a given factor combination may be obtained as the locus of points (x, y) where the fitted log odds for that combination equals 0 and geometric attributes (e.g., area) may be estimated from that outline.

In this article we apply and extend statistical shape-analytic methods, known collectively as *outline analysis*, to more directly model the geometry of the CSZ. Outline analysis is preferable to more well-known landmark-based shape-analytic methods when the objects under study are smoothly curvilinear and lack discrete homologous landmarks (such as joints of bones or corners of the eyes and mouth), as is the case here. Alternatively, the differential geometry-based planar shape analysis methodologies of Klassen et al. (2004) or Srivastava et al. (2005) could

be considered, but outline analysis is considerably simpler and more computationally efficient, is based on models whose parameters are easier to interpret and can deal with variation in location, size and orientation of the objects not only in their shapes.

Our outline analyses occur in three stages. First, the outline of a CSZ for each combination of a given set of factors is obtained by applying kernel discriminant analysis to the (x, y) locations of the called balls and strikes within that combination. This first-stage outline is regarded as a noisy object, subject to various sources of error including measurement error—PITCHf/x locations are claimed to be accurate only to within the nearest half-inch—and error due to the omission of important factors affecting the CSZ from those used to partition the set of called pitches. Second, a model is fit to the coordinates of points sampled along each first-stage outline. This produces estimates of model coefficients and a fitted outline that is considerably smoother than the first-stage outline. Last, the coefficients of the fitted second-stage outlines corresponding to years and all combinations of the other factors are analyzed using various classical statistical methods, depending on the nature of the question(s) being addressed. Specifically, we investigate effects of years, selected player attributes, game situation factors and umpires on the geometric attributes of the fitted outlines using standard multivariate and univariate regression, factorial analysis of variance and variance component estimation methods.

The models fit at the second stage are of two types: (1) truncated elliptic Fourier series expansions and (2) superellipses and generalizations thereof. The first class is well established within outline analysis, has closed-form least squares estimates and is very flexible in regard to the objects' shapes; its main disadvantage in this application is that only a few of its parameters correspond to readily discernible geometric features of CSZ outlines. The second class is introduced here for the first time. It has the advantage that all of its parameters correspond to discernible features of CSZ outlines; moreover, its relative parsimony turns out to be a virtue here because remarkably, apart from noise, CSZ outlines tend to be superelliptic in shape. However, superelliptic models must be fit by an iterative algorithm which we develop herein.

There are two main reasons why an outline analysis of CSZs may be a valuable supplement or alternative to an analysis based on models (such as GAMs) of the called strike probability. First, an outline analysis allows for more direct estimation and uncertainty quantification of important geometric attributes (e.g., centroid, height and width) of the CSZ and for the effects that years, umpires, various player traits and game situation factors have on those attributes. Second, because the RBSZ has a boundary, which itself is a closed curve belonging to the class of candidate outline models (provided that the candidate class is sufficiently large), an outline analysis facilitates inference comparing the CSZ to the RBSZ. Later, we will elaborate not only on these advantages but also on a disadvantage of outline analysis relative to analyses based on GAMs or other models of the called strike probability.

The remainder of the article is organized as follows. In Section 2 we describe the PITCHf/x data used to address our questions of interest, and we perform some basic exploratory analyses. Section 3 explains the kernel discriminant analysis approach used to determine CSZ outlines and broadly characterizes their geometry. Section 4 reviews and extends methodology for modeling the outlines via elliptic Fourier and (generalized) superelliptic models. Statistical analyses of the encoded outlines that address the main questions of interest are presented in Sections 5, 6 and 7. Section 8 contains a discussion of our findings and notes possible future work.

2. Data. Analyses presented in this article use PITCHf/x data from the nine-year period 2008–2016 (including postseasons). The data were scraped from Major League Baseball's Gameday website and manipulated using the `R` package `pitchRx` (Sievert (2014, 2015)). We used data corresponding to called pitches only. Among the available pitch characteristic variables, only the umpire's call of the pitch (ball or strike) and the pitch location were used, the latter being the (x, y) coordinates (from the perspective of the umpire behind home plate) of the baseball's center as it crossed the vertical plane containing the front edge of home plate. These coordinates are recorded in thousandths of a foot (or, equivalently, to the nearest 0.012 inch) on the Gameday website, but it is worth noting that they are measured with error. According to Sportsvision, the company that developed the PITCHf/x system, the location of a pitch determined by PITCHf/x is within one-half inch of its true location as it crosses the plate (Kagan (2009)). Among the available player attribute and game situation variables, we used only the handedness (right-handed or left-handed) of the pitcher and batter, the home-away status of the pitcher (henceforth called "venue"), the ball-strike count of the at-bat immediately prior to each called pitch (either zero, one, two or three balls and either zero, one or two strikes), and, for our last analysis only, the identity of the home plate umpire. These are the variables cited most often by previous studies as ones that affect the aforementioned probabilities. Additional variables could be considered, of course, but including them reduces the number of called pitches for some combinations of the factors too much to reliably determine first-stage outlines for those combinations.

The coordinates of each called pitch were compared to those defining the boundaries of the RBSZ to ascertain whether the pitch was truly a ball or a strike. The x-coordinates of the RBSZ's right and left boundaries may be computed trivially as $0 \pm (\frac{1}{2} \times \text{plate width} + \text{average ball radius}) = 0 \pm [\frac{1}{2}(17 \text{ in}) + 1.45 \text{ in}] \doteq \pm 0.829$ ft. Assigning y-coordinates to the lower and upper boundaries of the RBSZ is more problematic. The PITCHf/x data include y-coordinates of these boundaries as obtained by the PITCHf/x operator which ostensibly are the rule-book prescriptions as they apply to each batter. Unfortunately, however, these coordinates vary across plate appearances of the same batter by an amount much larger than could rea-

TABLE 1

Summary of the called pitches data by year. Each number in parentheses is the proportion for that category within the set of either all called pitches (for called balls, called strikes and overall misclassifications), all called balls (for called balls inside RBSZ) or all called strikes (for called strikes inside RBSZ)

Year	Called pitches	Called balls	Called strikes	Called balls inside RBSZ	Called strikes outside RBSZ	Overall misclassifications
2008	375,588	255,649 (0.6807)	119,939 (0.3193)	19,513 (0.0763)	30,630 (0.2554)	50,143 (0.1335)
2009	389,842	264,053 (0.6773)	125,789 (0.3227)	21,066 (0.0798)	30,443 (0.2420)	51,509 (0.1321)
2010	387,193	260,051 (0.6716)	127,142 (0.3284)	19,618 (0.0754)	30,040 (0.2363)	49,658 (0.1283)
2011	383,477	257,661 (0.6719)	125,816 (0.3281)	18,162 (0.0705)	29,517 (0.2346)	47,679 (0.1243)
2012	382,084	255,505 (0.6687)	126,579 (0.3313)	15,842 (0.0620)	29,173 (0.2305)	45,015 (0.1178)
2013	383,401	257,184 (0.6708)	126,217 (0.3292)	13,957 (0.0543)	28,197 (0.2234)	42,154 (0.1099)
2014	372,422	248,567 (0.6674)	123,855 (0.3326)	12,396 (0.0499)	28,299 (0.2285)	40,695 (0.1093)
2015	369,310	249,681 (0.6761)	119,629 (0.3239)	11,406 (0.0457)	27,495 (0.2298)	38,901 (0.1053)
2016	379,309	257,637 (0.6792)	121,672 (0.3208)	12,181 (0.0473)	26,131 (0.2148)	38,312 (0.1010)

sonably be expected by day-to-day variations in how much the batter crouches or how high they wear their pants (Fast (2011b)). Accordingly, to determine whether a pitch was called correctly for a given batter, we defined the lower and upper boundaries of a batter's RBSZ as the average value of the boundary's y-coordinates for all called pitches received by that batter from 2008–2016. However, for purposes of displaying a single RBSZ for all batters that can be compared visually to the various called strike zones that we obtain, we defined the "average RBSZ" as the rectangle with right and left boundaries as specified by the rule book and lower and upper boundaries computed as averages of the rule-book boundaries over *all* called pitches.

In aggregate, the data from the nine seasons consist of 3,426,165 called pitches corresponding to 1844 pitchers, 2511 batters and 122 home plate umpires. (The data are available for download from https://github.com/dalezimmerman/strikezone.) Of these pitches, 3539 were missing the batter's name. We included these 3539 pitches in our outline analyses but excluded them from the calculation of umpires' misclassification rates presented in this section because their upper and lower RBSZ boundaries could not be determined. Of the remaining 3,422,626 pitches, 2,305,988 (67.37%) were called balls and 1,116,638 (32.63%) were called strikes. Of the called balls, 144,549 (6.27%) lie inside the RBSZ, and of the called strikes, 261,452 (23.41%) lie outside the RBSZ. Overall, umpires misclassified 406,001 (11.86%) of called pitches. Table 1 breaks these statistics down by year and shows that umpire performance improved steadily over the nine-year period. Mills (2017) showed a similar improvement over a slightly shorter time frame (2008–2014) and argued that it was likely due to evaluations that Major League Baseball provided to umpires after every game, based on its own "in-house" analyses of the PITCHf/x data.

3. Outline determination. The CSZ differs fundamentally from tree leaves, hominid skulls and other objects to which methods of statistical shape analysis have previously been applied; it is not a tangible, physical object. Thus, one cannot simply take a photograph of a CSZ and digitize its outline. All first-stage CSZ outlines studied herein were obtained by applying a bivariate kernel discriminant analysis (KDA) procedure to subsets of the called pitches dataset, taking as inputs only the (x, y)-coordinates of the pitch. The subsets corresponded to cells in a multiway layout obtained by crossing the factors of interest (e.g., years, batter handedness and umpires) and marginalizing over the rest. The CSZ for a given subset was taken to be the set of points (x, y) where a locally smooth estimate of the probability density of called strikes exceeded a similar estimate of the probability density of called balls, following Green and Daniels (2014), Marchi and Albert (2014) and many others. Determining first-stage outlines of called strike zones in this way is not only easy to understand but also quite flexible, as it imposes no restrictions on the shape (apart from a degree of smoothness) or other geometric features of the objects produced. However, our experience indicates that in order to obtain reasonable, that is, undivided and not too "wiggly," outlines using well-accepted bandwidths, the subsets must contain several hundred called pitches, which limits the number of factors that can be considered simultaneously.

We implemented this approach using function `kda` in the `R` package `ks` with a bivariate spherical Gaussian kernel function and the unconstrained plug-in bandwidth selector of Wand and Jones (1994). This bandwidth selector is widely used, enjoys good optimality and practical properties for density estimation (Duong (2007), Wand and Jones (1994)) and has the same asymptotic order as a discriminant analysis-optimal selector (Hall and Kang (2005)).

Figure 1, introduced earlier, displays CSZs obtained in this manner, marginalized over all factors except year. The average RBSZ, defined in Section 2, is superimposed on each CSZ for comparison. It is evident that the center of each year's CSZ is shifted to the left of the RBSZ's center. Furthermore, each year's CSZ appears wider at its widest point than the RBSZ, though this has moderated recently. CSZs from all years differ from the RBSZ by having curved edges, with the upper margin often appearing to be slightly more rounded than the lower. The presence of curvature suggests that umpires call a pitch not merely by where the ball crosses the plate relative to prescribed rectangular boundaries but also by its distance from a central point. Though curved, the shapes of the outlines are not adequately described as elliptical, as some analysts (e.g., Carruth (2012), Roegele (2013a)) have described (and modeled) them. Rather, each outline more closely resembles a hybrid of an ellipse and a rectangle within a family of shapes known as *superellipses* (Gardiner (1965)). We defer further discussion of superellipses to Section 4.2, where we develop methodology to optimally fit them to CSZs.

For the data from years 2014–2016 only, we also obtained CSZs corresponding to the four combinations of pitcher and batter handedness (marginalized over all other factors) and to the 12 ball-strike counts (also marginalized over all other

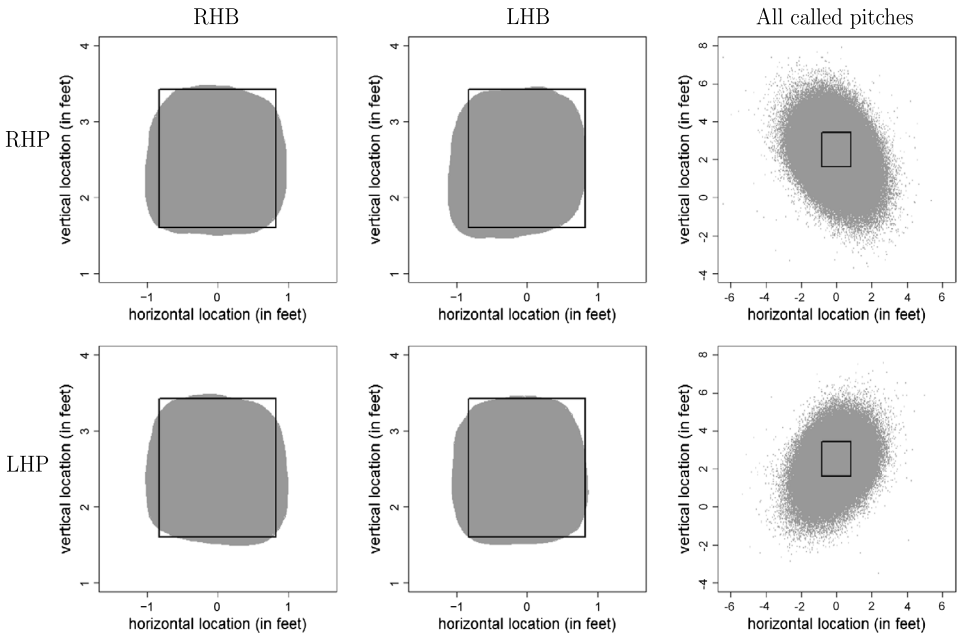

FIG. 2. *Called strike zones for the combined* 2014–2016 *data by pitcher and batter handedness* (*left two columns of panels*) *and all called pitches for the same time period by pitcher handedness* (*rightmost column of panels*), *superimposed upon the average RBSZ.*

factors). (We chose years 2014–2016 so that the results would be as current as possible, yet not specific to a single year.) CSZs corresponding to the handedness combinations (Figure 2) appear to be roughly similar in size and approximately centered vertically on the RBSZ but differ with respect to their horizontal center and shape. More specifically, while the CSZ for right-handed batters is approximately centered on the plate, that for left-handed batters is shifted discernibly to the left. In contrast, the handedness of the pitcher appears to have little effect on the centroid of the CSZ. As for shape, again superellipticity is evident, but with a more complicated form of asymmetry. Called strike zones for batter-pitcher combinations of the same handedness, like those for each year, have some latitudinal asymmetry but are approximately symmetric longitudinally. However, CSZs corresponding to batters and pitchers of opposite handedness are asymmetric longitudinally as well, being elongated so as to include more "down and away" pitches. The direction of elongation is not aligned with the axis of rotation of the pitcher's arm as he releases the pitch which is perhaps surprising in light of such an alignment when all called pitches by pitchers of the same handedness are considered (Figure 2, rightmost column of panels).

Figure 3 displays CSZs corresponding to the ball-strike counts. The display indicates that these CSZs are also superelliptic and vary considerably in size, growing larger as the ball count increases but smaller as the strike count increases.

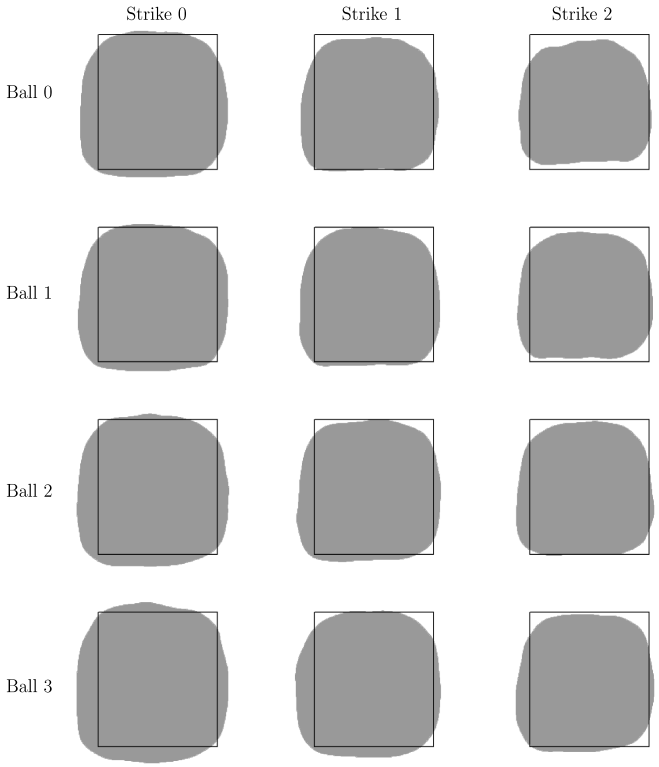

FIG. 3. *Called strike zones for the combined* 2014–2016 *data, by ball-strike count, superimposed upon the average RBSZ.*

However, no obvious systematic differences among their centroids or shapes are discernible.

Other nonlinear classifiers, for example, loess, nonparametric logistic regression and neural networks, could alternatively be used to obtain CSZs. Indeed, we obtained an alternative CSZ (not shown) for each year and factor combination using neural networks with various numbers of nodes. Those with four or more nodes turned out to be slightly smoother than our KDA-based outlines but were otherwise very similar, so we did not consider them any further. Likewise, we would expect outlines obtained using loess or nonparametric logistic regression to be very similar.

Once the KDA-based CSZ was determined, its outline was digitized using the `R` package `alphahull`. Coordinates were obtained for I points $\{(x_i, y_i) : i = 1, \ldots, I\}$ along each outline (I varies with the size and smoothness of the outline but typically lies between 300 and 600), starting arbitrarily with the uppermost of the two points where the outline intersects a vertical line passing through its centroid. Subsequently, we refer to these points as the *point configuration* and to the I-sided polygon whose vertices are the point configuration as the *discretized*

outline. Models of the two types described in the next section were then fit to each point configuration.

4. Outline modeling.

4.1. *Elliptic Fourier models*. An elliptic Fourier (EF) model (Kuhl and Giardina (1982)) expresses an outline's x and y coordinates as truncated Fourier series expansions in a curvilinear abscissa, t, that ranges from 0 to T where T is the perimeter of the outline. Specifically, the Nth-order EF model, or EF(N), for the outline $\{(x(t), y(t)) : 0 \le t < T\}$ is given by

$$x(t) = a_0 + \sum_{n=1}^{N}\left(a_n \cos\frac{2\pi nt}{T} + b_n \sin\frac{2\pi nt}{T}\right) + \epsilon_x(t), \tag{4.1}$$

$$y(t) = c_0 + \sum_{n=1}^{N}\left(c_n \cos\frac{2\pi nt}{T} + d_n \sin\frac{2\pi nt}{T}\right) + \epsilon_y(t), \tag{4.2}$$

where N is a positive integer less than or equal to $[I/2]$ (with $[\cdot]$ being the greatest integer function), and $\epsilon_x(t)$ and $\epsilon_y(t)$ are independent Gaussian white noise processes. The leading coefficients, a_0 and c_0, comprise the outline's centroid, and (a_n, b_n, c_n, d_n) is known as the nth *harmonic*. The locus of points (x, y) on the curve corresponding to the nth harmonic is an ellipse centered at the origin. Thus, the EF model describes the position of a point travelling (as t varies) around a series of N superimposed and successively smaller ellipses, as in a characterization of planetary orbits by Ptolemaic epicycles; for further details and a graphical illustration see Kuhl and Giardina (1982).

Least squares estimates $\hat{a}_0, \hat{c}_0, \{(\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{d}_n) : 1 \le n \le N\}$ of EF(N) model parameters [which minimize the integrals of squared x- and y-direction differences between the discretized outline and an outline of form (4.1)–(4.2)] exist in closed form as follows:

$$\hat{a}_0 = \frac{1}{T}\sum_{i=1}^{I} x_i \Delta t_i,$$

$$\hat{a}_n = \frac{T}{2\pi^2 n^2}\sum_{i=1}^{I}\frac{\Delta x_i}{\Delta t_i}\left(\cos\frac{2\pi nt_i}{T} - \cos\frac{2\pi nt_{i-1}}{T}\right),$$

$$\hat{b}_n = \frac{T}{2\pi^2 n^2}\sum_{i=1}^{I}\frac{\Delta x_i}{\Delta t_i}\left(\sin\frac{2\pi nt_i}{T} - \sin\frac{2\pi nt_{i-1}}{T}\right),$$

$$\hat{c}_0 = \frac{1}{T}\sum_{i=1}^{I} y_i \Delta t_i,$$

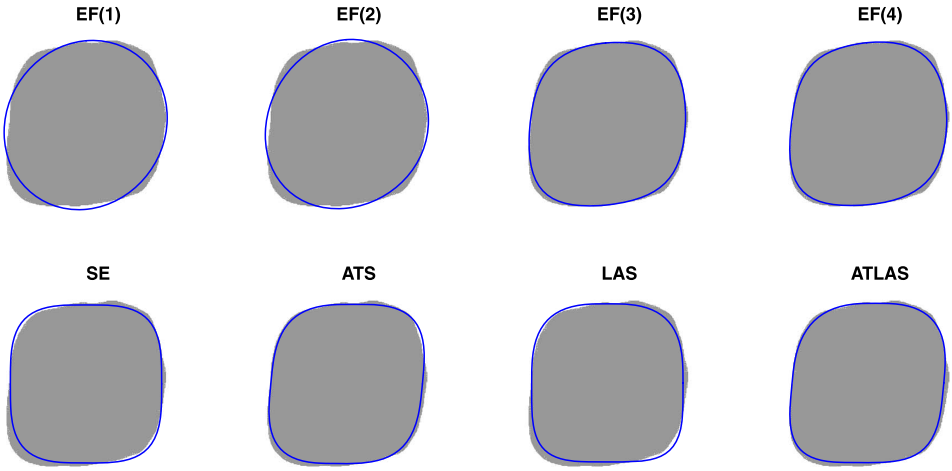

FIG. 4. *Fitted outline models* (*solid curve*) *to the CSZ for the RHP-LHB-away*-0-0 *factor combination for the combined* 2014–2016 *data*: *EF models of order* 1 *to* 4 (*top panel*), *and superelliptic models SE*, *ATS*, *LAS and ATLAS* (*bottom panel*).

$$\hat{c}_n = \frac{T}{2\pi^2 n^2} \sum_{i=1}^{I} \frac{\Delta y_i}{\Delta t_i} \left(\cos \frac{2\pi n t_i}{T} - \cos \frac{2\pi n t_{i-1}}{T} \right),$$

$$\hat{d}_n = \frac{T}{2\pi^2 n^2} \sum_{i=1}^{I} \frac{\Delta y_i}{\Delta t_i} \left(\sin \frac{2\pi n t_i}{T} - \sin \frac{2\pi n t_{i-1}}{T} \right).$$

Here, $\Delta x_1 = x_1 - x_I$ and $\Delta x_i = x_i - x_{i-1}$ for $i = 2, \ldots, I$ (with Δy_i defined similarly), Δt_i is the length of the linear segment between the $(i-1)$th and ith points, and t_i is the accumulated length of such segments at point i.

Although as many as $[I/2]$ harmonics may be estimated for any outline, the vast majority of the information about the outline typically is captured by just the first few. Figure 4 (top panel) displays a series of fits of EF models to one CSZ, using increasing numbers of harmonics, from which it appears that the original outline is reasonably well fit by a fourth-order model. A quantitative measure of information that morphometricians commonly use to choose the number of harmonics is the cumulative power P_N given by $P_N = \sum_{n=1}^{N} (\hat{a}_n^2 + \hat{b}_n^2 + \hat{c}_n^2 + \hat{d}_n^2) / \sum_{n=1}^{[I/2]} (\hat{a}_n^2 + \hat{b}_n^2 + \hat{c}_n^2 + \hat{d}_n^2)$, whose value approaches 1.0 as N increases. One can select the order of the EF model by choosing N as the value for which the cumulative power is, say, 0.999 or some other number deemed sufficiently close to 1.0. Alternatively, N may be chosen by cross validation.

The estimated EF coefficients $\{\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{d}_n : 1 \le n \le N\}$ are location free in the sense that they do not depend on the point configuration's centroid, $(\hat{a}_0, \hat{c}_0)$. However, they do depend on the starting point of the discretized outline, and they also depend on the outline's size and the orientation of its first harmonic's semi-

major axis. The coefficients of the first harmonic may be used to normalize the higher-order coefficients to render them invariant to starting point, size and orientation using the following normalizing transformation (Kuhl and Giardina (1982)):

$$\begin{pmatrix} \hat{A}_n & \hat{B}_n \\ \hat{C}_n & \hat{D}_n \end{pmatrix} = \frac{1}{\hat{\lambda}} \begin{pmatrix} \cos\hat{\psi} & \sin\hat{\psi} \\ -\sin\hat{\psi} & \cos\hat{\psi} \end{pmatrix} \begin{pmatrix} \hat{a}_n & \hat{b}_n \\ \hat{c}_n & \hat{d}_n \end{pmatrix} \begin{pmatrix} \cos n\hat{\theta} & -\sin n\hat{\theta} \\ \sin n\hat{\theta} & \cos n\hat{\theta} \end{pmatrix}. \tag{4.3}$$

Here, $\{\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n : 1 \le n \le N\}$ are the normalized coefficients, $\hat{\theta} = (1/2)\arctan[2(\hat{a}_1\hat{b}_1 + \hat{c}_1\hat{d}_1)/(\hat{a}_1^2 + \hat{c}_1^2 - \hat{b}_1^2 - \hat{d}_1^2)]$, $\hat{\psi} = \arctan(\hat{c}_1^*/\hat{a}_1^*)$ and $\hat{\lambda} = (\hat{a}_1^{*2} + \hat{c}_1^{*2})^{1/2}$ where $\hat{a}_1^* = \hat{a}_1\cos\hat{\theta} + \hat{b}_1\sin\hat{\theta}$ and $\hat{c}_1^* = \hat{c}_1\cos\hat{\theta} + \hat{d}_1\sin\hat{\theta}$. The normalized EF coefficients for the first harmonic satisfy $\hat{A}_1 = 1$, $\hat{B}_1 = \hat{C}_1 = 0$ and $|\hat{D}_1| \le 1$. Moreover, $\hat{\theta}$ is the angle between the starting point and the best-fitting ellipse's semimajor axis encountered in a clockwise direction, $\hat{\psi}$ is the orientation (in a counterclockwise direction) of that same axis relative to the positive x-axis, $\hat{\lambda}$ is the length of that axis and $|\hat{D}_1|$ is the ratio of minor axis length to major axis length (and thus is a measure of eccentricity).

The normalizing transformation (4.3) allows the nonlocation information in the first N harmonics of the outline to be reparameterized from $\{(\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{d}_n) : 1 \le n \le N\}$ to $\{\hat{\theta}, \hat{\psi}, \hat{\lambda}, |\hat{D}_1|\} \cup \{(\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n) : 2 \le n \le N\}$, where all information about starting point and orientation is contained in $\hat{\theta}$ and $\hat{\psi}$, respectively, all information about size is contained in $\hat{\lambda}$ and $|\hat{D}_1|$ and all shape information is contained in $|\hat{D}_1| \cup \{(\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n) : 2 \le n \le N\}$. In biological applications of morphometrics, "location" has no meaning, and the shapes of objects typically are of much greater interest than their size and orientation, hence statistical analyses of biological objects based on normalized EF coefficients usually use only the $4N - 3$ coefficients that remain after excluding $\hat{a}_0$, $\hat{c}_0$, $\hat{A}_1$, $\hat{B}_1$, and $\hat{C}_1$ (or equivalently $\hat{a}_0$, $\hat{c}_0$, $\hat{\theta}$, $\hat{\psi}$, and $\hat{\lambda}$). In our context, however, because it is important to compare called strike zones to the RBSZ—which has a well-defined location, size and orientation in addition to a well-defined shape—we retained nearly all of this information. The only quantity we excluded was $\hat{\theta}$, which is of no interest. However, we replaced $\hat{\lambda}$ with the more pertinent size variable $\hat{\kappa} = \pi\hat{\lambda}^2|\hat{D}_1|$ which is the area of the best-fitting ellipse. Therefore, our statistical analyses of fitted EF(N) CSZ outlines utilized the $4N + 1$ coefficients $\{\hat{a}_0, \hat{c}_0, \hat{\psi}, \hat{\kappa}, |\hat{D}_1|\} \cup \{\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n : 2 \le n \le N\}$. The first five of these correspond directly to discernible geometric features of CSZs, but the remainder do not. R code for fitting the elliptic Fourier models is provided in Section B of the Supplementary Material (Zimmerman, Tang and Huang (2019)).

4.2. *Superelliptic models*. The CSZ's shape, which appears (as noted in Section 3) to be a hybrid of a rectangle and an ellipse with possibly some axial asymmetries, leads us to consider, alternatively, representing its outline by the equation of a superellipse (SE) or some asymmetric generalization thereof. Such a model gains parsimony at the expense of flexibility relative to EF models of order two or higher, and all of its parameters correspond directly to discernible geometric

features of CSZ outlines. We consider only superellipses that are aligned with the (x, y)-axes. Such an object is a set of points (x, y) that satisfy the equation

$$\left|\frac{x - x_0}{a}\right|^{2r} + \left|\frac{y - y_0}{b}\right|^{2r} = 1,$$

where (x_0, y_0) is the center, a is the half-width, b is the half-height and $r > 0$ is the rectangularity index. The value of r strongly influences the superellipse's shape: $r < 0.5$ yields a four-armed star with concave sides; $r = 0.5$ corresponds to a rhombus; and $r > 0.5$ yields a convex, bilaterally symmetric (with respect to each of the x- and y-axes) object, with $r = 1$ corresponding to an ordinary ellipse. The "corners" of the superellipse become less rounded (more rectangular) as r increases; in the limit, as $r \to \infty$, the superellipse becomes a rectangle, the shape of the RBSZ. The top row of Figure 5 displays superellipses for which $a = 1.2$, $b = 1.0$ and $r \in \{1.0, 1.5, 2.5, +\infty\}$. Some generalizations of superellipses that allow for asymmetry of types observed in many CSZs are displayed in the bottom row. These generalizations are as follows:

- *Affine-transformed superellipse (ATS)*

$$\left|\frac{(x - x_0) + s(y - y_0)}{a}\right|^{2r} + \left|\frac{y - y_0}{b}\right|^{2r} = 1, \quad -\infty < s < \infty.$$

- *Latitudinally asymmetric superellipse (LAS)*

$$\left|\frac{x - x_0}{a}\right|^{2r_1} + \left|\frac{y - y_0}{b}\right|^{2r_1} = 1 \quad \text{if } y \geq y_0,$$

$$\left|\frac{x - x_0}{a}\right|^{2r_2} + \left|\frac{y - y_0}{b}\right|^{2r_2} = 1 \quad \text{if } y < y_0, r_1 > 0, r_2 > 0.$$

- *Affine-transformed latitudinally asymmetric superellipse (ATLAS)*

$$\left|\frac{(x - x_0) + s(y - y_0)}{a}\right|^{2r_1} + \left|\frac{y - y_0}{b}\right|^{2r_1} = 1 \quad \text{if } y \geq y_0,$$

$$\left|\frac{(x - x_0) + s(y - y_0)}{a}\right|^{2r_2} + \left|\frac{y - y_0}{b}\right|^{2r_2} = 1$$

$$\text{if } y < y_0, r_1 > 0, r_2 > 0, -\infty < s < \infty.$$

These models have five, six and seven parameters, respectively, making them more parsimonious than all EF models except EF(1). The parameters correspond directly to distinct geometric features: (x_0, y_0), a and b are defined as for a superellipse, while r_1 and r_2 are the top-half and bottom-half rectangularity indices, allowing the top half to be more rectangular (less rounded) than the bottom half or vice versa. The horizontal shear parameter, s, corresponds to an affine transformation on the superellipse that displaces every point horizontally by an amount proportional to

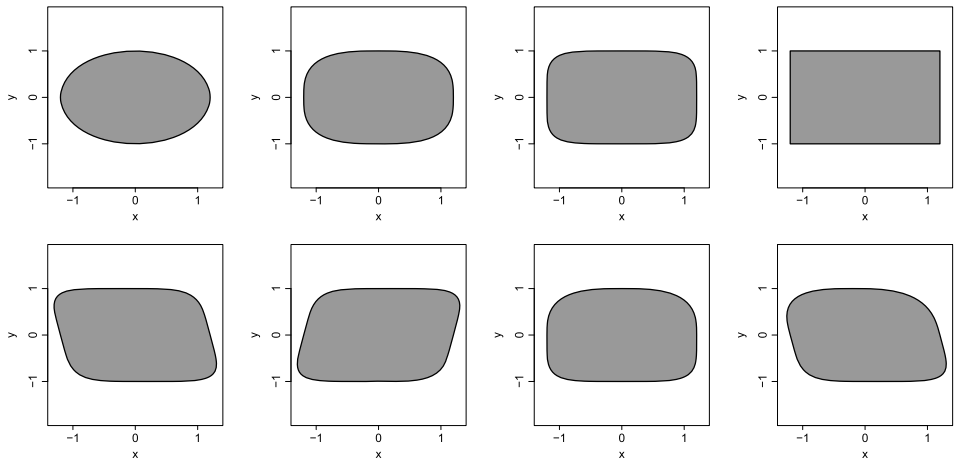

FIG. 5. *Examples of superellipses* (*top row*) *and generalizations* (*bottom row*: *first two plots*, *ATS*; *third plot*, *LAS*; *last plot*, *ATLAS*). *In all panels* $a = 1.2$ *and* $b = 1.0$. *Moving left to right in the top row*, $r = 1.0, 1.5, 2.5, +\infty$. *Moving left to right in the bottom row*, $(r_1, r_2, s) = (2.5, 2.5, 0.2)$, $(2.5, 2.5, -0.2)$, $(1.5, 2.5, 0)$ *and* $(1.5, 2.5, 0.2)$.

its y coordinate. This has the effect of elongating opposite corners of the superellipse in a way that is consistent with the opposite-handedness plots in Figure 2. Positive values of s elongate the upper-left and lower-right corners and shorten the others, while negative values do the opposite, creating an apparent "tilt" to the left or the right, respectively (see the second row of Figure 5). Two additional geometric features of importance for each generalized superellipse are its eccentricity, defined as $E = a/b$, and its area, which for the most general case of an ATLAS is given by

$$A = 4ab\left[\left(\frac{\{\Gamma[1+(2r_1)^{-1}]\}^2}{2\Gamma(1+r_1^{-1})}\right)+\left(\frac{\{\Gamma[1+(2r_2)^{-1}]\}^2}{2\Gamma(1+r_2^{-1})}\right)\right]$$

with appropriate simplification for the other cases. Note, importantly, that the area of an ATLAS is a monotone increasing function of each of a, b, r_1 and r_2.

Some methods for fitting superellipses to data are reviewed by Rosin (2000), but they are not optimal in any known sense nor do they extend easily to the ATS, LAS and ATLAS. We prefer a fitting approach that minimizes the sum of squared distances of the point configuration along normals to the superellipse and extends easily to its generalizations. Accordingly, we developed a distance-based orthogonal distance fitting (ODF) algorithm of a type described by Ahn (2004) for general closed curves which satisfies these criteria. Every (generalized) superellipse described above can be represented by an equation of the form $F(x, y, \boldsymbol{\theta}) = 0$ or, equivalently, in centered form by $F_c(x - x_0, y - y_0, \boldsymbol{\theta}_c) = 0$ where $\boldsymbol{\theta} = (x_0, y_0, \boldsymbol{\theta}_c)^T \in \Theta$ is the vector of model parameters (with Θ representing the parameter space) and $\boldsymbol{\theta}_c$ is the subvector of said parameters that ex-

cludes x_0 and y_0. Given a point configuration $\{(x_i, y_i) : i = 1, \ldots, I\}$, a distance-based ODF algorithm chooses the (generalized) superellipse $S(\boldsymbol{\theta}) = \{(x, y) \in R^2 : F_c(x - x_0, y - y_0, \boldsymbol{\theta}_c) = 0\}$ to minimize the sum of squared orthogonal distances from $S(\boldsymbol{\theta})$ to the point configuration, that is, to minimize

$$\Phi(\boldsymbol{\theta}) = \sum_{i=1}^{I} \big[\big(x_i - x_i'(\boldsymbol{\theta})\big)^2 + \big(y_i - y_i'(\boldsymbol{\theta})\big)^2\big],$$

where $\{(x_i'(\boldsymbol{\theta}), y_i'(\boldsymbol{\theta})) : i = 1, \ldots, I\} \subset S(\boldsymbol{\theta})$, over $\boldsymbol{\theta} \in \Theta$. We accomplish this by a "variable-separation" method consisting of two steps:

1. Given the current parameter estimate $\hat{\boldsymbol{\theta}}^{(k)}$, find the set of minimum distance points $\{(x_i', y_i') : i = 1, \ldots, I\}$ in $S(\hat{\boldsymbol{\theta}}^{(k)})$ using a generalized Newton method;

2. Given a set of minimum distance points $\{(x_i', y_i') : i = 1, \ldots, I\} \subset S(\hat{\boldsymbol{\theta}}^{(k)})$, update $\hat{\boldsymbol{\theta}}^{(k)}$ to $\hat{\boldsymbol{\theta}}^{(k+1)}$ using a Gauss–Newton algorithm.

Starting with an initial estimate $\hat{\boldsymbol{\theta}}^{(0)}$, Steps 1 and 2 are repeated until convergence. Details of the generalized Newton method of Step 1 and the Gauss–Newton algorithm of Step 2 and an `R` package for fitting the superelliptic models are provided in Sections C and D, respectively, of the Supplementary Material (Zimmerman, Tang and Huang (2019)).

The bottom row of Figure 4 displays minimum orthogonal distance fits of the SE, ATS, LAS and ATLAS models to the same CSZ for which some EF fits were displayed in the top row. It appears that the ATLAS is the best-fitting superelliptic model to this CSZ and that it also fits better than EF models of order up to (and including) four.

5. Evolution of the called strike zone from 2008–2016. EF models (up to order $[I/2]$) and superelliptic models were fitted to the first-stage outlines for each of the nine years, 2008–2016, which were displayed in Figure 1. Since there is no obvious common probabilistic model for residuals from the two types of models, we used the prediction sum of squares from tenfold cross validation as a model selection criterion with distance from points in the point configuration to the fitted outline measured orthogonally. Table 2 lists these values for EF models up to order 5 and for the four superelliptic models, which indicate that the ATLAS had better predictive capability than the other (generalized) superellipses and than all EF models of order less than 5. Thus, in order for an EF model to outperform the ATLAS, 21 EF parameters were required, compared to the seven parameters of an ATLAS.

Because the ATLAS was able to describe year-specific CSZs so much more parsimoniously than an EF model, we proceeded with statistical analysis of the fitted ATLAS coefficients only. Table 3 lists the estimated ATLAS coefficients

TABLE 2
Sums of squared prediction errors from tenfold cross validation for models fit to year-specific called strike zones

	Elliptic Fourier models					Superelliptic models			
Year	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$	SE	ATS	LAS	ATLAS
2008	1.242	0.953	0.092	0.071	0.022	0.264	0.253	0.039	0.030
2009	1.085	0.888	0.093	0.081	0.023	0.205	0.186	0.045	0.030
2010	1.114	0.906	0.103	0.086	0.022	0.216	0.191	0.050	0.028
2011	1.218	0.963	0.138	0.119	0.027	0.282	0.244	0.086	0.052
2012	1.215	0.959	0.133	0.108	0.015	0.263	0.230	0.061	0.030
2013	1.203	0.992	0.147	0.125	0.024	0.238	0.203	0.081	0.057
2014	1.280	1.099	0.163	0.147	0.017	0.199	0.175	0.060	0.042
2015	1.565	1.266	0.202	0.166	0.015	0.295	0.270	0.080	0.060
2016	1.424	1.199	0.188	0.164	0.017	0.243	0.209	0.071	0.046

(plus area and eccentricity) for each year, with the corresponding RBSZ values provided for comparison. Figure 6 displays plots of these quantities against year. Several trends are evident:

- The horizontal center of the CSZ, $\hat{x}_0$, shifted left considerably from 2008 to 2009 but after that trended back to the right; nevertheless, it was to the left of the RBSZ's center (by $\frac{3}{4}$ to $1\frac{1}{3}$ inches) over the entire nine-year period.
- The CSZ's vertical center, $\hat{y}_0$, was about one inch above that of the RBSZ in 2008, but then trended mostly downwards and by 2016 was slightly ($\frac{1}{7}$ inch) below that of the RBSZ. This 1.12-inch downward movement, in combination with a 1.49-inch increase in the half-height $\hat{b}$, implies that the bottom edge of the CSZ is estimated to have extended 2.61 inches lower in 2016 than in 2008.
- The half-width, $\hat{a}$, decreased monotonically, and the half-height, $\hat{b}$, increased monotonically over the period. Accordingly, the eccentricity decreased monotonically from about 1.30 to 1.01. Nevertheless, the CSZ was considerably wider than the RBSZ (which has eccentricity 0.908) over the entire nine-year period. At its narrowest, in 2016, the CSZ was still 17% wider than the RBSZ. In contrast, the CSZ was shorter than the RBSZ initially but taller after 2013; by 2016 it was 5% taller than the RBSZ.
- For all but the first and last years of the period, the annual increase in the CSZ's height was large enough relative to its decrease in width that its area increased monotonically. The CSZ was larger than the RBSZ over the entire period, about 10% larger by 2016.
- The upper- and lower-half rectangularity indices, $\hat{r}_1$ and $\hat{r}_2$, both tended to increase, indicating that the shape of the CSZ became more rectangular over time. Notably, $\hat{r}_2$ was consistently about 50% larger than $\hat{r}_1$, that is, the CSZ's lower half was consistently more rectangular than its upper half.

TABLE 3

Estimated ATLAS coefficients, plus eccentricity and area, for years 2008–2016. *Units of* $\hat{x}_0$, $\hat{y}_0$, $\hat{a}$ *and* $\hat{b}$ *are feet; units of* $\hat{A}$ *are square feet; the remaining coefficients are unitless. The corresponding RBSZ values are provided for comparison. The last row gives two-sided* P*-values corresponding to Mann–Kendall tests for trend*

Year	$\hat{x}_0$	$\hat{y}_0$	$\hat{a}$	$\hat{b}$	$\hat{E}$	$\hat{A}$	$\hat{r}_1$	$\hat{r}_2$	$\hat{s}$
2008	−0.085	2.592	1.087	0.838	1.297	3.176	1.160	1.832	−0.017
2009	−0.114	2.593	1.060	0.842	1.259	3.124	1.209	1.780	−0.023
2010	−0.098	2.570	1.055	0.853	1.237	3.153	1.217	1.813	−0.026
2011	−0.101	2.568	1.035	0.879	1.177	3.210	1.237	1.903	−0.031
2012	−0.096	2.535	1.027	0.900	1.141	3.262	1.236	1.915	−0.028
2013	−0.081	2.513	1.009	0.921	1.095	3.296	1.290	1.889	−0.024
2014	−0.081	2.481	0.993	0.945	1.051	3.354	1.346	1.955	−0.020
2015	−0.064	2.476	0.984	0.956	1.029	3.377	1.331	2.140	−0.020
2016	−0.062	2.498	0.970	0.962	1.008	3.346	1.348	2.052	−0.022
RBSZ	0	2.510	0.829	0.912	0.908	3.026	∞	∞	0
P-value	0.0091	0.0025	0.0003	0.0003	0.0003	0.0049	0.0012	0.0092	0.92

- The shear, $\hat{s}$, was quite small and did not exhibit a consistent trend over time.
- For several of the ATLAS parameters exhibiting trend, the trend moderated somewhat from 2014 to 2016, indicating that the strike zone was somewhat more consistent over the last three years than earlier in the nine-year period.

To quantify the strength of evidence for each trend noted above, we used a nonparametric approach, specifically a Mann–Kendall test, because of the small number of years involved and our reluctance to assume a particular form (e.g., linear) for the functional dependence of the ATLAS coefficients on time. P-values associated with these tests (last line of Table 3, unadjusted for multiplicity) indicated that the evidence against a no-trend hypothesis was strong ($P < 0.01$) for all ATLAS coefficients except shear.

In summary, from 2008–2016 the CSZ morphed from an object that was considerably wider than it was tall to one whose height and width were nearly equal to each other. By 2016 both the height and width of the CSZ were closer to those of the RBSZ, but the difference in widths was almost three times the difference in heights. Moreover, from 2009 forward the center of the CSZ moved much closer, both horizontally and vertically, to that of the RBSZ, and the CSZ became much more rectangular, implying that umpires were improving in correctly calling as strikes those pitches that were in the corners of the RBSZ. The only geometric attribute of the CSZ that became less like its RBSZ counterpart was area, which was larger than that of the RBSZ in 2008 and became even more so over time.

What impacts, if any, did the changes in the CSZ over the nine-year period 2008–2016 have on the game of baseball? Because a larger strike zone favors the pitcher over the batter, one might expect that the number of strikeouts per game

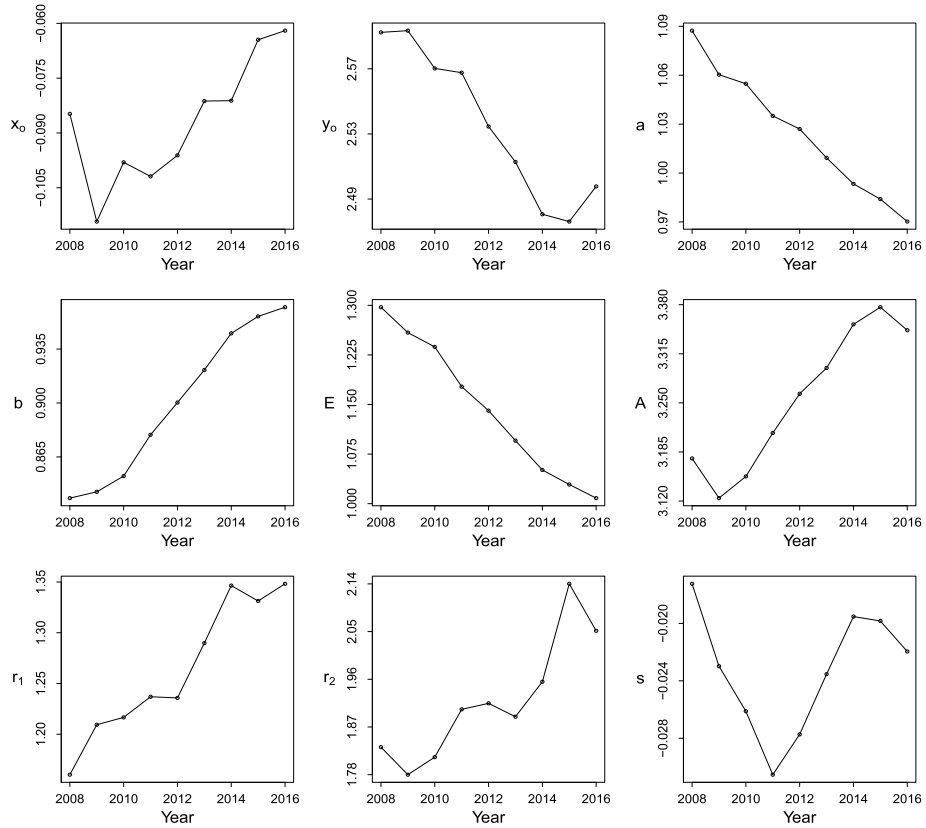

FIG. 6. *Estimated ATLAS coefficients* (*plus area and eccentricity*) *versus year.*

would have increased, and perhaps the numbers of walks and runs scored would have decreased over the period. Table 4 (first two columns) shows that indeed the number of strikeouts increased steadily. The numbers of walks and runs scored (third and fourth columns), however, exhibited less clearcut behaviors; both decreased until 2014 and then increased substantially over the remaining two years. Of course, many other relevant aspects of the game of baseball may have changed over the same period, so it is not an easy matter to determine how much of a change in such a variable can be attributed directly to changes in the CSZ. Nevertheless, Mills (2016a) established convincingly that the substantial downward extension of the lower boundary of the CSZ from 2008–2014 resulted in pitchers throwing a larger proportion of pitches below 21 inches in height and batters swinging at such pitches considerably more often and that 28% to 43% of the decline in runs scored from 2008–2014 could be directly attributed to changes in the CSZ. Two other analyses using different methods (Lindbergh (2015), Roegele (2014a)) attributed roughly the same amount of the decrease in run scoring to the expansion of the CSZ through 2014. What are we to make, however, of the turnaround in runs

TABLE 4

Numbers of strikeouts, walks and runs and proportions of certain types of pitches ($\times$ 100%) *by year. Proportions are defined as follows*: p_1 = *proportion of called pitches between* 18 *and* 21 *inches in height and lying between the left and right boundaries of the RBSZ that were called strikes*; p_2 = *proportion of pitches thrown that were within the RBSZ*; p_3 = *proportion of pitches thrown that were below* 21 *inches*; p_4 = *proportion of pitches thrown that batters swung at*; p_5 = *proportion of pitches below* 21 *inches that batters swung at*

Year	Strikeouts	Walks	Runs scored	p_1	p_2	p_3	p_4	p_5
2008	6.77	3.36	4.65	23.6	44.9	22.1	45.5	31.2
2009	6.91	3.42	4.61	23.4	45.1	22.3	44.9	31.1
2010	7.06	3.25	4.38	29.1	45.3	23.2	45.1	31.7
2011	7.10	3.09	4.28	34.2	45.2	23.8	45.7	32.6
2012	7.50	3.03	4.32	44.1	45.1	25.5	45.8	33.3
2013	7.55	3.01	4.17	51.4	45.0	26.3	46.1	33.9
2014	7.70	2.88	4.07	61.1	45.0	27.6	46.4	34.7
2015	7.71	2.90	4.25	65.8	44.4	28.5	47.1	35.4
2016	8.03	3.11	4.48	62.6	44.7	28.4	46.7	35.0

scored in 2015 and 2016? Table 4 shows that the proportion of called pitches lying between the left and right boundaries of the RBSZ and between 18 and 21 inches in height (roughly the lowest three inches of the RBSZ) continued to increase in 2015 but not in 2016. Similarly, the proportion of pitches thrown that were below 21 inches and the proportion of pitches below 21 inches that batters swung at continued to increase in 2015 but not in 2016. The abrupt halts to these trends in 2016 may be partially responsible for the increase in walks and runs in 2016, but it is likely that other factors are involved. We defer further discussion to Section 8.

6. Effects of player attribute/game situation factors. In order to study the effects of selected player attribute and game situation factors on the geometry of the called strike zone, KDA-based outlines corresponding to each of the 96 ($= 2 \times 2 \times 2 \times 4 \times 3$) combinations of pitcher handedness, batter handedness, venue, ball count and strike count were obtained for the data from 2014–2016. Sample sizes differed considerably across these 96 combinations, ranging from a minimum of 710 for the left-left-home-3-0 combination to a maximum of 76,414 for the right-right-home-0-0 combination. EF models (up to order $[I/2]$) and the four superelliptic models were fitted to the KDA-based outlines. The same ten-fold cross-validation procedure that was used to select among outline models for year-specific CSZs in Section 5 was applied to the fitted outlines. The ATLAS performed better than other superelliptic models for all but nine combinations. The ATLAS also performed better than the EF(2) model for all but two combinations and performed about equally as well as the EF(3) model. Since the EF(2) and EF(3) models have more parameters (nine and 13, respectively) than the ATLAS,

we again prefer the ATLAS. Displays of the first-stage outlines and the ATLAS model fits and lists of estimated ATLAS coefficients may be found in Section E of the Supplementary Material (Zimmerman, Tang and Huang (2019)). In what follows we report on analyses of ATLAS coefficients only; however, we repeated all analyses with an EF(3) model, obtaining similar results (not shown) for those EF coefficients that are analogous to ATLAS coefficients ($\hat{a}_0, \hat{c}_0, \hat{\kappa}, |\hat{D}_1|$).

6.1. *MANOVA of all coefficients.* Let $\mathbf{y}_{jklmq}$ denote the nine-dimensional vector comprising the fitted seven-coefficient ATLAS representation of the CSZ outline, plus the two derived variables area and eccentricity, for the jth level of pitcher handedness ($j = 1$ for right-handed pitchers, $j = 2$ otherwise), kth level of batter handedness ($k = 1$ for right-handed batters, $k = 2$ otherwise), lth level of venue ($l = 1$ for home, $l = 2$ for away), ball count m ($m = 0, 1, 2, 3$), and strike count q ($q = 0, 1, 2$). We performed a (weighted) multivariate analyses of variance (MANOVA) of the vectors $\{\mathbf{y}_{jklmq}\}$, based on a $2 \times 2 \times 2 \times 4 \times 3$ factorial model with an overall mean, main effects for each factor and all two-factor interactions. The residual vectors corresponding to the 96 factor combinations were assumed to be independent multivariate normal random vectors with mean vector zero. Owing to the large disparity among sample sizes, the variance-covariance matrix of the residual vectors was taken to be $n_{jklmq}^{-2/3}\boldsymbol{\Sigma}$ where n_{jklmq} is the number of called pitches that were used to create the outline for the $jklmq$th combination; this is in accordance with the fact that the asymptotic variance of a two-dimensional kernel density estimator with optimally chosen bandwidth, from a random sample of size n, is $O(n^{-2/3})$ (Wasserman (2006)). Note that we did not weight by sample size in the analysis of yearly CSZs presented in the previous section because sample sizes differed very little across years.

As is common practice for unreplicated factorial designs, we used the *effect sparsity* principle to produce experimental error for the MANOVA (and subsequent ANOVAs). Specifically, we applied Lenth's method (Lenth (1989)) to determine which of the effects are null. This method requires the design to be a 2^k factorial with no weighting, so for this purpose only we collapsed the three levels of "Strike" and four levels of "Ball" into two (by combining the called pitches on counts with zero or one strike, and doing likewise for pitches on counts with zero or one ball and two or three balls), making our design a complete 2^5 factorial. The entire three-step estimation and modeling process was then repeated for this design without weighting, and estimates of all $2^5 - 1 = 31$ contrasts were determined to be null or nonnull according to Lenth's procedure. We found no evidence that any three-, four- or five-factor interaction effects were nonnull, and on this basis those effects were pooled in the weighted $2 \times 2 \times 2 \times 4 \times 3$ MANOVA and ANOVAs to produce experimental error. The experimental error can be regarded as an amalgamation of several distinct error components, including PITCHf/x instrument error, errors in representing the actual CSZ as the 50% outline of a kernel discriminant analysis,

TABLE 5
Weighted level means of estimated ATLAS coefficients, plus area and eccentricity, corresponding to each player attribute and game situation factor and selected two-factor combinations, for the combined 2014–2016 *data. Units are identical to those in Table* 3

Factor	Levels	$\hat{x}_0$	$\hat{y}_0$	$\hat{a}$	$\hat{b}$	$\hat{A}$	$\hat{E}$	$\hat{r}_1$	$\hat{r}_2$	$\hat{s}$
Pitcher	RHP	−0.081	2.481	0.943	0.931	3.156	1.016	1.396	2.010	−0.061
	LHP	−0.060	2.490	0.944	0.927	3.121	1.022	1.319	2.031	0.041
Batter	RHB	−0.028	2.487	0.960	0.926	3.195	1.040	1.391	2.117	0.009
	LHB	−0.130	2.481	0.924	0.935	3.082	0.991	1.345	1.895	−0.071
Venue	Home	−0.074	2.484	0.949	0.934	3.173	1.018	1.359	2.042	−0.029
	Away	−0.074	2.484	0.939	0.925	3.115	1.018	1.381	1.992	−0.025
Ball	0	−0.072	2.491	0.936	0.920	3.092	1.021	1.377	2.025	−0.025
	1	−0.073	2.485	0.944	0.915	3.102	1.034	1.353	2.110	−0.029
	2	−0.078	2.478	0.946	0.941	3.188	1.007	1.376	2.015	−0.027
	3	−0.079	2.463	0.969	0.993	3.405	0.978	1.381	1.729	−0.029
Strike	0	−0.076	2.479	0.985	0.996	3.498	0.989	1.408	1.805	−0.021
	1	−0.074	2.476	0.928	0.913	3.050	1.018	1.409	2.129	−0.025
	2	−0.071	2.502	0.893	0.840	2.669	1.066	1.263	2.237	−0.038

Batter × Strike:	Batter	Strike	$\hat{x}_0$	Batter × Pitcher:	Batter	Pitcher	$\hat{s}$
	RHB	0	−0.022		RHB	RHP	−0.018
	RHB	1	−0.022		RHB	LHP	0.053
	RHB	2	−0.045		LHB	RHP	−0.105
	LHB	0	−0.141		LHB	LHP	0.020
	LHB	1	−0.136				
	LHB	2	−0.103				

residuals from the fitted ATLAS model to points on the KDA-based outline and errors due to the omission from the model of covariates that influence the CSZ.

Wilk's lambda tests from the MANOVA indicated very strongly that all main-effect vectors and several second-order interaction effect vectors were nonnull. Furthermore, for all but one pair of ATLAS coefficients, there was no evidence of cross correlation among residuals, and for the exceptional pair ($\hat{y}_0$ and $\hat{r}_1$) the evidence was only moderate ($P = 0.010$, with a Bonferroni-based adjustment for multiplicity). Accordingly, we proceeded with a (weighted) univariate factorial ANOVA of each ATLAS coefficient plus area and eccentricity. Results of these analyses are summarized in subsequent subsections; Table 5 lists the (weighted) factor level means of each coefficient; the corresponding standard errors and ANOVA tables with P-values unadjusted for multiplicity may be found in Section E of the Supplementary Material (Zimmerman, Tang and Huang (2019)). In our summaries below, we discuss only those effects for which the evidence against the null hypothesis of no effect is "strong" ($0.001 < P < 0.01$) or "very strong"

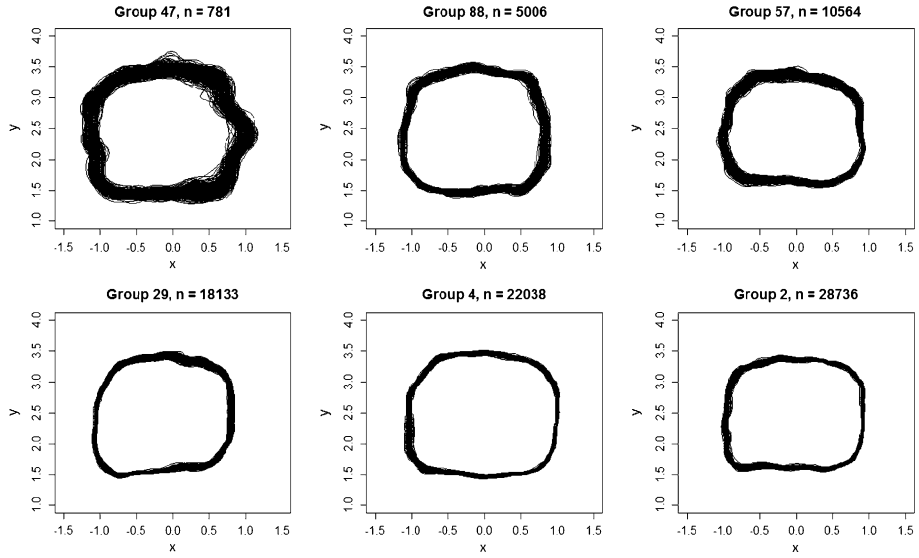

FIG. 7. *KDA-based CSZ outlines for* 299 *bootstrap samples, for six groups* (*factor combinations*) *having a range of sample sizes.*

($P < 0.001$) following Bland (2000). We deemed the sizes of all other effects, including some that are statistically significant in conventional terms ($P < 0.05$), as too small to be practically relevant. Even some effects for which the evidence against the null is very strong are not practically relevant, as will be seen.

In conjunction with each ANOVA, residual analyses including an assessment of normality were carried out. Evidence against normality existed for some coefficients, but in all such cases the distribution of fitted residuals was heavy tailed rather than skewed, so, taking into account the substantial number (63) of degrees of freedom for error, we judged the P-values obtained from the ANOVAs to be trustworthy.

Also, as a check for the appropriateness of weighting by $n^{2/3}$ in the ANOVAs, we took bootstrap samples of size 299 from the data for each of six factor combinations, chosen from among the 96 factor combinations so as to have sample sizes that range from very small ($n = 781$) to moderately large ($n = 28{,}736$). KDA-based outlines corresponding to each bootstrap sample are displayed in Figure 7. Sample variances of the widths and heights of the outlines closely adhere to the assumed inverse dependence on $n^{2/3}$. Furthermore, it is shown in Section E of the Supplementary Material (Zimmerman, Tang and Huang (2019)) that this inverse dependence on $n^{2/3}$ propagates to the ATLAS parameter estimates obtained via the second stage of analysis, and yields bootstrap-based variances of those estimates that, when pooled over factor combinations, are similar in magnitude to the residual mean squares from the ANOVAs. These results establish that using an ANOVA

with weights $n^{2/3}$ to estimate standard errors of ATLAS parameter estimates and the effects that player attribute/game situation factors have on them is sensible.

For the MANOVA/ANOVAs just described, we marginalized over years 2014–2016. One might question whether the results of this analysis are affected by changes in the CSZ over years that were documented in Section 5. Therefore, we repeated the analysis using a third-stage model with all the same effects plus main effects of years and interaction effects of year with every other factor. Results of this analysis are given in Section E of the Supplementary Material (Zimmerman, Tang and Huang (2019)). Although these year-by-year analyses indicate that some additional effects are nonzero for some ATLAS parameters, none of them are large enough to be practically relevant, and all conclusions based on the analysis marginalized over years (to be described in what follows) still hold. Numerical estimates of effects, however, must be interpreted as averages over the three years.

6.2. *Center.* The ANOVA of horizontal center ($\hat{x}_0$) provided very strong evidence of a nonzero batter handedness main effect and nonzero batter $\times$strike interaction effect. These two effects explained 88% and 6%, respectively, of the overall variation. The (weighted) level means of $\hat{x}_0$ were -0.34 (s.e. 0.02) inches for right-handed batters (RHBs) and -1.56 (s.e. 0.02) inches for left-handed batters (LHBs). Together with a half-width ($\hat{a}$) for LHBs that was 1.14 inches larger than its rule-book prescription, this shift implies that the CSZ for LHBs extends 2.70 (s.e. 0.03) inches to the left of the left boundary of the RBSZ. This feature, known as the "lefty strike," is undoubtedly practically relevant and has been a topic of much discussion among baseball analysts (e.g., Roegele (2013b)); we defer further discussion of it to Section 8. The effect size of the batter $\times$strike interaction was considerably smaller and manifested itself primarily as horizontal centers for LHBs and RHBs that were 0.73 (s.e. 0.07) inches closer to each other on a two-strike count than when there were no strikes. Although there was strong or very strong evidence of a nonzero pitcher handedness effect and some additional interaction effects, their effect sizes were less than one-third inch, that is, only about 1.5% of the total width of the RBSZ, which may not be practically relevant.

The ANOVA of vertical center ($\hat{y}_0$) indicated very strong evidence for three nonzero main effects, but only those of ball and strike count, which manifested as a lowering of the centroid as the strike count increased and a slight raising as ball count increased, were practically relevant. The centroid of the CSZ for an 0-2 count was 0.66 (s.e. 0.09) inches higher than that for a 3-0 count. Sizes of the remaining effects were less than one-eighth inch which is much less than 1% of the height of the RBSZ.

6.3. *Width, height and area.* We found very strong evidence that all main effects on the half-width ($\hat{a}$) save that of pitcher handedness were nonzero with the strike, batter handedness, ball and venue effects explaining 74%, 13%, 4% and 1%, respectively, of the overall variation. There was also some evidence for three

nonzero two-factor interactions, but their effect sizes were not practically relevant. The difference in mean half-widths for RHBs and LHBs was 0.43 (s.e. 0.03) inches, indicating that RHBs had a CSZ almost one inch wider than that of LHBs. Thus, while our analysis of the horizontal center indicated that the CSZ of LHBs extended further off the outside of the plate than that of RHBs, this analysis indicates that LHBs actually had a substantially narrower CSZ. The strike and ball main effects were of opposite sign: as the strike count increased, the CSZ became narrower by an average of about 1.10 (s.e. 0.07) inches per strike, but as the ball count increased it became wider by an average of about 0.26 (s.e. 0.09) inches per ball. Thus the average decrease in width with each additional strike was about four times the average increase in width with each additional ball. Although the evidence of a nonzero venue main effect was very strong and indicated that pitchers had a wider strike zone when they pitched at home, the difference in width was only 0.24 (s.e. 0.03) inches, hence perhaps not very important from a practical standpoint.

The results for half-height ($\hat{b}$) were similar in many respects to those of half-width. Evidence of nonzero main effects was very strong for all factors save pitcher handedness, and no two-factor interaction was practically relevant. However, in this case the strike and ball counts explained 85% and 10%, respectively, of the overall variation, and the batter handedness effect was not practically relevant. The strike and ball main effects were again of opposite signs, but their magnitudes were somewhat greater than the magnitudes of their effects on $\hat{a}$. More specifically, as the strike count increased the CSZ became shorter by an average of 1.87 (s.e. 0.09) inches per strike, and as the ball count increased it became taller by an average of about 0.58 (s.e. 0.11) inches per ball. We found very strong evidence that pitchers had a taller strike zone when they pitched at home, but the difference of 0.22 (s.e. 0.03) inches was not very large.

Owing to the similarities between the results for $\hat{a}$ and $\hat{b}$ just presented, the results for the derived variable, area, are similar. Only the strike, ball and batter handedness main effects were practically relevant. Particularly noteworthy was the difference in the CSZ's area for the most extreme ball-strike counts, 0-2 and 3-0; means for those combinations were 2.42 (s.e. 0.02) and 4.00 (s.e. 0.03) square feet, respectively. Thus, the CSZ was 65% larger when the count was 3-0 than when it was 0-2! Some possible explanations for this effect are offered in Section 8. Finally, we note that while there was very strong evidence of home-field advantage in the form of a larger strike zone for a pitcher at home, the difference in areas at the two venues was only 0.06 (s.e. 0.01) square feet (1.9% of the CSZ's area).

6.4. *Eccentricity, rectangularity and shear.* The ANOVA of $\hat{E}$ provided very strong evidence of nonzero batter handedness, strike and ball main effects. CSZs for RHBs were more eccentric than those for LHBs. Strike and ball count effects had opposite signs, with eccentricity consistently increasing as the strike count increased or ball count decreased.

Turning our attention to rectangularity, the evidence was strong and very strong, respectively, that the main effects of pitcher handedness and strike on $\hat{r}_1$ were nonzero. The top half of the CSZ was more rounded for LHPs than RHPs and, regardless of pitcher handedness, became more so as the strike count increased. However, the effect sizes were rather small and not discernible in Figures 2 and 3. For $\hat{r}_2$, evidence of nonzero batter handedness and strike main effects was very strong, but the first of these was not large. The bottom half of the CSZ became more rounded as the ball count increased from two to three, but more rectangular (i.e., more like that of the RBSZ) as the strike count increased. The bottom-half effects of the ball-strike count on shape are large enough to be discernible in Figure 3.

Finally, there was very strong evidence of nonzero pitcher and batter handedness main effects and their interaction effect on $\hat{s}$. Shear coefficients for CSZs corresponding to RHBs or LHPs were positive on average while those corresponding to LHBs or RHPs were negative. Shear coefficients for all four pitcher-batter handedness combinations were quite different with the largest difference existing between pitcher-batter combinations of opposite handedness in a manner that expands the CSZ in a down-and-away direction, as Figure 2 illustrated.

7. Variability between and within umpires. The analyses reported in the previous two sections considered called strike zones that were marginalized over umpires. Students of the game of baseball are aware, however, that umpires seem to differ in how they call balls and strikes. This section compares the variability of CSZ outlines corresponding to different umpires (inter-umpire variability) to variability within individual umpires (intra-umpire variability). Of course, obtaining a separate outline for each umpire greatly reduces the number of called pitches from which the outlines may be determined. Furthermore, examining within-umpire variability requires that we randomly split the called pitches for a given umpire to create replicates, further reducing the sample sizes. So as to have sufficient data to reliably determine KDA-based outlines that can serve our purpose, we subsetted the called pitches and restricted our focus as follows. First, we used data over merely the three-year period 2014–2016 (during which time the CSZ was more homogeneous than earlier) and from only the 86 umpires who called at least 500 pitches during this period for each subset. Second, the only player attribute/game situation factors that we accounted for were batter handedness and count; we marginalized over the remainder. The analysis in Section 6 justifies these marginalizations, as it indicated that the effects of venue on CSZ were not practically relevant and that pitcher handedness affected only the rectangularity and shear coefficients, which are of lesser importance for comparing umpires than other ATLAS coefficients and will be excluded from this analysis. Even with these marginalizations, however, after random splitting of each umpire's called pitches into replicate halves, only those outlines corresponding to a 0-0 count were based on sufficient sample sizes to be reliable, so it is those to which an ATLAS model was fit. Figure 8 shows KDA-based and optimally fitted ATLAS outlines for four umpires,

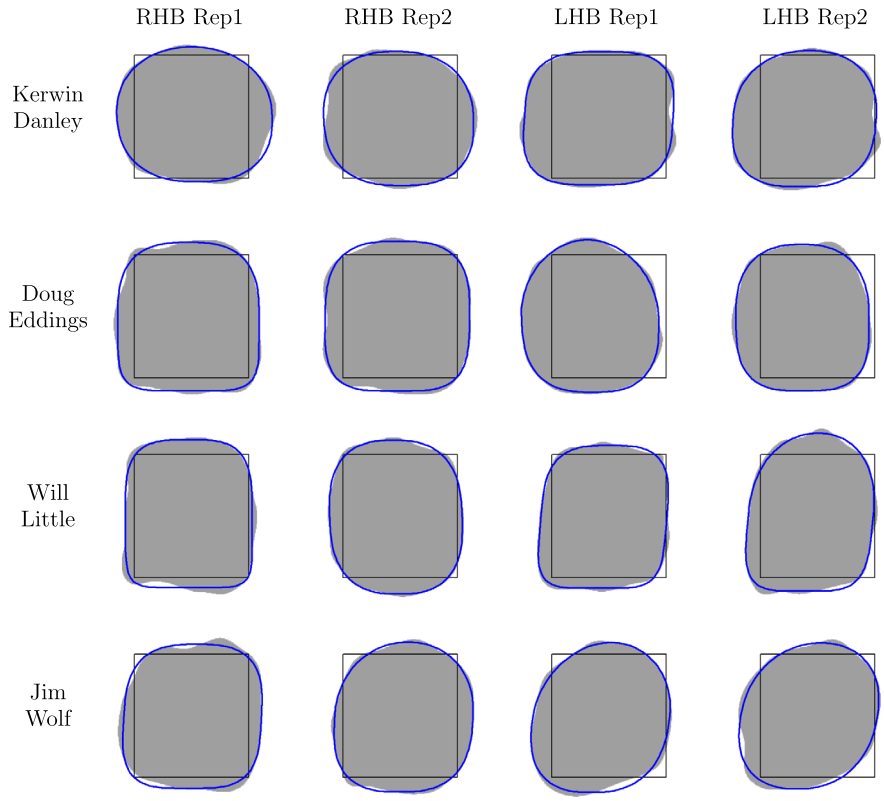

FIG. 8. *KDA-based (shaded regions) and fitted ATLAS called strike zones (bold outlines) on* 0-0 *counts from* 2014–2016 *for four umpires superimposed upon the average RBSZ.*

illustrating typical differences between umpires and between the two replications of the same umpire. For outlines corresponding to the same batter handedness, greater geometric variability in the CSZ is evident between umpires than between replicates from the same umpire.

Let y_{ijk} denote an arbitrary fitted coefficient from the ATLAS representation of the CSZ outline corresponding to the ith level of batter handedness, jth umpire, and kth replicate. The following two-factor crossed mixed effects model with interaction was fit to each such coefficient:

$$y_{ijk} = \mu + \beta_i + u_j + (\beta u)_{ij} + e_{ijk}. \tag{7.1}$$

Here, μ is an overall fixed effect, β_i is the fixed effect of the ith level of batter handedness, u_j is the random effect of umpire j, $(\beta u)_{ij}$ is the random interaction effect between the ith level of batter handedness and umpire j and e_{ijk} is the random effect of the kth replicate. We assumed that the u_js, $(\beta u)_{ij}$s, and the e_{ijk}s are mutually independent random variables with $N(0, \sigma_u^2)$, $N(0, \sigma_{\beta u}^2)$ and $N(0, \sigma_e^2)$ distributions, respectively. Note that we regard the effects of umpires

TABLE 6

Minimum variance quadratic unbiased estimates ($\times 10^6$) *of variance components in the mixed effects model* (7.1) *for selected ATLAS coefficients* (*plus area and eccentricity*) *and the corresponding point estimate and* 99% *upper confidence interval estimate of the proportion* γ *of variability attributable to umpires*

ATLAS coefficient	$\hat{\sigma}_u^2$	$\hat{\sigma}_{\beta u}^2$	$\hat{\sigma}_e^2$	$\hat{\gamma}$	Confidence interval
x_0	459	1238	987	0.731	$(0.634, \infty)$
y_0	199	360	945	0.372	$(0.207, \infty)$
a	1301	549	987	0.652	$(0.526, \infty)$
b	1122	209	996	0.572	$(0.424, \infty)$
A	14,830	10,510	13,500	0.652	$(0.530, \infty)$
E	2961	824	1920	0.663	$(0.538, \infty)$

(and thus also their interaction with batter handedness) as random variables. This seems more reasonable for our purposes than treating them as fixed effects because the question being addressed here pertains to the variability of these effects rather than their values for individual umpires.

Minimum variance quadratic unbiased estimates of the variance components under this model are listed in Table 6 for each ATLAS coefficient. (Because the data are balanced, these estimates are also residual maximum likelihood, i.e., REML, estimates.) Two features in the table are especially noteworthy. First, the estimates of σ_e^2 for the CSZ centroid's coordinates and its width and height are very similar, indicating that individual umpires were about equally consistent horizontally and vertically in their calls. Second, the estimated between-umpire variability, viz. $\hat{\sigma}_u^2 + \hat{\sigma}_{\beta u}^2$, is larger than the estimated within-umpire variability, $\hat{\sigma}_e^2$, for all ATLAS coefficients except vertical center. Estimates of the proportion of total variability of an ATLAS coefficient attributable to umpires, that is, $\hat{\gamma} = (\hat{\sigma}_u^2 + \hat{\sigma}_{\beta u}^2)/(\hat{\sigma}_u^2 + \hat{\sigma}_{\beta u}^2 + \hat{\sigma}_e^2)$, are also listed in the table, together with 99% upper confidence intervals for the corresponding true proportion. We derived the form of the confidence interval by manipulating and specializing a general expression for confidence intervals given by Lu, Graybill and Burdick (1987) for certain other ratios of variance components in random and mixed linear models; for details see Section F of the Supplementary Material (Zimmerman, Tang and Huang (2019)). It may be noted that the lower endpoint of the interval exceeds 0.5 for all geometric attributes except vertical center and half-height. These results provide strong evidence that, on 0-0 counts at least, the horizontal attributes (horizontal center and width) of the CSZ, plus its area and ellipticity, exhibit more variability between umpires than they do for replicates within umpires. No such statement can be made about the vertical attributes of the CSZ.

8. Discussion and future work. In this work we applied and extended methods of outline analysis in order to characterize the geometry of the called strike

zone and perform inference on it. The new affine-transformed latitudinally asymmetric superelliptic (ATLAS) model introduced here fit the data remarkably well, generally much better than existing elliptic Fourier models with a similar number of parameters. Another distinct advantage of the ATLAS is that all of its parameters, unlike those of EF models, correspond directly to geometric attributes of the CSZ that are of interest. In particular, the rectangularity coefficients of the ATLAS allow us to directly measure the adherence (or lack thereof) of the top and bottom halves of the called strike zone to the rectangular shape prescribed by the rule book, and the shear coefficient accounts for the observed longitudinal asymmetry in the CSZ for pitcher-batter combinations of opposite handedness. Additional contributions of this work to outline analysis include the development of an orthogonal distance fitting algorithm for (generalized) superelliptic models and an expansion of follow-up methodologies for outline model coefficients to, specifically, tests for time trend, factorial analyses of variance for determining moderators of outline geometry and mixed effects models/variance component estimation for characterizing outline variability among and between umpires. The superelliptic models, fitting algorithm, and follow-up methodologies developed here may in the future be applied, when appropriate, to outlines of other objects.

In Section 1, we briefly noted some advantages of an outline analysis relative to previously published analyses based on generalized additive models (GAMs) of the log odds of a called strike. We give a more detailed comparison of the two methodologies now. With one exception, authors presenting GAM-based analyses of the CSZ have not produced estimates of geometric attributes or provided confidence regions for the CSZs they display. The exception is Mills (2016a), who gave numerical values for the endpoints of a confidence interval for the CSZ's area and displayed a confidence region for the CSZ's outline but did not describe how he actually obtained them. In contrast, because the parameters of the model fit to the outline at its second stage directly represent important geometric attributes of the CSZ and because the factorial ANOVA or other third-stage analysis yields an estimate of error, an outline analysis results in straightforward inference on those attributes and straightforward comparison to the corresponding attributes of the RBSZ. Moreover, previous authors using GAMs have imposed pure additivity on the model, ignoring the possibility of interactions between factors, some of which we found to be important via outline analysis. It is possible, of course, to add interaction terms to a GAM but only with a corresponding increase in the effects of the curse of dimensionality. Thus, an outline analysis has several advantages over a GAM-based analysis. However, a distinct disadvantage of an outline analysis relative to GAM-based analyses is that the player attribute/game situation factors that can be considered are limited, as a consequence of requiring a sufficient number (several hundred) of called pitches in subsets corresponding to factor combinations in order to obtain reasonable kernel discriminant analysis-based outlines at the first stage. A GAM, when embedded within a Bayesian hierarchical framework, can borrow strength across factor combinations to produce useful inferences even

when the subsets contain very few pitches. For example, Deshpande and Wyner (2017) used hierarchical Bayesian GAMs to address questions about effects of individual catchers on the CSZ. Some of their models included effects of ball-strike count and individual catchers, umpires, pitchers and batters with factor combinations numbering in the millions. Nevertheless, they obtained meaningful results though it may be noted that inferences for a model with so many parameters are likely to be sensitive to prior specification and to the validity of their no-interaction assumption.

One main objective of our work was to perform inference on changes in the geometry of the called strike zone since the inception of the PITCHf/x system in 2008. We found that the CSZ evolved remarkably smoothly from 2008–2016, and that most of its geometric attributes were more like those of the RBSZ at the end of this period than at its beginning. The lone exception was the CSZ's size, which increased over time to the point that it was about 10% larger than that of the RBSZ by 2016. Two factors were mainly responsible for the increase in size: (1) a drop in the CSZ's lower boundary, and (2) an increase in its rectangularity. However, these changes actually made the CSZ more like the RBSZ, not less. The primary reason the CSZ was so much larger than the RBSZ by 2016 was its excessive width which, though it moderated over time, was still 17% larger in 2016 than that prescribed by the rule book. For comparison, the CSZ in 2016 was only about 5% taller than the rule-book prescription. Reports from a few years ago (e.g., Passan (2015)) indicate that Major League Baseball's Competition Committee, in hopes of reducing the size of the (called) strike zone and thereby increasing offensive production, was contemplating raising the lower boundary of the RBSZ from the hollow beneath the hitter's kneecap to the top of the knee. While such an action would likely reduce the size of the CSZ, it would seem that if adherence to the RBSZ is desired, it is the width, not the height, of the strike zone that should be receiving most of the Committee's attention. In any case, increased offensive output since the middle of the 2015 season appears to have rendered the issue moot, for now. Mills (2016b), seeking an explanation for the recent dramatic increase in run scoring despite a continued increase in strikeouts and only a very small decrease in the size of the CSZ, documented several changes by batters and pitchers over this period, including an increase in exit velocity (the speed at which a struck ball leaves the bat), an increase in the proportion of pitches thrown to the inner half of the plate (where batters can drive the ball with more power) and an increase in average launch angle (the vertical angle at which the ball leaves the bat after being struck). But as of this writing, the extent to which these changes are responsible for the increase in run scoring is unknown.

Our second main objective was to investigate how selected player attribute and game situation factors affect the CSZ's geometry. For the data from 2014–2016, we found that pitcher handedness, batter handedness and ball-strike count all had practically relevant effects on at least one geometric attribute of the CSZ. Of these effects, we regard four as worthy of additional discussion; two have been discussed

extensively by baseball analysts and two have not. Perhaps the most important of these is the effect that the ball-strike count has on the CSZ's size. In short, the CSZ expands as the number of balls increases and shrinks as the number of strikes increases. One possible explanation, termed the "Compassionate Umpire Theory" by Walsh (2010), is that umpires attempt (perhaps unconsciously) to help the pitcher by expanding the strike zone when the batter is "ahead in the count" (i.e., the count has more balls than strikes) and attempt to help the batter by shrinking the strike zone when he is behind in the count. Another, more plausible explanation, put forward by Mills (2014) and Green and Daniels (2015), could be called the "Reluctant Arbiter Theory." According to this theory, the umpire, preferring to let the pitcher and batter determine the outcome of an at-bat via game action (a swinging strike or a pitch struck and put into play), adjusts his strike zone with each pitch to reduce the chance that he will eventually have to make a decisive call of a third strike or fourth ball. A similar phenomenon has been observed in other sports such as football and basketball, where referees appear to be less likely to call certain penalties or fouls late in the game compared to earlier; see, for example, Moskowitz and Wertheim (2011). On the other hand, Lopez and Mills (2018) present evidence of an opposite effect in the bottom half of extra innings of baseball games, wherein umpires (again perhaps unconsciously) act in a way that increases the probability that the game will end after that inning, by calling more borderline pitches balls if the batting team is tied and has at least one man on base, while calling more such pitches strikes if the batting team is behind. Still another explanation for the ball-strike count's effect on the CSZ, advocated by Molyneux (2016) and called "Bayesian instinct" by Green and Daniels (2017), is that when the batter is ahead in the count, the pitcher needs to throw a strike, so the umpire's prior belief of a strike is higher; when the pitcher is ahead in the count, he can better afford to throw a ball, so the umpire's prior belief of a strike is lower. The umpire then calls the pitch by integrating this "rational expectations" prior with his observation of the pitch's location.

The other extensively documented noteworthy effect is the "lefty strike," that is, a strike zone for left-handed batters that extends outside the plate more than 2.5 inches beyond its rule-book prescription. A theory for this phenomenon has emerged (Fast (2011a), Roegele (2013b)), which is based on the notion that the visual cues the umpire uses to call a pitch include not only its location but also the movements (or lack thereof) of the batter and catcher it induces. According to this theory, greater movement by either player suggests to the umpire that the pitch was less likely to be in the RBSZ, increasing the likelihood that he calls it a ball. One component of the theory is that pitches near the inside edge of the RBSZ may appear to the batter as if they will hit him, possibly causing him to move out of harm's way, while pitches near the outside edge, if not swung at, will induce little movement. A second component is the fact that virtually all catchers catch the ball with their left hand, with the consequence that a pitch caught near the left edge of the RBSZ results in considerably less movement of the catcher's arm than

a pitch caught near the right edge (the former requires only a slight movement to the side while the latter requires the catcher to reach across his body). Together, these two components imply that a pitch to a left-handed batter that is caught near the left edge of the RBSZ will result in less movement by the batter and catcher than the other three possibilities. A final component of the theory is "pitch framing," whereby the catcher, in agreement with the pitcher, sets his glove in a predetermined location prior to receiving the pitch, providing a target for the pitcher and resulting in little catcher movement if the pitch is on target. If catchers set up outside the midpoint of the plate more often to left-handed batters than to right-handed batters, their arm will tend to move less for outside pitches to lefthanders than for outside pitches to righthanders. It is worth noting that 65.7% of called pitches to left-handed batters from 2014–2016 were outside the midpoint of the plate, compared to 61.9% of called pitches to right-handed batters. Recent in-depth analyses of pitch framing and its impacts are presented by Brooks, Pavlidis and Judge (2015) and Deshpande and Wyner (2017).

Two previously undocumented features of the CSZ revealed by our analysis are its latitudinal asymmetry, specifically the top half being smaller and more rounded than the bottom half, and that its elongation toward its low outside (to the batter) corner, noted previously by baseball experts, is attributable to pitcher-batter combinations of opposite handedness only. We speculate that both of these can be explained by a "movement" theory similar to the one offered for the lefty strike. In regard to the latitudinal asymmetry, catchers may set their target below the vertical midpoint of the RBSZ more often than not with the consequence that pitches caught in the top half tend to cause greater arm movement; furthermore, the umpire's eyes are located at this higher level so the catcher's body is less likely to obscure the precise location where the ball is caught if it is above the midpoint. Some supporting evidence for the first of these points is that 62.4% of called pitches in 2014–2016 crossed the vertical plane containing the plate's front edge below the vertical midpoint of the RBSZ. Unfortunately, the theory does not explain the substantial increase in rectangularity in the bottom half of the CSZ as strike count increases, for which we have no plausible explanation. As for the low-outside-corner elongation, we reiterate that the batter does not have to move out of the way of an outside-edge pitch; furthermore, the momentum of a pitch received on the side of the plate coinciding with the pitcher's handedness will tend to carry the catcher's mitt (and arm) into his body, resulting in little sideways arm movement, whereas the momentum of a pitch received on the opposite site will tend to carry the catcher's arm away from his body, resulting in greater arm movement.

Our final analytical objective was to characterize the variability in the geometry of called strike zones between umpires. For pitches on 0-0 counts from 2014–2016, we found that the horizontal center, width, area and eccentricity of the CSZ (but not its vertical center and height) exhibited greater variability between umpires than within umpires. Put another way, on 0-0 counts the (horizontal) variation in an individual umpire's CSZ from pitch to pitch is smaller than the (horizontal)

variation in CSZs of different umpires. This is probably desirable, as players and managers appear more willing to tolerate variability in strike zones called by different umpires than they will tolerate inconsistency in calls from the same umpire, particularly if the umpire is inconsistent over the course of a mere inning or game.

Outline-based methodologies similar to those used herein could be used to address many other interesting questions about the geometry of the called strike zone, such as whether pitch type and speed, or umpire age and experience, affect it. Moreover, rather than focusing merely on geometric attributes of the CSZ, future studies could also consider its precision, as measured for example by the average gradient along the KDA-based outline or, more simply, by the the area between KDA-based or ATLAS-based outlines corresponding to called strike probabilities of $(50-\alpha)\%$ and $(50+\alpha)\%$, where $0 < \alpha < 50$. In particular, it would be of interest to see whether the precision of the CSZ, not merely its accuracy, improved from 2008–2016, and, if so, whether the improvement occurred equally on all four sides of the strike zone, whether it was affected by the ball-strike count and whether individual umpires' improvements in precision over time were correlated with their improvements in accuracy. Answers to these questions will be reported elsewhere.

Our outline-analytic methods for characterizing the geometry of the strike zone moderated by the effects of years, umpires and player attribute/game situation factors could be used to predict whether a (called) pitch in any given location would be called a strike (or a ball), though this was not an inferential objective of the current article. Methods of analysis based on GAMs or other logistic regression models for the called strike probability can do likewise, so, in principle, it would be possible to compare the predictive performance of our modeling approach to that of those other approaches. This is a topic for future investigation.

SUPPLEMENTARY MATERIAL

Supplement A to "Outline analyses of the called strike zone in Major League Baseball" (DOI: 10.1214/19-AOAS1285SUPPA; .zip). We include in this supplementary material an animation depicting the evolution of the called strike zone from 2008–2016.

Supplement B to "Outline analyses of the called strike zone in Major League Baseball" (DOI: 10.1214/19-AOAS1285SUPPB; .zip). We include in this supplementary material `R` code for fitting elliptic Fourier models.

Supplement C to "Outline analyses of the called strike zone in Major League Baseball" (DOI: 10.1214/19-AOAS1285SUPPC; .pdf). We include in this supplementary material details of the algorithm for fitting generalized superellipses and other closed curves.

Supplement D to "Outline analyses of the called strike zone in Major League Baseball" (DOI: 10.1214/19-AOAS1285SUPPD; .zip). We include in this supplementary material an `R` package for fitting generalized superelliptical models.

Supplement E to "Outline analyses of the called strike zone in Major League Baseball" (DOI: 10.1214/19-AOAS1285SUPPE; .pdf). We include in this supplementary material displays of KDA-based and fitted ATLAS called strike zones and a list of estimated ATLAS coefficients corresponding to the 96 factor combinations described in Section 6. We also give the MANOVA table, ANOVA tables, and standard errors of weighted level means described in Section 6, and give similar results for a model that includes main effects of year and interactions of all other factors with year. Finally, we list bootstrap variances of ATLAS parameter estimates of called strike zones for selected factor combinations.

Supplement F to "Outline analyses of the called strike zone in Major League Baseball" (DOI: 10.1214/19-AOAS1285SUPPF; .pdf). We include in this supplementary material a derivation of the form of the confidence interal for the proportion of variability in geometric attributes of called strike zone outlines attributable to umpires.

REFERENCES

AHN, S. J. (2004). *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space* **3151**. Springer, Berlin.

BELONGIE, S., MALIK, J. and PUZICHA, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 509–522.

BLAND, M. (2000). *An Introduction to Medical Statistics*, 3rd ed. Oxford Univ. Press, Oxford.

BOOKSTEIN, F. L. (1997). *Morphometric Tools for Landmark Data*. Cambridge Univ. Press, Cambridge. MR1469220

BROOKS, D., PAVLIDIS, H. and JUDGE, J. (2015). Moving beyond WOWY: A mixed approach to measuring catcher framing. Available at https://www.baseballprospectus.com/news/article/25514/moving-beyond-wowy-a-mixed-approach-to-measuring-catcher-framing/.

CARRUTH, M. (2012). The strike zone. Available at https://www.lookoutlanding.com/2012/10/29/3561060/the-strike-zone.

CLAUDE, J. (2008). *Morphometrics with R*. Springer, New York.

DEL CASTILLO, E. and COLOSIMO, B. M. (2011). Statistical shape analysis of experiments for manufacturing processes. *Technometrics* **53** 1–15. MR2791943

DESHPANDE, S. K. and WYNER, A. (2017). A hierarchical Bayesian model of pitch framing. *J. Quant. Anal. Sports* **13** 95–112.

DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis*. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, Chichester. MR1646114

DUONG, T. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *J. Stat. Softw.* **21** 1–16.

FAST, M. (2011a). Spinning yarn: The real strike zone. Available at http://www.baseballprospectus.com/article.php?articleid=12965.

FAST, M. (2011b). Spinning yarn: The real strike zone, part 2. Available at http://www.baseballprospectus.com/article.php?articleid=14098.

GARDINER, M. (1965). The superellipse: A curve that lies between the ellipse and the rectangle. *Scientific American* **213** 222–234.

GREEN, E. and DANIELS, D. P. (2014). What does it take to call a strike? Three biases in umpire decision making. Available at http://www.sloansportsconference.com/wp-content/uploads/2014/02/2014_SSAC_What-Does-it-Take-to-Call-a-Strike.pdf.

GREEN, E. and DANIELS, D. P. (2015). Impact aversion in arbitrator decisions. Available at https://dx.doi.org/10.2139/ssrn.2391558.

GREEN, E. and DANIELS, D. P. (2017). Bayesian instinct. Available at https://site.stanford.edu/sites/default/files/3624-ssrn-id2916929_2.pdf.

HALL, P. and KANG, K.-H. (2005). Bandwidth choice for nonparametric classification. *Ann. Statist.* **33** 284–306. MR2157804

JÄGER, J. M. and SCHÖLLHORN, W. I. (2012). Identifying individuality and variability in team tactics by means of statistical shape analysis and multilayer perceptrons. *Human Movement Science* **31** 303–317.

KAGAN, D. (2009). The anatomy of a pitch: Doing physics with PITCHf/x data. *The Physics Teacher* **47** 412–416.

KIM, J. W. and KING, B. G. (2014). Seeing stars: Matthew effects and status bias in Major League Baseball umpiring. *Manage. Sci.* **60** 2619–2644.

KLASSEN, E., SRIVASTAVA, A., MIO, M. and JOSHI, S. H. (2004). Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 372–383.

KUHL, F. P. and GIARDINA, C. R. (1982). Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* **18** 236–258.

KURTEK, S., SRIVASTAVA, A., KLASSEN, E. and DING, Z. (2012). Statistical modeling of curves using shapes and related features. *J. Amer. Statist. Assoc.* **107** 1152–1165. MR3010902

LENTH, R. V. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics* **31** 469–473. MR1041567

LINDBERGH, B. (2015). The strike zone time travel test: How much of baseball's anemic offense really stems from the swollen zone? Available at http://grantland.com/the-triangle/mlb-larger-strike-zone-declining-offense/.

LOPEZ, M. and MILLS, B. (2018). Everyone wants to go home during extra innings—maybe even the umps. Available at https://fivethirtyeight.com/features/everyone-wants-to-go-home-during-extra-innings-maybe-even-the-umps/.

LU, T.-F. C., GRAYBILL, F. A. and BURDICK, R. K. (1987). Confidence intervals on the ratio of expected mean squares $(\theta_1 + d\theta_2)/\theta_3$. *Biometrics* **43** 535–543. MR0909757

MARCHI, M. and ALBERT, J. (2014). *Analyzing Baseball Data with R*. CRC Press, Boca Raton, FL.

MILLS, B. M. (2014). Social pressure at the plate: Inequality aversion, status, and mere exposure. *Managerial and Decision Economics* **35** 387–403.

MILLS, B. M. (2016a). Policy changes in Major League Baseball: Improved agent behavior and ancillary productivity outcomes. *Economic Inquiry* **55** 1104–1118.

MILLS, B. M. (2016b). Are the umpires at it again? Available at https://www.fangraphs.com/tht/are-the-umpires-at-it-again/.

MILLS, B. M. (2017). Technological innovations in monitoring and evaluation: Evidence of performance impacts among Major League Baseball umpires. *Labour Econ.* **46** 189–199.

MOLYNEUX, G. (2016). Prospectus feature: Umpires aren't compassionate; they're Bayesian. Available at https://www.baseballprospectus.com/news/article/28513/prospectus-feature-umpires-arent-compassionate-theyre-bayesian/.

MOSKOWITZ, T. J. and WERTHEIM, L. J. (2011). *Scorecasting*: *The Hidden Influences Behind How Sports Are Played and Games Are Won*. Random House, New York.

PARSONS, C. A., SULAEMAN, J., YATES, M. C. and HAMERMESH, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *Am. Econ. Rev.* **101** 1410–35.

PASSAN, J. (2015). Sources: MLB could alter strike zone as response to declining offense. Available at https://sports.yahoo.com/news/sources-mlb-could-alter-strike-zone-as-response-to-declining-offense-232940947.html.

ROEGELE, J. (2013a). A simple strike zone formula and calculating plate discipline stats. Available at https://www.beyondtheboxscore.com/2013/8/5/4576622/simple-strike-zone-formula-calculating-plate-discipline-stats-pitchfx-sabermetrics.

ROEGELE, J. (2013b). Investigating the "Lefty Strike." Available at https://www.beyondtheboxscore.com/2013/6/7/4391656/investigating-the-lefty-strike-pitchfx-sabermetrics.

ROEGELE, J. (2013c). The living strike zone. Available at https://www.baseballprospectus.com/news/article/21262/baseball-proguestus-the-living-strike-zone/.

ROEGELE, J. (2014a). The strike zone during the PITCHf/x era. Available at https://www.fangraphs.com/tht/the-strike-zone-during-the-pitchfx-era/.

ROEGELE, J. (2014b). The strike zone expansion is out of control. Available at https://www.fangraphs.com/tht/the-strike-zone-expansion-is-out-of-control/.

ROEGELE, J. (2015). The 2015 strike zone. Available at https://www.fangraphs.com/tht/the-2015-strike-zone/.

ROEGELE, J. (2016). The 2016 strike zone. Available at https://www.fangraphs.com/tht/the-2016-strike-zone/.

ROSIN, P. L. (2000). Fitting superellipses. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 726–732.

SIEVERT, C. (2014). Taming PITCHf/x data with XML2R and pitchRx. *R J.* **6** 5–19.

SIEVERT, C. (2015). pitchRx: Tools for harnessing MLBAM Gameday data and visualizing PITCHf/x. R package version 1.7.

SRIVASTAVA, A. and KLASSEN, E. P. (2016). *Functional and Shape Data Analysis. Springer Series in Statistics*. Springer, New York. MR3821566

SRIVASTAVA, A., JOSHI, S. H., MIO, W. and LIU, X. (2005). Statistical shape analysis: Clustering, learning, and testing. *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 590–602.

TAINSKY, S., MILLS, B. M. and WINFREE, J. A. (2015). Further examination of potential discrimination among MLB umpires. *Journal of Sports Economics* **16** 353–374.

TEUTSCH, C., BERNDT, D., TROSTMANN, E. and WEBER, M. (2013). Real-time detection of elliptic shapes for automated object recognition and object tracking. In *Proceedings of SPIE—The International Society for Optical Engineering*.

VACCA, J. R. (2007). *Biometric Technologies and Verification Systems*. Elsevier, Amsterdam.

WALSH, J. (2010). The compassionate umpire. Available at https://www.fangraphs.com/tht/the-compassionate-umpire/.

WAND, M. P. and JONES, M. C. (1994). Multivariate plug-in bandwidth selection. *Comput. Statist.* **9** 97–116. MR1280754

WASSERMAN, L. (2006). *All of Nonparametric Statistics. Springer Texts in Statistics*. Springer, New York. MR2172729

ZELDITCH, M. L., SWIDERSKI, D. L. and SHEETS, H. D. (2012). *Geometric Morphometrics for Biologists: A Primer*, 2nd ed. Elsevier, Amsterdam.

ZIMMERMAN, D. L., TANG, J. and HUANG, R. (2019). Supplement to "Outline analyses of the called strike zone in Major League Baseball." DOI:10.1214/19-AOAS1285SUPPA, DOI:10.1214/19-AOAS1285SUPPB, DOI:10.1214/19-AOAS1285SUPPC, DOI:10.1214/19-AOAS1285SUPPD, DOI:10.1214/19-AOAS1285SUPPE, DOI:10.1214/19-AOAS1285SUPPF.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242
USA
E-MAIL: dale-zimmerman@uiowa.edu
jun-tang-1@uiowa.edu
rui-huang-2@uiowa.edu