A Novel Method for Calculating Player-Level Park Effects

Johnny Nienstedt
(Dated: September 14, 2024)

In baseball, the term 'park effect' refers to some quality of a ballpark which meaningfully changes the function of the game. Some park effects are well understood; we know that in the thin air of Colorado, balls fly roughly 9% farther than they do at sea level. We also know that Tropicana field enhances the movement of pitches due to a combination of humidity and salt in the air. Calculating these effects is straightforward, since the properties of a ball in flight are dependent only on the environment. But what about metrics like release height, extension, or bat speed? These are properties of the player alone, and so it is not so easy to isolate the effects of the park. This research introduces a novel method for calculating player-level park effects which aims to remove player bias by use of aggregated pairwise analysis. This model is extremely effective at quantifying player-level effects (i.e. pitch metrics, swing metrics), but may be vulnerable to selection bias for interaction statistics.

INTRODUCTION

One of the great and unique joys of baseball is the inconsistency of the playing field: all 30 parks have different dimensions, providing such lovable quirks as the Green Monster at Fenway, the short porch at Yankee Stadium, and many others (RIP Tal's Hill - gone but never forgotten). Disentangling these effects has long been a goal of analysts wishing to better understand team and player performance. Should the Rockies stock up on groundball pitchers to mitigate the effects of playing on the moon? Will Anthony Santander fare better from the right side if he leaves Walltimore in free agency?

However, it can be difficult to separate the trends caused by each ballpark from the talents of the players themselves. Consider this: in 2022, LoanDepot Park in Miami boasted a fairly high 23.4% strikeout rate. In 2023, that rate fell nearly all the way back to average, at 22.6%. So, is it a strikeout park or not? Well, part of that change can be attributed to one man: Luis Arraez. Acquired by the Marlins prior to the 2023 season, Arraez struck out just ten times at home in 318 plate appearances - a 3.1% clip. Such an outlier drags down the overall numbers significantly.

You may have already thought of a way around this conundrum. If one considers visiting players only, this kind of bias can be avoided, since no single player can have such an outsized effect. Unfortunately, there are still multiple ways for bias to seep in. First, visiting players are *not* chosen at random; divisional rivals see a lot more time at each other's parks compared to inter-division or interleague foes (especially prior to 2023). Secondly, the talents of the home pitching staff have not been accounted for. These confounding variables are enough to invalidate this method.

Clearly, in order to truly isolate park effects from player effects, one must find a park-neutral baseline to compare against. Doing so, however, is far from simple. A naive baseline choice could be a player average for the stat one wishes to measure. However, if half of their plate appearances come at a single park, this will hardly be a neutral measurement. The next step would be to find averages at each park the hitter has played at, and then average each of those averages. I believe this is the method that MLB uses to calculate their park factors; it's mostly solid, but still not ideal. Unless the batter has played at all 30 parks (and played enough to provide a reasonable cache of data), his baseline will still be skewed by which parks he has and has not visited. We can do better.

METHODS

The solution comes most easily (as it often does) when one considers the simplest case: two parks only. Let's use the Oakland Coliseum and Angel Stadium.

Pitch-Level Effects

Up until this point I have been using strikeout rate as an intuitive example, but this explanation demands a player-level metric, so I will switch to the original inspiration for this research: release height, the height from which the pitcher releases the ball. This is generally an innate quality of each pitcher's delivery, so any significant changes can be attributed to park effects - namely the height and structure of the mound.

So, let's consider pitchers who have thrown at least 50 pitches at both of our parks since 2021. That's a rather large set of pitchers. For each

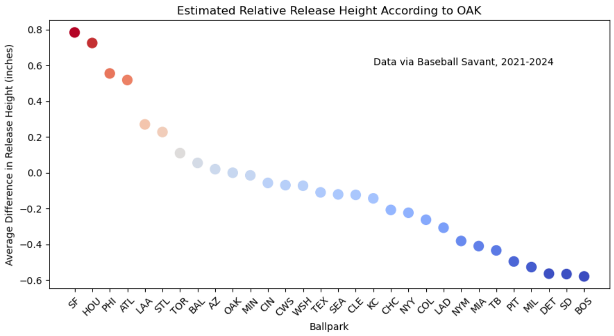


FIG. 1: *One iteration of pairwise analysis, in which each ballpark is compared directly to Oakland. Oakland was chosen for its middling effect on release height.*

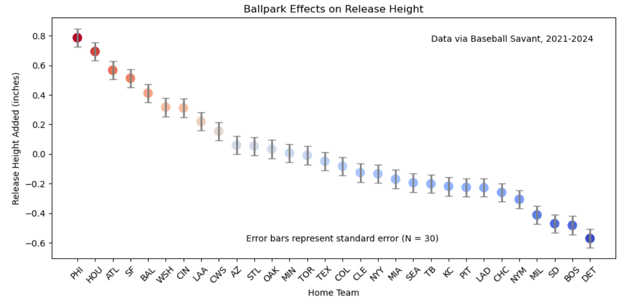


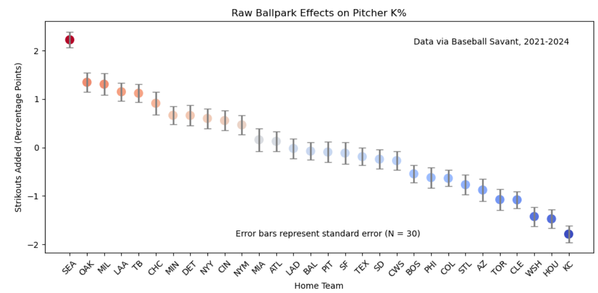
FIG. 2: *Full pairwise analysis of release height effects. Citizens Bank Park inflates release height the most, while Comerica Park depresses it. The range is 1.4".*

pitcher, let's compute his average release height at each park, and take the difference between the two.

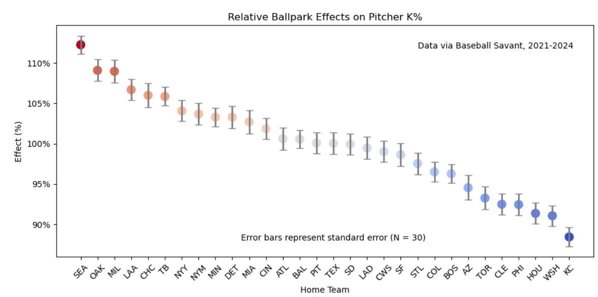
Now we can average the pitchers' differences and come up with a fairly confident estimate for the actual difference in release height between these two parks. Since each pitcher's baseline is formed solely from events at these parks, there is no bias.

Then, let's repeat the process with Oakland and Seattle, Oakland and Houston, and so on for the rest of the non-Oakland ballparks. Then we could have a fairly good idea of each park's effect on release height relative to Oakland's (see FIG. 1).

To get a better estimate, we can repeat this process for the other 29 parks, totalling 435 ballpark pairs, to come up with 30 distributions of relative mound height. Each of those distributions have the host park at zero, and the other parks grouped around it. Then all 30 distributions can be averaged directly, leaving me with what I believe is a truly bias-free measurement of (relative) release heights at each ballpark (see FIG. 2).



(a) *Raw pairwise analysis of strikeout rate effects. A pitcher is expected to increase his strikeout rate by just over 2% pitching in Seattle vs. Atlanta.*



(b) *Relative pairwise analysis of strikeout rate effects. A pitcher with a 25% K-rate in Atlanta would be expected to strike out 28% of batters in Seattle.*

FIG. 3

Relative Effects

However, most such statistics are best analyzed in terms of a relative change, rather than absolute. The simplest method for measuring such effects is to use a percentage difference. For something like strikeout rate, this is ideal: strikeout pitchers are more affected by their park's strikeout tendencies than contact pitchers. (see FIGs. 3a and 3b).

One can confirm the above hypothesis by considering the range of each distribution. The raw model displays a range of ±2 percentage points. Using the league average 22.5% K-rate, the relative model has a range of about ±2.5 percentage points. Thus the raw model undersells the magnitude of park effects on strikeout rate.

It is important to note, however, that the above analysis is incomplete; notice that the charts are labeled Pitcher K%, not just K%. Since the current iteration of my model is unable to quantify pitcher-batter interactions, I have opted to treat K% as a pitcher-level stat, and neglect the fact that batters influence strikeouts as well. I chose to use pitchers since they tend to exert slightly more control over K% than hitters (which, incidentally, is why strikeout pitching is so vital in the playoffs).

On the bright side, error is still relatively low, which means that the parks generally agree with each other about their relative strikeout inducing capabilities. See 'Discussion' for more.

Metrics Centered Near Zero

There are two more limitations to the method as described above, both of which can be exposed by considering the metric induced vertical break (iVB). Fastballs, with lift created by backspin, have positive iVB; curveballs, with topspin, generally have negative iVB.

The first problem is that iVB is not truly a player-level metric, like release height or K%, which ought to be stable for each player; iVB varies between pitch types as well. This is fairly easy to overcome, since we can essentially break each pitcher down into multiple pitchers, each with just one pitch type. The second problem is more tricky, and requires a real-world example.

Logan Gilbert, a pitcher for the Seattle Mariners, boasts a devastating slider. He throws the pitch with almost pure gyro spin - that is, if the ball was sitting on the ground, it would roll directly to the right (from his POV) rather than forward, like a curveball, or backward, like a fastball. Because of this spin, there is very little induced vertical movement on the pitch - it averages exaclty 0" iVB.

Let's say Gilbert pitches in Tampa Bay, and his slider averages -1" iVB. How should we quantify that change? Clearly a percentage difference won't work, and raw isn't ideal either; his big-breaking curveball will probably be much more effected than the slider, and we want to capture them equally[1].

My solution was to measure differences based on standard deviation. If Gilbert's slider averages 0" iVB with a standard deviation of 1", then Tampa Bay would be assigned a park effect of -1 standard deviation for that pitch. Then the curveball, which might have an average iVB of -4" and standard deviation of 2", can be judged on the same scale. See FIG. 4 for pairwise analysis of positive iVB.

[1] This is actually a very poor way to measure iVB effects, since spin axis plays such a huge role in pitch break. But that is not the focus of this discussion, so we'll let it slide.

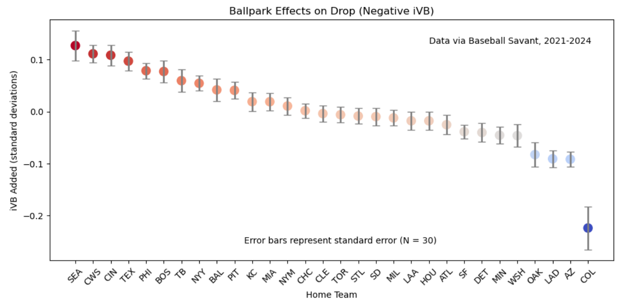


FIG. 4: *Pairwise analysis of negative induced vertical break (i.e. breaking balls only). Park effects are measured in terms of standard deviations from the mean, to avoid low frequency noise. Here a positive value represents more break.*

This isn't an ideal method - standard deviations are very player dependent, and it's far less intuitive to understand the true magnitude of the effect - but it's the best way I've found to measure park effects of metrics centered near zero[2].

ANALYSIS

You've now seen several examples of park effects calculated by pairwise analysis, but now it's time to examine their impact on the game itself. The park effect effects?

Let's start with release height. On first glance it seems like a fairly useless statistic; wherever the pitcher releases the ball from, he still has to throw it over the plate. However, in recent years we've learned about the importance of a pitch's vertical approach angle (VAA) in determining its effectiveness - specifically its ability to induce swings and misses.

So what does this have to do with release height? Well, if a pitcher releases the ball from a higher slot, he naturally has to throw it at a steeper downward angle to reach the strike zone. Some quick geometry tells us that our range of 1.4" in release height should correspond to just over a tenth of a degree in VAA. Astonishingly, even such a seemingly miniscule effect is borne out in the data (See FIG. 5). And the range is right on the money as well!

[2] The other solution I found was to filter out iVBs with magnitude less than 3". While this had the desired effect, it may not be possible for other such statistics.

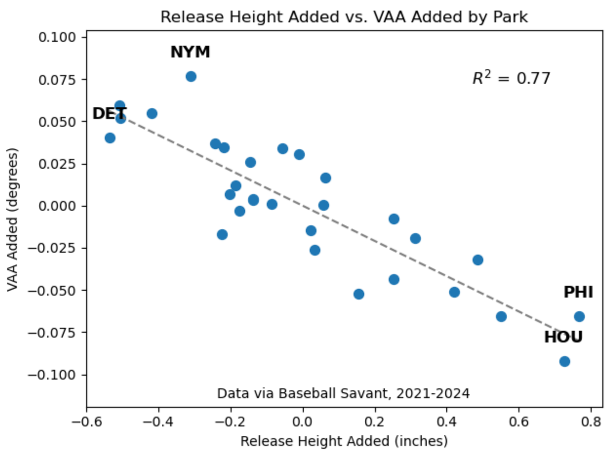


FIG. 5: *Parks with higher mounds generally decrease a pitcher's VAA. This supports the argument that mound height is an important park effect.*

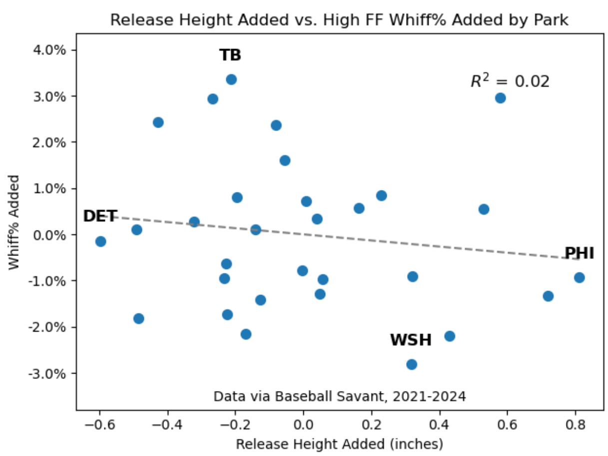


FIG. 6: *Mound height has no direct relationship with park effects on high fastball whiff rate. Effects of velocity, iVB, and deception probably outweigh the relatively small changes in VAA due to release height. Even in ideal conditions, a 0.1° decrease in VAA only corresponds to a 0.7 percentage point boost in Whiff%.*

But still, to anyone outside of pitching analytics, VAA itself is an obscure and irrelevant statistic. Can we find a relationship with actual results? If we can, it will probably be by isolating the poster boy of VAA: the high fastball. Research has shown that the flatter the angle a fastball comes in on, the more likely batters will be to swing and miss.

VAA alone can only explain 19% of the variance in Whiff% for high fastballs, so that's the absolute ceiling. Unfortunately, going back to mound height seems to be one level too far to see any direct relationship (See FIG. 6).

Interestingly, while the park effects of release height and overall VAA share a robust 0.77 R^2 correlation, release height only correlates at $R^2 = 0.46$ with high fastball VAA. This implies that high fastballs are less subject to the whims of mound height than other pitches, a conclusion which I did not expect.

DISCUSSION

While this method seems to be quite effective at quantifying player-level statistics, the goal of lowering bias for interaction statistics has not yet been met, and its application to actual results is lacking.

I am confident that more research will reveal significant relationships between release height park effects and pitch results, so I'll focus on the other issue. While we can learn a lot from player-level metrics, we can learn a lot more from interaction statistics; therefore, I will continue to test new methods until I achieve this goal.

It seems clear that we must first find both a pitcher and hitter baseline for the statistic in question, and compute a tandem baseline for every matchup. Unfortunately, that is much easier said than done; strikeout rate (along with most result-based statistics) simply does not stabilize quickly enough to get meaningful data from individual players at individual parks. I am currently working to solve this conundrum, and I hope to update this report when I have.