Pitching performance in baseball is notoriously unpredictable, making it difficult to accurately evaluate the performance of pitchers who do not throw a large number of innings. To address this challenge, I have developed a new metric known as a "stuff model" that uses detailed pitch-related data to evaluate the performance of individual pitches. This metric takes into account the interactions between pitches, known as "tunneling," to better predict a pitcher's overall performance. In this article, we will describe the construction of this stuff model and discuss its potential benefits over existing metrics for evaluating pitcher performance in small samples.

Pitching performance has historically been highly unstable in small sample sizes. There are 89 pitchers who threw 50-100 innings, a typical reliever workload, in both 2021 and 2022. The year over year correlation of ERA in this sample is miniscule with a p-value of .44. There is no statistical significance in the association. To combat this lack of “stickiness”, analysts have created metrics on the ERA scale that take more than just earned runs into account like xERA, FIP and xFIP. Unfortunately xFIP, which uses strikeout rate, walk rate, and flyball rate to create a ERA like metric also have a p-value of over .05, and thus is not statistically significant year over year. While xERA, which uses xwOBA to estimate ERA, and FIP, which uses strikeout rate, walk rate, and home run rate, are both have a statistically significant association with 2022 ERA, the correlation is miniscule, with r squared values of .063 and .069 respectively. Therefore to evaluate pitchers who don’t throw 200 innings a season, a new metric that stabilizes in lower sample sizes is necessary.

“Stuff” models have become a typical way of evaluating pitcher performance in very small sample sizes. Using individual pitch data, we can evaluate every pitch on its own, creating a large sample of events even in a relatively small sample of baseball. A popular model that I took a lot of inspiration from is pitchingbot http://pitchingapp.pitchingbot.com/ made by Cameron Grove.

My stuff model uses run value as the response variable. You need some way to evaluate the impact of a pitch relative to other pitches, so I created weighted run value (wRV) by finding the difference a ball or a strike made to the eventual woba of a plate appearance on average by count. I then subtracted the mean of each counts’ wRV from itself so that the wRV values would be count independent. For balls in play, the wRV of that pitch is equal to the xwOBA of that ball in play.

Independent variables:

Pitch velocity or “release_speed”

IVB (induced vertical break) or “pfx_z”

HB (horizontal break) or “pfx_x”

Release point horizontal “X” or “release_pos_x”

Release height “Z” or “release_pos_z”

Extension “Y” or “release_extension”

“SSW” or seam-shifted-wake

“d” or average distance from the primary pitch

“dvelo” or velocity difference from the primary pitch

“dmovex” or difference in the horizontal movement from the primary pitch

“dmovez” or difference in the vertical movement from the primary pitch

Seam shifted wake (SSW) is not available in the baseball savant data I scraped using baseballr, thus I had to make it using the given information. If we define SSW as the difference between the observed and inferred spin axis, we can get an exact value. The baseballr package has spin axis, so we just need to make an inferred spin axis based on movement. The baseball savant csv documentation (https://baseballsavant.mlb.com/csv-docs) defines spin axis as “The Spin Axis in the 2D X-Z plane in degrees from 0 to 360, such that 180 represents a pure backspin fastball and 0 degrees represents a pure topspin (12-6) curveball.” In the 2d X-Z plane we can infer the spin axis by taking the inverse tangent of the quotient of the IVB divided by the HB. Now if the HB is negative we add 90 degrees and if it’s positive we add 270. Finally the difference between inferred and observed spin axis can be calculated by the absolute value of the difference between the observed and inferred spin axis if this number is less than 180 degrees, or 360 minus the absolute value of the difference between the observed and inferred spin axis if that number is greater than 180 degrees.

Pitches do not exist in a vacuum. They have meaningful interactions with each other, thus it would be wrong to only examine the data from a pitch type to conclude how effective that pitch type will be. These relationships between pitches are often conceptualized as “tunnels”, where 2 pitches start in the same tunnel, then at the last second break in different directions. If we stick to

this definition of pitch interactions, we can define how well a pitch tunnels based on a few factors: distance between the pitches in the initial flight of the ball, difference in ultimate vertical movement, difference in ultimate horizontal movement, and difference in velocity. While secondary tunnels matter, tunnels with a pitcher's primary pitch are much more important, as that's what we'd assume a batter would be looking for especially early in counts. For this project, we'll define a primary pitch as a pitcher's most thrown pitch that's over 86 mph to get rid of the slow curveballs which a few starters like Adam Wainwright throw more than any fastball. The difference in vertical and horizontal break is quite easy to calculate between 2 pitches, as is the difference in velocity, however it is more difficult to figure out the difference in the initial flight of the ball with the available data. I first found the idea of approximating ball trajectories after reading an article by Walter Limm
(https://wlimm.wordpress.com/2022/08/24/a-tunneling-sampler/)

I realized that if you assume that in the y direction, the jerk, or derivative of acceleration, is 0, which isn't true because the air resistance would likely become slightly less strong as the ball decelerated towards the plate, but close enough because the flight of the ball is so brief, you can get a very solid estimation of where the ball is at any given time. Baseball Savant, and thus baseballr, provides the release velocity of the pitch, the release extension of the pitch, then the acceleration and velocity in the x,y, and z directions at y=50ft. We can estimate the time it will take for the ball to get to the "commit point", which Walter Limm defines as "the time at which a batter has to commit to swinging or taking." This is 167 milliseconds prior to the ball crossing the plate. We can approximate this time by the equation

$$commitTime = -.167 + \frac{-velo(5280/3600) + \sqrt{(velo(5280/3600))^2 + 2ay(60.5 - extension)}}{ay}$$

where velo is the release velocity, extension is the release extension, and ay is the acceleration in the y directions at y=50. Using a similar equation we can approximate the time it takes the ball to reach y=50:

$$50Time = \frac{-velo(5280/3600) + \sqrt{(velo(5280/3600))^2 + 2ay(10.5 - extension)}}{ay}$$

Utilizing these 2 equations as time and t50 respectively, we can use calculus to derive that the average distance d between 2 pitches before the commit time is approximately

$$d = \frac{\int_0^{time_1} \sqrt{((\frac{ax_1t^2}{2} + t(vx_1 - ax_1(t50_1)) + X_1) - (\frac{ax_2t^2}{2} + t(vx_2 - ax_2(t50_2)) + X_2))^2 + ((\frac{az_1t^2}{2} + t(vz_1 - az_1(t50_1)) + Z_1) - (\frac{az_2t^2}{2} + t(vz_2 - az_2(t50_2)) + Z_2))^2}dt}{time_1}$$

Where az and ax are the accelerations at y=50, vz and vx are the velocities at y=50, and X and Z are the release positions X and Z respectively and $pitch_1$ and $pitch_2$ are the two pitches being compared. Differently from Mr. Limm, instead of using the mean of the distances between all instances of two pitches, I use the fifth percentile distance. If I were to use the mean, location would be much more important, as pitches tend to naturally start in different places just based on their terminal location. By taking a small percentile of the distances, you get a value of how close together those pitches can be. This is more meaningful because it's not about the distance between two pitches, it's about how similar they look. Even if you rarely throw a glove side fastball, your glove side changeup will look like a fastball out of hand if the trajectories are more similar. It's not about how close together they actually are; it's about how close they appear. Then using the difference in where they end up in both the X and Z plane, along with the difference in velocity, we can approximate the effect of a tunnel.

To predict run value, I created 2 Extreme Gradient Boost (XGB) models with optimized hyperparameters using r. The first uses pitch velocity, IVB, HB, release point horizontal, release height vertical, extension, and seam shifted wake to model the expected run value of a primary pitch. The second model uses these same variables along with the tunneling variables (d, dvelo, dmovex, dmovez) to account for how well that secondary tunnels with that pitcher's primary pitch against a handedness. Note that some pitchers throw different primary pitches to different handedness batters.

One player who truly shows the tunneling concept, and why a low percentile of the average difference in ball trajectory is superior to the mean, is Sandy Alcantara. He has the lowest average distance between his primary and secondaries of any pitcher. Against right handed hitters, his tunnels are also very location based. He tends to start pitches middle middle to middle away, then they break in different directions. However, against lefties, Alcantara’s tunneling no longer makes sense based on location. He starts his slider way farther outside than any other pitch. He starts his high fastball higher than any other pitch he typically throws. Both of these pitches are still highly effective though, as you don’t necessarily need location for tunneling. Just the threat of a pitch being a different pitch is tunneling. Alcantara still has elite tunnels against lefties because his pitches stay so incredibly close together, that he doesn’t need to have a consistent starting location for the tunnels to be relevant.

Results: Top primary pitches against righties

player_name	pitch_type
Peralta, Freddy	FF
Strider, Spencer	FF
Díaz, Alexis	FF
Hendriks, Liam	FF
Bush, Matt	FF
Gose, Anthony	FF
Kimbrel, Craig	FF
Pérez, Cionel	FF
Cleavinger, Garrett	FF
Díaz, Edwin	FF

Top primary pitches against lefties

player_name	pitch_type
Hader, Josh	SI
Helsley, Ryan	FF
deGrom, Jacob	FF
Strider, Spencer	FF
Karinchak, James	FF
Bush, Matt	FF
Snell, Blake	FF
Quijada, José	FF
Kikuchi, Yusei	FF
Clase, Emmanuel	FC

Top secondary pitches against righties

player_name	pitch_type
Ashby, Aaron	SL
Scott, Tanner	SL
Cortes, Nestor	SL
Lange, Alex	CU
Pérez, Cionel	SL
Márquez, Germán	KC
McCullers Jr., Lance	SL

Moore, Matt	KC
Martin, Brett	CU
Phillips, Evan	SL

Top secondary pitches against lefties

player_name	pitch_type
Burnes, Corbin	CU
Muñoz, Andrés	SL
Sanmartin, Reiver	SL
Díaz, Edwin	SL
Ohtani, Shohei	SL
Webb, Logan	SL
Hughes, Brandon	SL
Cole, Gerrit	KC
Castillo, Luis	SL
Gray, Josiah	CU

The results pass the common sense test here. Most of these pitches are known for being nasty. One pitch that stands out is Matt Bush's 4 seam fastball. It's the only pitch to show up on 2 different lists, and for good reason. It has a very low release height, combined with 97 MPH velocity and nearly 20 inches of induced vertical break. This leads to a very flat horizontal ascent angle, which is similarity between most of the top ranked fastballs. Bush's curveball also plays well off this fastball, averaging 5.5 inches away from the fastballs trajectory in the tunnel zone, but ultimately ending up roughly 30 inches lower. All in all, this model makes Bush's 1.8 million dollar contract seem like a bargain. While he put up a 3.5 ERA in 2022, and is getting paid off this, he has the stuff to be an elite relief pitcher.

One thing that this model is very good at investigating is where pitchers should be in relation to the rubber. Looking at what the model selects as the top 2 curveballs in the game, Gerrit Cole's knuckle curve and Corbin Burnes' curve, Burnes' is a little slower and breaks a little less. Surprisingly, however, the model favors Burnes' because Cole's release point x is over a foot farther to the third base side than Burnes'. This release makes the ball start outside to a lefty, and not initially appear to be a strike when it jumps out of his hand, as opposed to Burnes' which starts on the middle of the plate. The obvious solution would be to move Cole a foot over to the first base side of the rubber, but looking deeper, the model says this would negatively affect his fastball. Burnes' fastball works from that release because it's a hard cutter, but this release point would be detrimental to Cole's high IVB 4 seam fastball.

Overall, I am very happy with how well the model lines up with real life observational knowledge. Stuff models such as this one have many use cases such as moving around players on the rubber, evaluating individual pitches in small sample sizes, and finding better combinations of pitch types and trajectories for tunneling. Evaluating tunnels is extraordinarily difficult. There is so much more room to grow when it comes to tunneling. Secondaries tunnel with each other. Location matters more than I gave it credit for. Spin axis and picking up pitches through spin, like the classic dot on the slider is crucial, but difficult with current data. Some other smaller things that could be improved about the model is the hyperparameter training was a bit lackluster, as the model takes a relatively long time to train, leading to hours of hyperparameter optimization when running the model hundreds of times. Also the run value is certainly not a perfect description of the outcome of the pitch. It is partially based off of xwOBA, which is a highly flawed statistic, as it assumes randomness of spray angle which is not true on a macro level, and certainly not true in a singular pitch leading to one batted ball. While there's a lot more room for improvement, this model is a good starting point for quantifying the effects of an elusive baseball idea, tunneling.