Preliminary Report of the Committee Studying Home Run Rates in MLB

Jim Albert[1], Anette (Peko) Hosoi[2],
Alan Nathan[3], and Lloyd Smith[4]

[1]Bowling Green State University
[2]MIT
[3]University of Illinois at Urbana-Champaign
[4]Washington State University

December 10, 2019

Abstract

A preliminary report is presented analyzing possible causes of the change in home run rate in Major League Baseball during the period 2017-2019 and the features of the baseball that affect its aerodynamic properties. Analysis of StatCast data shows that the increase in home run rate between 2018 and 2019 was due in part to a change in launch conditions and in part to a change in the baseball drag. The increase due to changes in launch conditions was determined to be due to a change in player behavior rather than to changes in the baseball. The increase due to changes in drag was confirmed in both the StatCast data and laboratory measurements, both of which show a significant ball-to-ball variation in the baseball drag that is large compared to the year-to-year change in the average drag. The laboratory experiments, using newly developed techniques, show a correlation between drag and seam height, with the average seam height in 2019 smaller than that in 2018 by less than 0.001 inches. Other properties of the baseball contributing to the variation in drag were not identified, although various hypotheses suggested in the media were ruled out. No evidence

was found that changes in baseball performance were due to anything intentional on the part of Rawlings or MLB and were likely due to manufacturing variability.

1 Introduction

The number of home runs per season in MLB has changed dramatically during the period 2014-2019, as shown in Fig. 1, After a steady downward trend during the period 2000-2014, home runs have increased rapidly since then, with the exception of the drop in 2018. Indeed, in 2019 a new record was set for the most home runs in a season, breaking the previous record set only two years earlier in 2017.

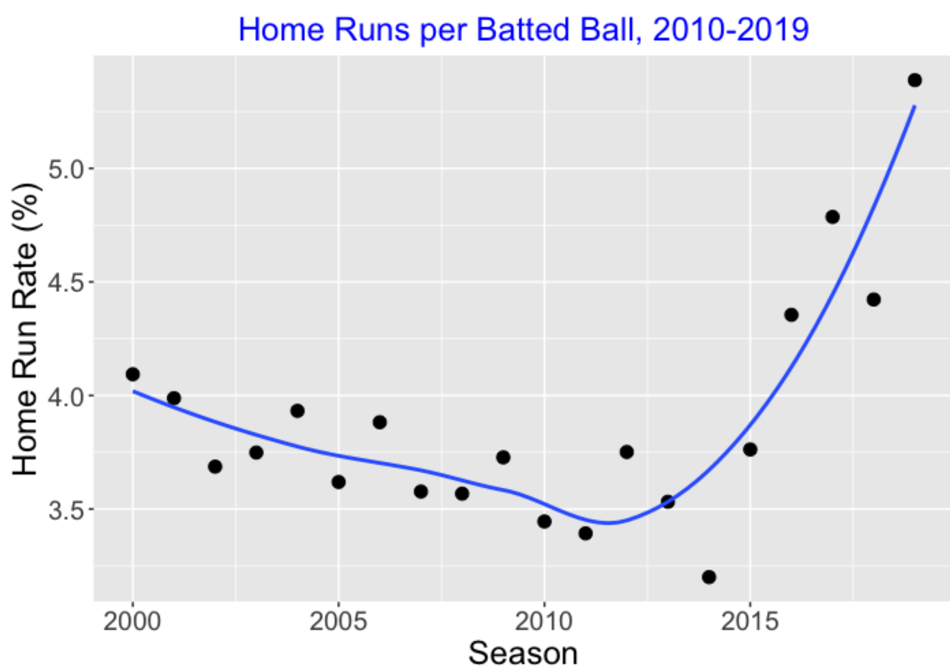

Figure 1: Yearly home runs per batted ball during the period 2000-2019 along with a smooth trendline.

In 2017 MLB commissioned an independent study to investigate the causes of the home run increase during the period 2015-2017. The results of the study were presented in a May 2018 report [1], with the following principal findings:

- The primary factor causing the increase in home runs during the 2015-2017 period was a change in the aerodynamic properties of the baseball, specifically a reduction in the average drag coefficient C_d, resulting in baseballs carrying farther for given launch parameters (e.g., exit velocity EV, launch angle LA, etc.) The increase in home runs was not due to a change in the launch conditions themselves.

- In any given year, the ball-to-ball variation in C_d is much larger than the relatively small change in the mean C_d that accounts for the change in the home run rates between seasons.

- Neither the year-to-year changes nor the significant ball-to-ball variation in C_d could be associated with any specific physical property of the ball, such as seam height, surface roughness, size, etc. Nor could these factors be explained by any changes in the manufacturing materials or processes.

In view of latter finding as well as the significant increase in home run rates during the 2019 season, members of the original committee were reconvened in 2019. The committee set its own agenda and arrived at its own conclusions and recommendations, independent of MLB. The primary goals of the committee were to seek answers to the following questions, all of which are addressed in Sec. 3:

1. What are the physical properties of the baseball that lead to both a large ball-to-ball variation and to a smaller year-to-year change in its drag properties? This question will be addressed in Sec. 3.1.

2. What are the underlying causes of the changes in home run rate during the 2018 and 2019 seasons? This question is divided into the following sub-questions and will be addressed in Sec. 3.2:

   - How much of the changes in home run rate are due to changes in the ball that affect its carry?

   - How much of the changes in home run rate are due to changes in launch parameters? Is there evidence that changes in launch parameters are due to changes in the baseball as opposed to changes in player behavior?

3. Did the home run rate change during the 2019 postseason, and if so what is driving that change? This question will be addressed in Sec. 3.3.

This report is organized as follows. A brief section on methodology is presented in Sec. 2, followed by Sec. 3, where the questions posed are specifically addressed and recommendations are given. Some of the material that supports these findings are presented in Sec. 4. A brief biosketch of the committee membership is given in Appendix A and references are given at the end. While the present report is meant to be a summary of the important findings, a more detailed report will be forthcoming.

2 Methodology

To answer the questions posed, the following testing and analysis were undertaken:

1. Using the Sports Science Laboratory at Washington State University [2], extensive laboratory testing was undertaken of 65 dozen unused Major League baseballs from the 2013 to 2019 seasons, including 20 dozen each from 2018 and 2019. The 2018 and 2019 baseballs all came from the beginning of their respective production cycles. In addition the drag coefficient and seam height of 20 dozen baseballs from the 2019 postseason were measured.

2. These investigations have been time-intensive and complex. They have involved the development of new methods and instrumentation to measure properties of the ball with improved precision, since the effects under investigation are very subtle.

   (a) The apparatus used previously to measure the drag and seam height were significantly modified, resulting in an improvement to both the precision and accuracy of the measurements. These improvements, when combined with the large sample sizes for the 2018 and 2019 seasons, allowed a statistically significant determination of differences of yearly mean values as small as 0.005 for C_d and 0.0016 inches for seam height.

   (b) Novel tests of the baseball were devised and conducted in the attempt to identify the factors that lead to both ball-to-ball and year-to-year variation in the drag properties.

3. StatCast data from the 2016-2019 seasons were analyzed.

   (a) By modeling the dependence of home run probability on launch parameters, the fraction of the year-to-year home run changes due to changes in the launch parameters and the fraction due to changes in the drag were determined.

   (b) StatCast trajectory data for fly balls were analyzed to determine C_d, including the dependence on spin and the relationship to fly ball distance.

   (c) The publicly available StatCast trajectory data for pitched baseballs (the so-called 9P parametrization) were analyzed to determine C_d, including determining the dependence on and controlling for the transverse spin [3].

4. The quality control data on Major League baseballs from Rawlings and UMass/Lowell for the 2010-2019 period were reviewed and analyzed.

3 Findings

3.1 What Baseball Properties Affect Drag?

1. The improved experimental measurements reveal a ball-to-ball correlation between C_d and seam height (Fig. 2) with $R^2 \approx 0.35$ in addition to a strong correlation between the yearly-average C_d and seam height for the four periods encompassing the years 2013-2019 (Fig. 3).

2. Other than seam height, none of the other alternate hypotheses discussed in the media (e.g., roundness, surface roughness, lace thickness) are correlated with C_d. Factors other than seam height account for roughly 65% of the ball-to-ball differences in C_d, and have been relatively constant between 2013 and 2019. An example is shown in Fig. 4

3. Given the large ball-to-ball variation and the comparatively small year-to-year changes in both C_d and seam height (Fig. 5), large sample sizes ($\approx$ 20 dozen) are needed to compare yearly changes using current test methods.

4. The aerodynamic flow over a baseball is complex [4], resulting in drag properties that depend not only on seam height but also on spin rate,

spin axis, seam orientation, application of mud (see Fig. 6), and possibly other factors not yet identified. While we have learned much from our studies as well as those of other investigators [5], there is much that is not yet understood.

3.2 What Drives the Home Run Changes?

1. The StatCast data were analyzed using the technique described by Albert [6] to separate the home run changes into two parts: a part due to changes in carry and a part due to changes in launch conditions. The following results are obtained (see Figs. 7-8 and Table 1):

    - For 2016-2017, the increase in home runs is primarily due to an increase in carry, as previously concluded in the 2018 report [1].
    - For 2017-2018, the change in home runs is due to two opposing effects: a change in launch conditions, which would have increased the number of home runs; and a decrease in carry, which would have decreased the number of home runs. The combined effect was a decrease in home runs.
    - For 2018-2019, approximately 60% of the home run increase is due to an increase in carry and 40% to a change in launch conditions. As noted above, only 35% of the increase in home run rate attributable to greater carry is due to a change in the seam height.

2. The part of the home run changes due to carry (both the decrease in 2018 and the increases in 2017 and 2019) is consistent with changes in C_d values determined from multiple sources, including the analysis of StatCast fly ball and pitched ball trajectories (Fig. 9). These measurements indicate a decrease in C_d in the 2016-2017 transition, an increase in the 2017-2018 transition, and another decrease in the 2018-2019 transition.

3. The laboratory C_d shows a somewhat smaller gap in the 2018-2019 transition than those from StatCast, perhaps due to the fact that the laboratory baseballs were sampled from very early in each production season whereas the StatCast baseballs were sampled throughout the entire season. Moreover, the pattern of changes during the period

2016-2018 from laboratory measurements (Fig. 3), with much smaller samples sizes in 2016 and 2017, do not follow the pattern from StatCast data (Fig. 9). These discrepancies are puzzling and perhaps point to the need for larger sample sizes, such as those available for 2018-2019.

4. Both C_d values and fly ball distances show considerable variation about their mean values (Fig. 10), qualitatively consistent with the ball-to-ball variation in drag found in laboratory experiments. However, for given exit velocity and launch angle, the variation of fly ball distance with rate of backspin and sidespin on the batted ball is much larger than the variation due to C_d (Table 2). Whether year-to-year changes in those spin rates play a role in the changes in home run rates is currently under investigation.

5. Some of the home run increase is due to the changes in launch conditions.

    - Not only is there a general increase in home run hitting, but the rates of hitting home runs has shown a general increase across players of all slugging abilities (see Fig. 11).
    - The parameters contributing to this change are exit velocities, launch angles (see Fig. 12), and spray angles.
    - From analysis of the Rawlings and UMass/Lowell test data (Fig. 13), there is some evidence of a small increase in the CCOR of the baseball between 2018 and 2019. However, that increase was too small to play a significant role in the home run increase in 2019.
    - Lacking strong evidence that the change in launch conditions are due to changes in the baseball, we conclude that they are due to a change in player behavior.

3.3 Did the Home Run Rate Change in the 2019 Postseason?

1. It is the understanding of the committee that Rawlings uses the same manufacturing process to create the baseball used in the postseason as they do to create the ball used in the regular season, save for the application of the postseason stamp. There would therefore be no reason to

suspect a change in the performance properties of the baseball between the regular and postseason.

2. The analysis of StatCast pitched ball trajectories indicates an increase in C_d in the 2019 postseason (Fig. 14), leading to a decrease in both fly ball distances (Fig. 15) and home runs (Fig 16). While the analysis controlled for home field, the sample size was far too small to control for players in any meaningful way.

3. The laboratory testing showed a comparable increase in C_d but no change in seam height for the 2019 postseason baseballs compared to the 2019 regular season baseballs. Therefore, the reason for the change in C_d for the postseason baseballs is not known.

4. It should be noted that there are far fewer games played in the postseason than during any given week of the regular season. For example, there were 37 postseason games in the 2019 postseason, compared to typically ~ 90 in any week of the regular season.

3.4 Recommendations

Based on its investigations, the committee makes the following recommendations:

1. Rawlings should develop a system to track the dates on which balls are manufactured and shipped to clubs. Clubs should log which batches of baseballs are used in which games or homestands.

2. To facilitate determination of drag and other properties affecting performance from in-game data, MLB should install atmospheric tracking systems at field level in all 30 parks, including temperature, pressure, relative humidity, and wind conditions.

3. Since changes in drag play a major role in driving changes in home runs, MLB should codify the current procedures used to monitor the drag, whether in the laboratory or with in-game data, sampling baseballs manufactured throughout the production cycle.

4. Similarly, the monitoring of other baseball properties (especially the COR and CCOR) at UMass/Lowell, which is currently being done

three times each season, should be expanded to sample baseballs manufactured throughout the production cycle.

5. In view of the apparent dependence of drag on the applied mud based on measurements of a small sample, a more extensive study should be performed with much larger sample.

6. MLB should study the viability of employing humidors in all 30 parks to reduce the variability in storage conditions across the league.

3.5 Summary

The laboratory data show a correlation between seam height and drag. They further show a large ball-to-ball variation in those quantitites and much smaller year-to-year changes in their averages. Indeed, the change in mean seam height between 2018 and 2019 is less than one-thousandth of an inch. Statcast data show that a significant part of the home run increase in 2019 was due to a reduction in the drag properties of the baseball, although some was due to a change in launch parameters. The Rawlings quality control data show a slight increase in the CCOR for 2019, but that increase did not play an important role in the home run increase. The committee found no evidence in either this study or the May 2018 report that the changes in baseball performance were due to anything intentional on the part of Rawlings or MLB and were likely due to normal manufacturing variability.

4 Supporting Material

4.1 Laboratory Analysis

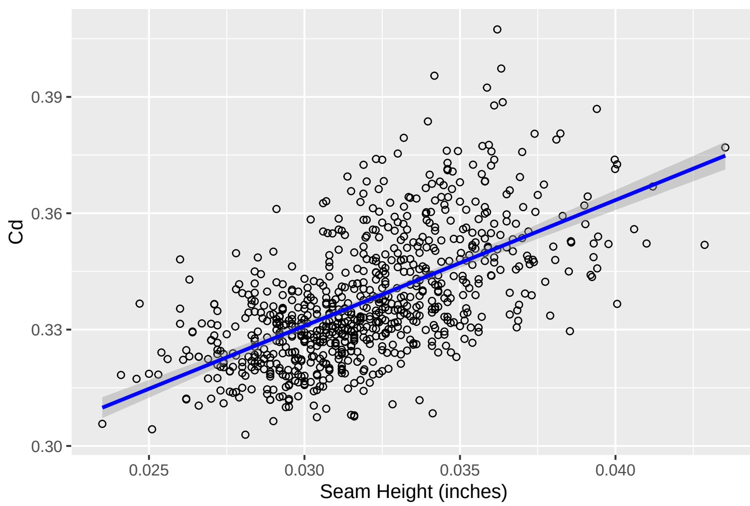

Figure 2: C_d as a function of seam height for 65 dozen baseballs from the period 2013-2019, along with a linear regression fit and error band ($R^2 \approx 0.35$)

.

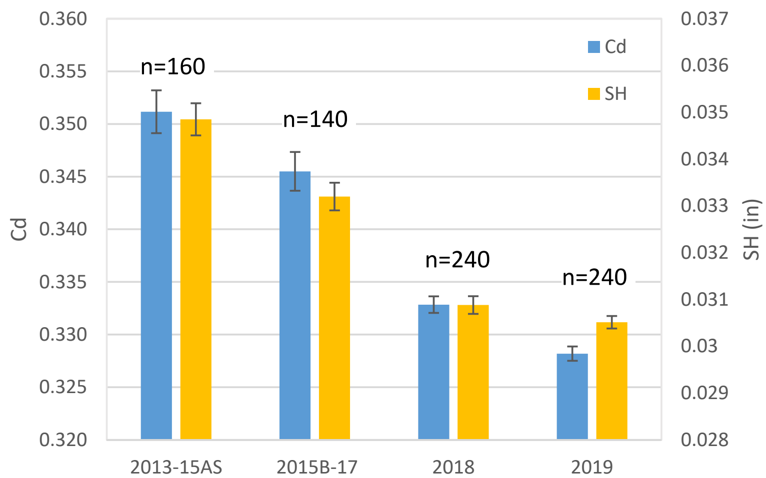

Figure 3: Yearly averages of C_d vs. seam height, with standard errors indicated by the bars, for the four periods encompassing the years 2013-2019, where 2015AS and 2015B refer to pre- and post-All Star Game in 2015, respectively. The sample sizes are indicated.

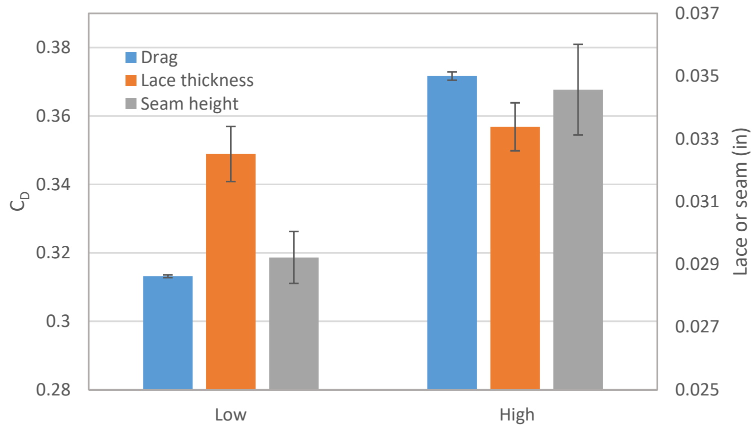

Figure 4: Comparision of C_d, seam height, and lace thickness for a group of one dozen each of high-drag and low-drag baseballs. These data show a significant difference between the two groups in seam height but not in lace thickness.

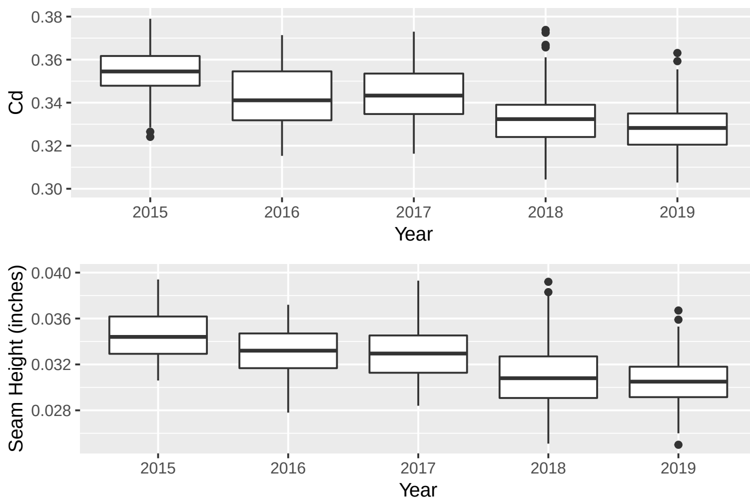

Figure 5: Box plot of the distribution of yearly averages of C_d (top) and seam height (bottom), showing that the year-to-year variation of mean values are significantly smaller than the variation within any given year. The sample size was 20 dozen for both 2018 and 2019 and 5 dozen for each of the earlier years.

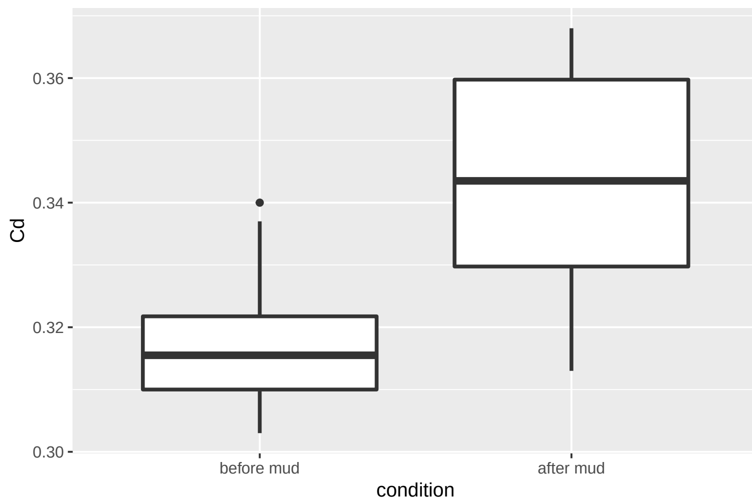

Figure 6: Boxplot of C_d measurements on a one-dozen sample of new baseballs before and after the application of mud. Even with this limited sample, the data show a significant increase in both the mean and the variation in C_d when the mud is applied.

4.2 StatCast Analysis

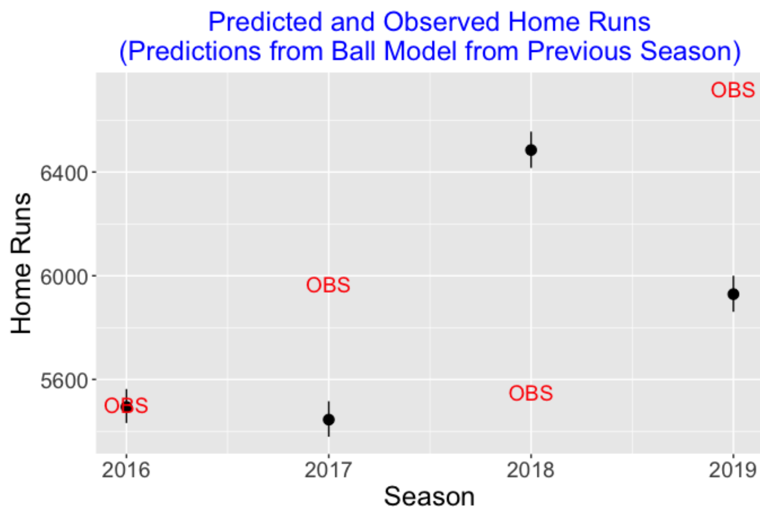

Figure 7: Predicted and observed home run counts for each season. The black bar represents a 90% confidence interval for the count using a model trained with data from the previous season and the red label shows the location of the actual home run count. The model was based on exit velocity, launch angle, and spray angle. Including additional variables to the model, such as park, do not alter the conclusions in any substantial way. The differences between the observed and predicted values are assumed to be due to a change in the carry of the ball, presumably due to a change in drag. For example, a year in which the number of observed home runs (shown in red) is larger than the predicted number (shown in black) would indicate that the drag coefficient of the baseball decreased from the previous season.

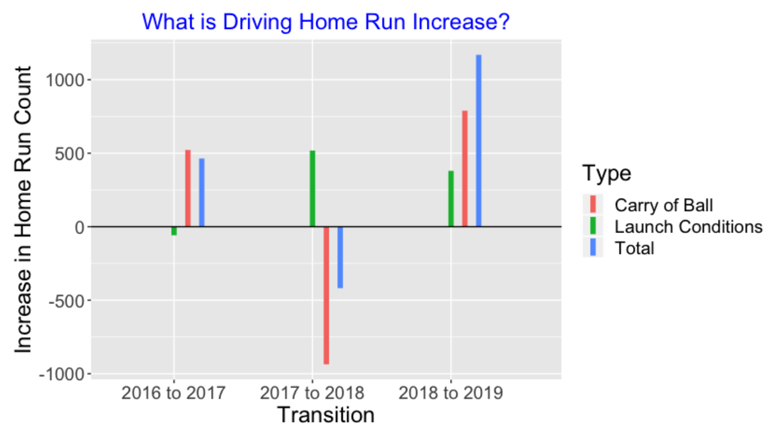

Figure 8: Change in the total number of home runs, the number attributed to the carry of the ball, and the number due to the launch conditions for the seasons 2017, 2018, and 2019. These results are summarized in Table 1.

Table 1: Total increase in home runs, increase due to carry, and increase due to changes in launch conditions, for the three yearly transitions.

Transition	Total Increase in Home Runs	Increase Due to Carry	Increase Due to Launch Conditions
2016 to 2017	464	522	-58
2017 to 2018	-418	-936	518
2018 to 2019	1169	789	380

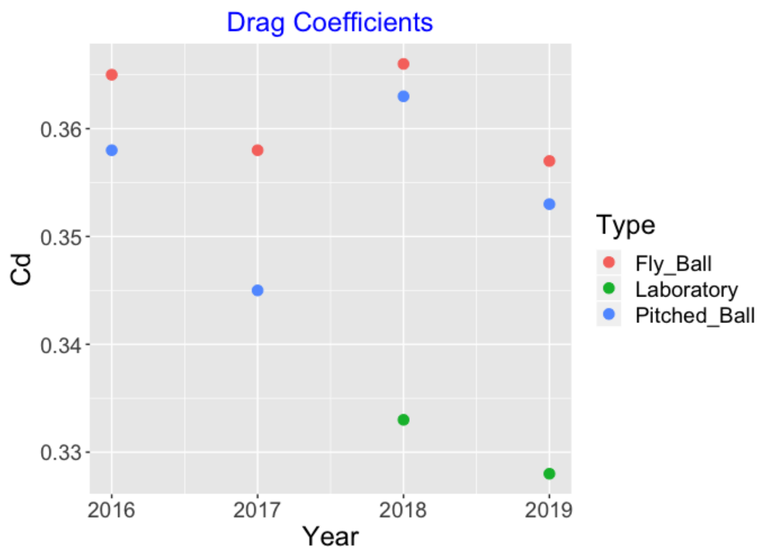

Figure 9: Yearly mean drag coefficients from analysis of StatCast fly ball and pitched ball data and from laboratory measurements. The latter only include data from 2018 and 2019, where a sufficiently large sample size was available. The StatCast data had a transverse spin around 2500 rpm, and the pitch data had a release velocity 92-95 mph. The difference between the laboratory and StatCast data might be due to the application of mud to the latter. Nevertheless, all data show the same yearly trends.

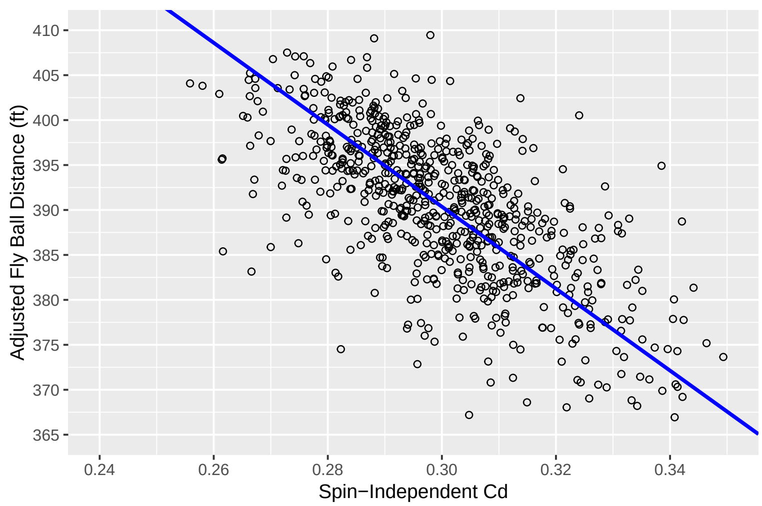

Figure 10: Relationship between fly ball distances and the spin-independent drag coefficient. The distances are for 100 mph exit velocities and launch angles 25°-30° and have been controlled for air density, spin rate, and spin axis. The plot shows that the adjusted distances are inversely proportional to the spin-independent part of the drag coefficient, as expected. The total spread in distances, ∼8 ft rms, is due primarily to the ball-to-ball variation in C_d (∼6 ft), with the remainder due to experimental noise.

Table 2: Root-mean-square(rms) scatter of fly ball distances for the 2019 season, for covered stadiums. All distances have been controlled for air density and are for launch conditions with exit velocity 100 mph and launch angles 25°-30°. Column 1 is the total rms scatter and columns 2-4 are the contributions of spin, drag, and noise, respectively, with $\sigma_T^2 = \sigma_{spin}^2 + \sigma_{drag}^2 + \sigma_{noise}^2$. These results show that the contribution of spin far outweighs the contribution of drag.

σ_T	σ_{spin}	σ_{drag}	σ_{noise}
17.1	15.1	6.1	5.2

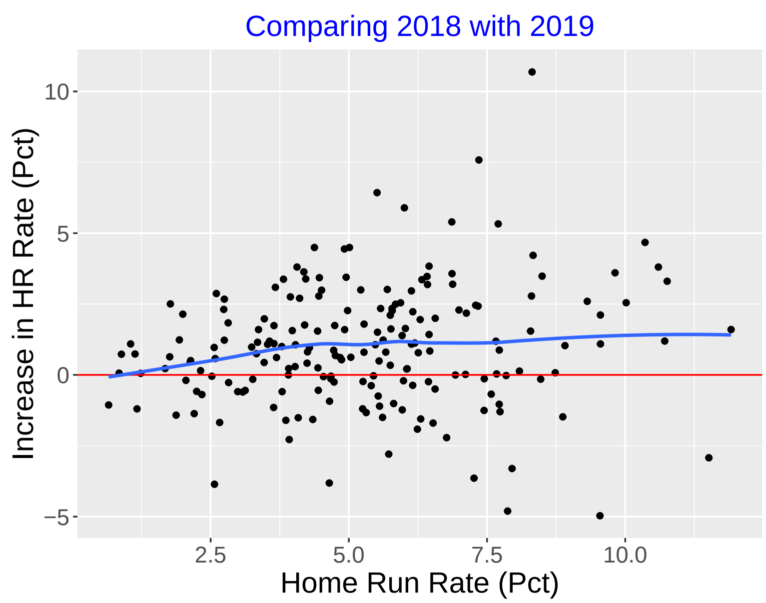

Figure 11: Scatterplot of 2018-to-2019 change in home run rate vs. average home run rate for the two seasons, the latter being a measure of slugging ability. Each point represents a single batter with at least 200 batted balls in each of the two seasons. The blue curve is a trend line. These data show that the increase in home run rate was widespread across batters of different slugging ability.

Figure 12: Scatterplot of 2018-to-2019 change in mean exit velocity vs. change in mean launch angle. Each point represents a single batter with at least 200 batted balls in each of the two seasons. The size of each point indicates the magnitude of the change in home runs per batted ball (in percent) and the color indicates whether the rate increased (blue) or decreased (red). The largest increase in home run rate appears in the upper right quadrant, corresponding to an increase in both launch angle and exit velocity. Conversely, the largest decrease occurs in the lower left quadrant, corresponding to a decrease in both launch angle and exit velocity.

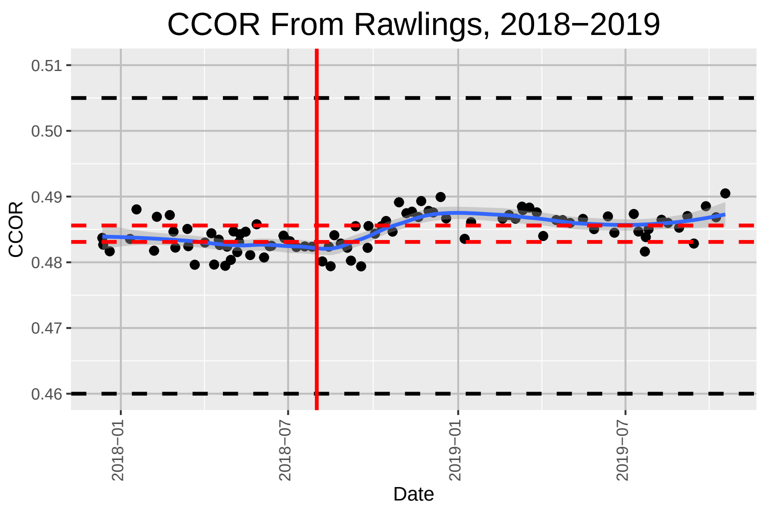

Figure 13: Test data over time from Rawlings on the high-speed CCOR of the baseball for the period 2018-2019. Each point represents the mean of 6 or 12 baseballs. Although there is considerable scatter to the data, there is a definite upward trend over time, as indicated by the smoothing curve and error band. The vertical red line marks the separation between balls tested for use in the 2018 and 2019 seasons. The lower and upper dashed lines are the mean CCOR for 2018 and 2019 balls, respectively, and suggest an increase in the CCOR across the two seasons of approximately 0.0025, which is well within the MLB specifications indicated by the dashed black horizontal lines. This increase in CCOR corresponds to an increase in exit velocity of about 0.3 mph, an increase in fly ball distance by about 1.5 ft, and an increase in home run probability by less than 4%. (see Tables 7-8 in [1]).

4.3 Postseason Analysis

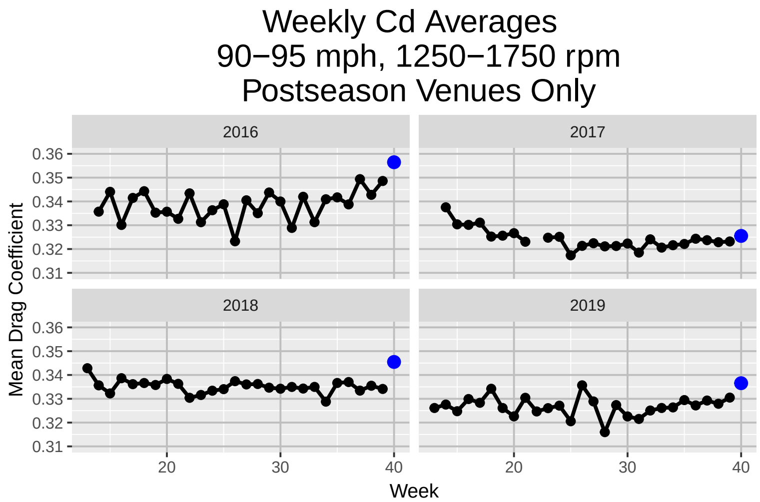

Figure 14: Weekly average of drag coefficients from StatCast pitched ball data, all fastballs with release velocity 90-95 mph and transverse spin 1250-1750 rpm. The black and blue points are for the regular season and postseason, respectively, with the latter all lumped into week 40. The regular season data only include home venues that coincided with the postseason for that year. There is a small uptick in C_d in the 2019 postseason.

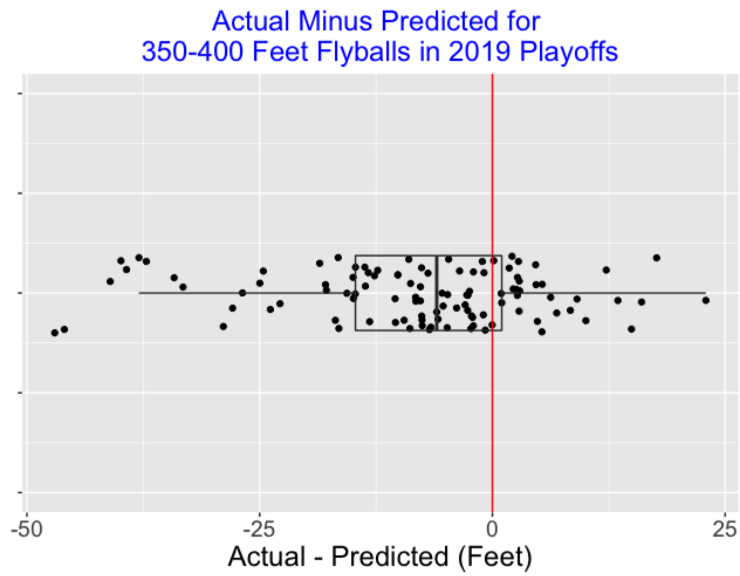

Figure 15: Actual minus predicted fly ball distance for the 2019 postseason. The predicted value is based on a model determined from the 2019 regular season, including the effect of exit velocity, launch angle, home field, and tempeartures. The dots are individual batted ball hits in the distance range 350-400 ft. The median of ∼-6 ft shows that fly balls had less carry in the postseason compared to the regular season, consistent with the increase in drag and the decrease in home runs.

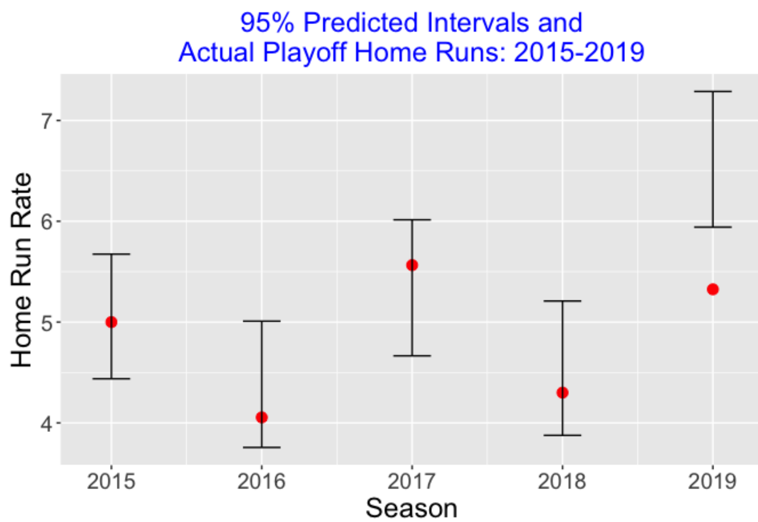

Figure 16: Predicted and observed home run counts for each season of playoffs. The black bar represents a 95% prediction interval for the home run rate using model determined from the regular season, including the effect of temperature, and the red dot shows the location of the actual home run rate.

Appendix A The Committee

- **Jim Albert** is Distinguished University Professor at Bowling Green State University. He has written many published papers and authored four books on the interface of statistical thinking and baseball. He is active in the Section on Statistics and Sports in the American Statistical Association and is former editor of the Journal of Quantitative Analysis of Sports. He contributes regularly on the blog `https://baseballwithr.wordpress.com` that illustrates the use of the statistical system R in exploring baseball data.

- **Anette (Peko) Hosoi** is the Neil and Jane Pappalardo Professor of Mechanical Engineering and Professor of Mathematics at MIT. She is a Fellow of the American Physical Society (APS) and co-founder of the MIT Sports Lab. Her research interests include fluid mechanics and biomechanics, particularly the intersection of engineering, applied mathematics, and athletic performance.

- **Alan Nathan (chair)** is Professor Emeritus of Physics at the University of Illinois at Urbana-Champaign and a Fellow of the APS. After a successful career doing experimental nuclear/particle physics, he has spent much of the past two decades doing research in various aspects of the physics of baseball. He has written numerous articles, both for academic journals and for the popular media, and runs an oft-visited website devoted to the topic, `http://baseball.physics.illinois.edu`.

- **Lloyd Smith** is a Professor in the School of Mechanical and Materials Engineering at Washington State University and is a Fellow of the ASME and ISEA. He is the director of the Sports Science Laboratory (`https://ssl.wsu.edu/`), which specializes in measuring and modeling equipment performance and personal protection.

References

[1] Jim Albert, *et al.*, "Report of the Committee Studying Home Run Rates in Major League Baseball", May 2018 https://atmlb.com/35xv40i.

[2] https://ssl.wsu.edu/

[3] Alan Nathan, "All Spin Is Not Alike," Baseball Prospectus, March 31, 2015, http://www.baseballprospectus.com/article.php?articleid=25915

[4] F. Alam, H. Ho, L. Smith, A. Subic, H. Chowdhury, and A. Kumar, "A study of baseball and softball aerodynamics," Procedia Eng., **34**, 8691 (2012).

[5] Bart Smith, Baseball Aerodynamics web site, http://www.baseballaero.com.

[6] Jim Albert, "What is Causing the Home Run Explosion," August 19, 2019: http://bit.ly/2XSnoTx.