Dew Point and Pitching Take-Home Assignment

Jacob Engelberg

October 2023

1 Introduction

The goal of this project is to determine how dew point, a measure of water vapor in the air, affects pitching performance. When the dew point exceeds 65 degrees, it is considered to be uncomfortable to the pitcher and therefore has an impact on the pitches thrown. In addition, a higher dew point can have an effect on a ball in flight. By considering key pitching performance metrics and analyzing a data set of pitches thrown, I attempted to predict the probability that any given pitch was affected by dew point.

2 Methods

The pitch data provided included 26 different performance measures for 9889 different pitches. In my analysis, I focused on four key measures - velocity, spin rate, horizontal break, and induced vertical break. I picked these four measures because I believe, out of all the possible options, that these are the measures that are most subject to influence by environmental factors such as dew point and by the "comfortability" of the pitcher.

I made a couple of key assumptions about each of these four measures and their relationship to a high dew point. For velocity, I made the assumption that, as dew point increases and a pitcher becomes more uncomfortable, velocity decreases. Interestingly, this is at odds with the laws of physics, which say that a ball should move faster through thinner (more humid) air. However, I believe that the uncomfortability that a pitcher feels during times of high humidity has a bigger impact on velocity than air density. For spin rate, I made the assumption that in more humid air, spin rate increases because the moisture in the air makes a baseball easier for the pitcher to grip (similar to the sticky-substances that pitchers used to put on the baseball). For horizontal and vertical break, I made the assumption that as humidity increases, a baseball breaks less. This is because the more humid the air, the thinner the air is, meaning that there is less opportunity for air friction to affect the ball in flight.

3 Approach

The first step of my analysis was to clean and normalize the key data points that I decided to investigate. I extracted the measures that I deemed significant for this project, which were the unique pitch ID (PID), the MLB ID number for each pitcher (PITCHER_KEY), pitch type (PITCH_TYPE_TRACKED_KEY), pitch velocity (RELEASE_SPEED), spin rate (SPIN_RATE_ABSOLUTE),the horizontal break (HORIZONTAL_BREAK) and vertical break of a pitch (INDUCED_VERTICAL_BREAK). The four key measures discussed previously were normalized, with the averages and standard deviations being calculated within specific pitcher and pitch type groups.

The probability of dew point effect was calculated using cumulative probability density functions (CDF) for each of the four measures. The probability of dew point affecting a given pitch was calculated by averaging the CDF value for each of the four "key" measures, with a pitch that exactly matches a pitcher's average velocity, spin rate, horizontal break, and vertical break for a specific pitch type having a 50% chance of being affected by the dew point.

4 Conclusion

The data that my analysis produced is in the table below:

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0.03012	0.40791	0.48724	0.48620	0.56456	0.89700

This data represents the breakdown of the dew point affected percentages I calculated for all 9889 pitches. The minimum of .03 (3%) represents a pitch that had the lowest likelihood of being affected by a dew point higher than 65 degrees, and the maximum of 90% represents a pitch that was very likely (90% probability) affected by humidity. The mean and median both being around 50% makes sense because I made the determination in my analysis that a pitch that has the exact mean velocity, spin rate, and horizontal/vertical break for a given pitch type for a specific pitcher has a 50% chance of being affected by humidity.

One thing that requires further investigation is whether a series of pitches all deviating from the mean across multiple performance measures is more strongly indicative of dew point effect than a single pitch deviating from the mean on a single measure. Since the dew point is unlikely to change dramatically over the course of an innings, it stands to reason that if a high dew point does begin to affect one or more performance measures, that measure would likely be affected across a series of pitches rather than just on a single pitch here and there. So, if a series of pitches clustered together deviates from the mean significantly, and if those deviations from the mean occur across multiple measures of performance, there is a heightened probability that those pitches were affected by high dew point.