Consistency in uncertain times: the differences between the 2020 and 2019 MLB seasons are not affecting pitches.

Michael Petersen and Erich Rentz

When the dust settles from the 2020 MLB season, the statistics will look radically different from the 2019 MLB season. But is the shortened 2020 season actually statistically different from the 2019 season? The rates at which pitchers are striking out and walking batters do not look appreciably different. The long-term trends in number of home runs does not appear to be broken. Is the 2020 season a statistical continuation of 2019?

A myriad of effects on the 2020 season might logically lead one to conclude that we should be able to identify differences. Most obviously, there are only 60 games in the season. But other changes may also play a role: National League pitchers never bat. Travel distances are shorter, on average. More games are played in fewer days, on average (and many were rescheduled with short notice). Can we detect the effect the changes have had on pitcher's performance? Unfortunately, at 60 games, the sample sizes are too small to discern trends 'by eye' owing to the inherent randomness in the game of baseball. Fortunately, techniques exist specifically to identify statistical differences, even in small data sets.

Through the machine-learning technique principal component analysis (PCA), we can tease out subtle changes to a pitch, even across a small sample size. PCA works to group the most common attributes between given pitches, leaving behind only differences, which may be correlated with other changes -- such as differences to the structure of the MLB season. PCA can accept a variety of data streams. MLB Statcast provides geometric data for pitch release points, pitch movement, pitch velocity, and spin rate, all of which can be used to generate unique signatures for the pitches in a pitcher's arsenal. Because the PCA fingerprints of a pitcher's arsenal may be trained on any subset of data and then provide a rigorous metric to compare with another subset, we can slice the data to address specific questions about changes to the 2020 MLB games versus 2019. One may ask specific questions about individual pitchers or larger trends by dividing up the data, such as: Does Gerrit Cole's reduced K/9 suggest that his pitches have changed? (No, see Figure 1.) Does playing more games in fewer days impact pitcher's ability to perform? (No.) Did Zack Greinke roll out a completely new pitch arsenal in 2020? (Yes.) Does the altered travel schedule, leading to more pitching appearances in a small number of ballparks, affect pitchers? (No.)

Using the PCA pitch fingerprinting technique, we show that pitcher's performances are, on average, unaffected by the changes to the MLB rules in 2020. We will discuss the most significant changes to the 2020 season as reflected in pitch fingerprints and outcomes, and make predictions for pitching trends in the 2021 season. (We also expect to analyse the postseason in advance of the paper due date to compare with the 2019 season)
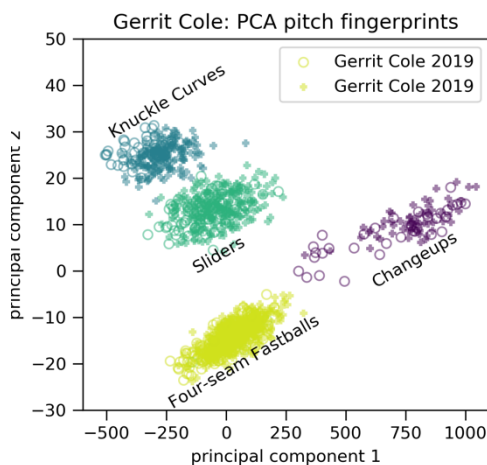

Figure 1. Comparison of Gerrit Cole's 2019 and 2020 PCA pitch fingerprints. That the circles and pluses lie on top of each other suggests that Gerrit Cole's 2020 pitches are indistinguishable from his 2019 pitches.