For the past several seasons, I have been in two separate dynasty leagues. I use the term Dynasty lightly because while they are deep in roster size, the number of prospects actually owned trends to the lower side. This casually sparked my interest in prospects and as a result led me to really dive into prospect content. One of my favorite things I found in recent seasons was KATOH from Chris Mitchell at Fangraphs. He built a stat based model which predicted player WAR. Unfortunately, he had to take his model down as he moved into the private sector and all of the fantastic analysis was no more. At the time of this writing a similar system has been developed for pitchers by Connor Kurcon of Six Man Rotation. If you have not checked that out I recommend you do.

This past season, I joined a 30 team dynasty league. Each team can hold a minors roster of 20 players, which when combined with the league depth means there are over 500 minor league eligible players rosters. I was overwhelmed and realized I did not nearly have the level of prospect knowledge of some of the others in the league. I also realized I did not really know how to attack these lesser named players who may be on my wire. I needed to know how their stat lines translated to fantasy.

Most prospect content does not focus specifically on the fantasy side of things. Typically, the goal of prospect research is to predict the best baseball players, so things such as defense and position are much more important. For fantasy, we do not care quite as much. We Just care about the numbers themselves and defense really does not matter at all. Thus I wanted to set out to build a model that looked at things entirely through the lens of fantasy baseball.

# Process

## Building a Sample

The first step to building this model was I needed a sample to work with. Fortunately, I had a general idea of the data I wanted to work with from a minor league perspective and I knew Fangraphs had this relevant information. However, they only have batted ball data going back to 2007, which proved to be useful in my overall modeling process meaning I could not include seasons prior to that in my model. I want to be able to find this information elsewhere going forward and see if it can possibly help to make the model even stronger in the future.

As for developing the minor league sample, I needed to also determine each player’s fantasy value. To do this I found the career numbers of all players who over the last 20 years had at least 250 MLB plate appearances. For each of these players I scaled their stats to a 600 PA sample as a proxy for a single season. Then using the SGP factors I developed from the 2018 Online Championship standings from NFBC I had a per 600 PA value for each player in my sample. Those who never reached the majors were obviously given a value of 0.

One thing I noticed initially was this overvalued older players who stole bases, as steals are worth more individually now in a SGP sense. I did perform some minor adjusting of stolen base totals in order to try to bring them down to 2018 standards. There are still some very high rankings for speedsters in the overall model especially those who may have been used in a pinch running role. This is something I plan to adjust in my future iterations but it works well as a starting point.

## Determining What Matters

After I had my pairs of seasons and the subsequent MLB fantasy value for each player, I was able to begin the model building process. I selected over 20 different metrics provided by Fangraphs and then divided each player season up based on the level it took place at. Right now I do not work with rookie ball data this is another update for the future. For each level split, I found the correlation between the metric and the total fantasy value and pulled each of the ones that had a significant correlation with value. \

This left me with anywhere from five to seven variables per level. These metrics tended to remain the same across levels including K/BB, Age, and ISO among others. In this initial run none of these metrics were scaled to take into account park factors or anything else. Going forward, I plan to test this with scaled metrics to take overall league context into effect as things like the juiced AAA baseball may help to explain some of the results I saw.

## Building The Model (NERD WARNING)

Now that I was able to determine the statistics that matter it was time to put 2019 to the test. I have long been a fan of using comparison based models for tasks like this because I think it is important to find the range of outcomes for a particular player type. My favorite type of comparison model uses mahalanobis distance. I have previously used this model for a number of different things including a relief pitcher model as well a expected fantasy points model in fantasy football and have loved the results. However, for this I found a new tool that will only help make it better.

One thing I’ve always wanted to incorporate into the model was a weighting system for the factors. Each factor has a relative importance and I wanted my model to reflect that. Luckily for me, I found a function that allowed me to provide factor weights in the mahalanobis distance calculation. Using this function I compared a given player season to all of the seasons at the same level in my sample. It then generated a distance between the seasons, the smaller the distance the more similar the season was. This then allowed me to sort by the most similar seasons. I took the 100 most similar seasons and then determined the weighted mean value of the MLB careers for each of those players. The weighting was based on how similar a given season was. This then provided me a value for each player at each level where they had at least 50 PAs in 2019. Then for players who played at multiple levels in 2019, I weighted each level’s value by the plate appearances they saw at that level and was able to come up with a single player value.

# Results

After my initial run, I was extremely pleased with the results and the players the model identified at the top. Below are the top 10 players according to the model (under 29 years of age).

|  |  |  |
| --- | --- | --- |
| Name | Age | xVal |
| Bryan Reynolds | 24 | 7.64 |
| Yordan Alvarez | 22 | 7.60 |
| Francisco Mejia | 23 | 7.45 |
| Keston Hiura | 22 | 7.34 |
| Josh VanMeter | 24 | 7.26 |
| Austin Riley | 22 | 6.90 |
| Kevin Cron | 26 | 6.88 |
| Luis Urias | 22 | 6.85 |
| Jared Walsh | 25 | 6.83 |
| Mac Williamson | 28 | 6.82 |

Several of the players on this list were called in in 2018 and went on to have extremely productive fantasy seasons including future Rookie of the Year Yordan Alvarez. It is interesting that the model thought so highly of Reynolds to place him at the top despite his relatively unheralded profile before the call-up. VanMeter is a really interesting guy who could be a fantasy stud if he gets a chance but is blocked currently in Cincinnati.

While I am happy with the early results I realized that the top section of my list was peppered with AAA guys. My belief is that since reaching AAA tends to mean that you have a pretty good chance of getting an MLB shot, these players have a lower probability of no value. This is evidenced by consensus #1 prospect, Wander Franco ranking 92nd despite some really impressive low minors numbers.

This led me to try a few different things to try to factor in level when making these rankings. My goal was to see how much better or worse a prospect was projected than the other prospects at their level. To do this I found the average projection and subtracted it from each level’s value. So the baseline projection in AAA was much higher than that in A. Below are the top 10 prospects based on this adjustment.

|  |  |  |
| --- | --- | --- |
| Name | Age | xVal\_Adj |
| Dylan Carlson | 20 | 4.79 |
| Josh Stephen | 21 | 4.73 |
| Gavin Lux | 21 | 4.25 |
| Brennen Davis | 19 | 4.14 |
| Jeter Downs | 20 | 4.12 |
| Bryan Reynolds | 24 | 3.87 |
| Wander Franco | 18 | 3.87 |
| Jarred Kelenic | 19 | 3.86 |
| Yordan Alvarez | 22 | 3.84 |
| Isaac Paredes | 20 | 3.77 |

As you can see the list is almost 100% different than the pure comparison based one. Both Reynolds and Alvarez appear here again though. The had truly special seasons in 2019. Many of the other names are top prospects who have yet to truly get their MLB chance namely Lux and Carlson. Through the rest of the list we see several lower level prospects begin to crack the top 10 including the aforementioned Franco. Brennen Davis, Jeter Downs, and Isaac Paredes appear to be fantasy studs in the making.

A really interesting guy is Josh Stephen who despite little love within prospect circles and stat profile that does not jump off the page he has a truly impressive ranking. He is an interesting case study and one I will need to dive deeper into but I do know that his high value projection is driven by a high ranking comparison to Jose Altuve in AA. Altuve made some significant changes upon reaching the big leagues so it may be important to look at that rank with some hesitation.

Overall, I think that both lists provide a significant amount of value. The pure value list could be useful in season when deciding if a player is worth spending FAAB money on or adding as they called up. On the other hand, the adjusted list could be more useful for offseason research and trade comparisons for dynasty leagues. It can help to contextualize what the offensive lines mean for different players at different ages and different levels.

Since both provide a certain level of value I wanted to make a blended rank of sort. This is the pure average between the two values and the top 10 are as follows.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Age | xVal | xVal\_Adj | Avg |
| Bryan Reynolds | 24 | 7.64 | 3.87 | 5.76 |
| Yordan Alvarez | 22 | 7.60 | 3.84 | 5.72 |
| Dylan Carlson | 20 | 6.44 | 4.79 | 5.61 |
| Francisco Mejia | 23 | 7.45 | 3.68 | 5.57 |
| Keston Hiura | 22 | 7.34 | 3.58 | 5.46 |
| Josh Stephen | 21 | 6.13 | 4.73 | 5.43 |
| Gavin Lux | 21 | 6.54 | 4.25 | 5.39 |
| Josh VanMeter | 24 | 7.26 | 3.49 | 5.37 |
| Austin Riley | 22 | 6.90 | 3.13 | 5.01 |
| Kevin Cron | 26 | 6.88 | 3.12 | 5.00 |

As you can see this list mirrors the original a little bit more than the adjusted but it does a good job in my opinion of displaying both instant impact and value compared to the average minor leaguer at their levels.

# Additional Notes/Applications

All of this information can be found at the link provided [here](https://pmammino18.shinyapps.io/prospectreport/). This has the full leaderboard and all three of the discussed values for each minor league player in 2019. It also has two other useful tools.

## Comparison List

This is the driving force behind the model and for each player season you can see the top comparable players. This will be the best comps for the player at each level as well as levels split depending on if they played for multiple teams. So for a player like Taylor Trammell, he will have comps for AA with both the Reds and the Padres. This can extremely helpful when evaluating the high ranking of a guy like Stephen. Looking at him we can see the incredible favorable comparisons that have led to such lofty expectations.

## Probability Chart

This in my opinion is the best part of the site and the model. Again for each player season split by level and team, I can create a chart with the range of outcomes. I shared this on Twitter after his extension but here is the chart for Luis Robert in AA.

A screenshot of a cell phone

Description automatically generated

As you can see there is a fairly significant change of no impact still for Robert, but the extension signing makes that a much smaller percentage. Something catastrophic would need to occur at this point. Of the chances where his comps made the majors, you can see that he has significant fantasy upside which is evidenced by the fact that Mike Trout and Ronald Acuna both showed up in his top comps. These is an extremely valuable tool because it can help to show you the risk in a given player and helps to visualize the high upside/downside prospect types we talk about frequently in fantasy.

## Flaws and To Do

As is the case with any first run at something, there are some flaws with my analysis. At the moment I am including all seasons from 2007- 2018 in my model so that means that some of the zero entries will be because the player has yet to reach majors but is still playing minor league baseball. Going forward I plan on finding an appropriate way to remove or account for these players.

Additionally, this is only a single season snapshot of each of these players. I plan on using previous seasons of data to develop these projections in order to account for down seasons or injury impacted seasons. As I mentioned in the beginning, I also want to find a way to adjust for league and park context for these players. The fact that so many AAA guys rank highly could be due to the insane juiced ball that inflated numbers across all of AAA. In order to account for this, I may need to adjust the current and previous seasons to the league as a whole.

This is just a sampling of the different things I want to incorporate going forward and the improvements that I already have planned as well as building a similar model for pitchers. It is important to note that is an entirely stat based approach and should never be used as a replacement for actual scouting. I am not where near as talented as those who scout for a living and would never think I can replace them. It is also important to remember this is entirely based on fantasy value. Just because a player ranks highly on here does not mean they are going to be a great real life baseball player, fantasy is all that matters here.

# Conclusion

Overall, I am extremely excitied about the prospects of this model and I think it could be an invaluable tool for the fantasy baseball community. Looking back at the 2019 results, it would have directed us to take a look at Bryan Reynolds who certainly flew under many people’s radars upon his call-up. I am eager to address any questions, concerns, feedback, or other thoughts that you may have on the model and the tools I have built to interact with it.