

MiLB Sprint Speed Model Workflow & Analysis

Reasoning

Until recently, minor league sprint speeds have been unavailable due to lack of tracking technology. Although most AAA affiliates have changed to Hawkeye in recent seasons, most of the lower levels still lack this technology. The following minor league sprint speed prediction model leverages proxies for sprint speed that are independent but proven to still be predictive of speed. The model is built on 2022-2023 MLB data, using Hawkeye-tracked sprint speeds as our target.

Proxies for Sprint Speed

The following factors were used to try and predict minor league sprint speed:

- **Stolen Base Opportunities Taken (SBOT)** – proportion of the number of stolen bases attempted over the number of opportunities to steal
- **Triples Rate** – number of triples over the total number of at bats
- **First to Third Rate** – proportion of the number of times a baserunner went from first to third on a ball in play over the number of opportunities to go first to third
- **Double Play Avoidance Rate*** – proportion of the number of times a batter beat out the backside of a double-play ground ball over the number of plays in which a double play was predicted based on batted ball characteristics
- **Age** – a player's age on 4/1 of the focus year

** Double play probabilities were found using a separate xgboost model*

The Model

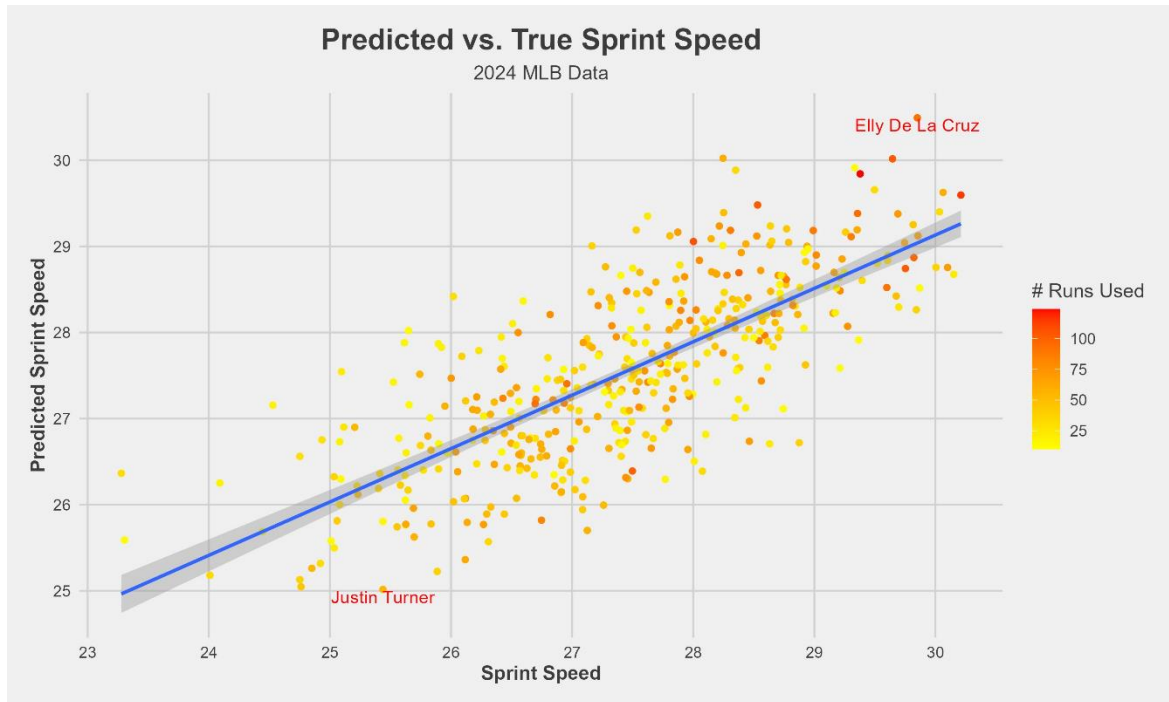
To build the model, each of the metrics listed above were calculated for every qualified player (>10 tracked sprints per season) from the 2022-2023 MLB season. A two-year weighted average, weighted by year and plate appearances per season, was found for each stat for each player. To build a proxy for sprint speed, a generalized additive model (GAM) was produced, with all predictors having their own smooth term and the true average sprint speed (in ft/sec) as the target variable. Below are the significance levels for each smooth term.

Approximate significance of smooth terms:				
	edf	Ref. df	F	p-value
s(AGE_401)	8.406	8.900	11.008	< 2e-16 ***
s(SBOT)	3.818	4.744	46.686	< 2e-16 ***
s(Triples_Rate)	1.000	1.000	17.041	4.26e-05 ***
s(f_t_rate)	4.366	5.441	6.261	7.92e-06 ***
s(Non_DP_Rate)	7.161	8.181	8.632	< 2e-16 ***

All predictors were found to be significant, but especially age, SBOT, and double play avoidance rate. Based on the smooth term plots (not shown), all had a positive relationship with sprint speed except for age, which had a negative relationship.

MiLB Sprint Speed Model Workflow & Analysis

Once the model was built, it was tested using 2024 MLB data. The model fits the data reasonably well, as can be seen in the graph below.

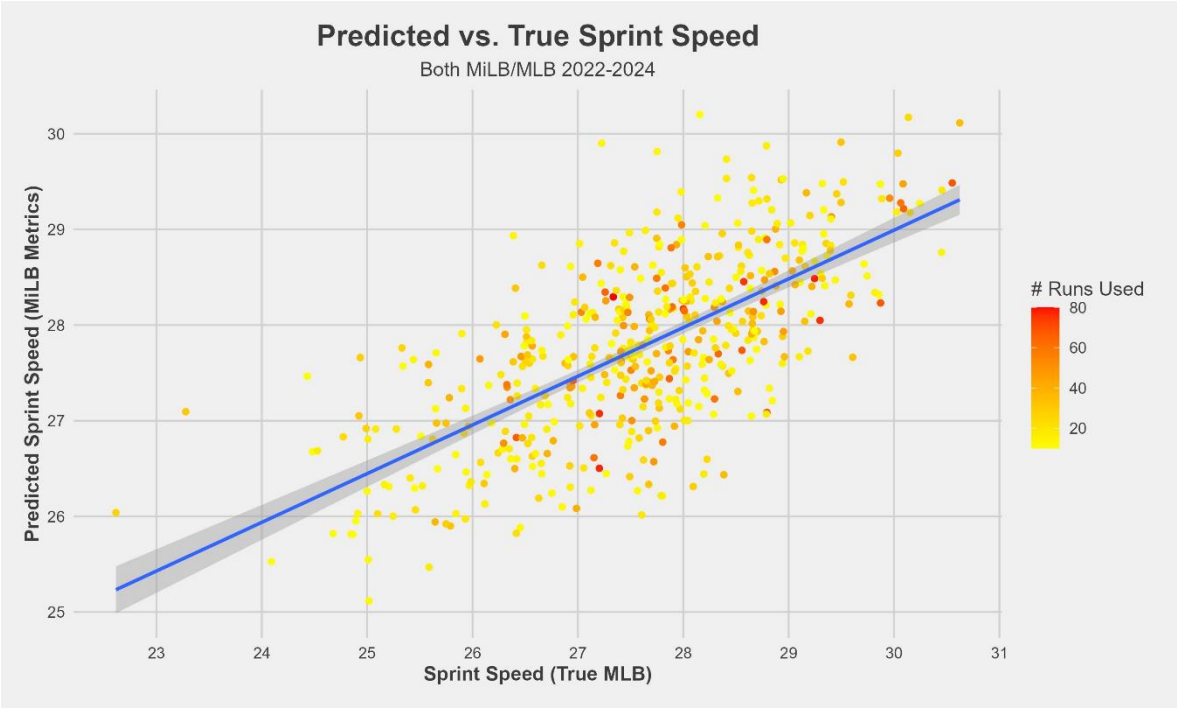


The RMSE for the testing set was 0.84, while the standard deviation of the true sprint speeds was 1.27. The model error was within one standard deviation, which signals the model is doing a solid job of predicting sprint speed. The coefficient of determination (R^2) was 0.57, indicating a “moderate” correlation between predicted and true sprint speed. The fastest predicted player was Elly De La Cruz (30.5 ft/sec) and the slowest predicted was Justin Turner (25.0 ft/sec).

Testing On Players at Both Levels

To further check the model’s reliability, predictions were made on all players who qualified at both the minor and major league level in 2022, 2023, and 2024. The results were similar to the test data and are shown below.













MiLB Sprint Speed Model Workflow & Analysis



With an RMSE of 0.93 and standard deviation of 1.26, we can again say the model is doing a good job of predicting sprint speed based on minor league speed metrics and true tracked sprint speed at the big-league level.

Here are all **RANGERS** players from 2022-2024 who qualified at both levels (as of 9/17/24):

Predicted Sprint Speeds for Players that Qualified at Both MiLB/MLB

2022-2024						
Player	Year	Org	AGE 4/1	MiLB PA	True Sprint Speed	Predicted Sprint Speed
 Brad Miller	2023	TEX	33.4	66	25.8	26.0
 Sam Huff	2022	TEX	24.2	274	26.9	26.3
 Andy Ibanez	2022	TEX	28.9	315	27.5	27.7
 Josh Jung	2024	TEX	26.1	29	27.5	27.8
 Nick Solak	2022	TEX	27.2	259	28.5	28.0
 Josh Smith	2022	TEX	24.6	261	28.6	28.0
 Ezequiel Duran	2024	TEX	24.8	86	28.9	29.0
 Ezequiel Duran	2022	TEX	22.8	355	29.1	28.4
 Leody Taveras	2022	TEX	23.5	221	29.4	29.1
 Evan Carter	2023	TEX	20.5	501	29.4	29.1
 Derek Hill	2024	TEX	28.2	176	29.5	29.5
 Bubba Thompson	2022	TEX	23.8	375	30.6	30.1

MiLB Sprint Speed Model Workflow & Analysis

2024 MiLB Leaders

Finally, here are the top 10 MiLB players in terms of predicted sprint speed in 2024 (as of 9/17):

Predicted Sprint Speed Top 10 in MiLB (2024)

Player	Year	Comp Level(s)	Org(s)	AGE 4/1	MiLB PA	Predicted Sprint Speed
Luis Valdez	2024	AA,A,A+	BAL	24.1	121	31.0
Yhoswar Garcia	2024	A	MIL	22.5	394	30.9
Tavian Josenberger	2024	AA,A+	BAL	22.4	372	30.7
John Peck	2024	A+,A	DET	21.7	282	30.7
Alexander Vargas	2024	A+,AA	NYN	22.4	272	30.7
Yosy Galan	2024	A+	TEX	22.9	326	30.7
Bryce Teodosio	2024	AAA	ANA	24.7	462	30.7
Jonah Cox	2024	A+,A	SF	22.6	460	30.6
Elis Cuevas	2024	A	BAL	19.3	81	30.6
Matthew Etzel	2024	AA,A+	BAL,TB	21.9	484	30.5

And here are the fastest predicted players in the **RANGERS** organization in 2024 (as of 9/17):

Predicted Sprint Speed Top 10 in TEX Org (2024)

Player	Year	Comp Level(s)	Org(s)	AGE 4/1	MiLB PA	Predicted Sprint Speed
Yosy Galan	2024	A+	TEX	22.9	326	30.7
Sebastian Walcott	2024	AA,A+	TEX	18.0	515	30.1
Chandler Pollard	2024	A	TEX	19.9	326	30.0
Anthony Gutierrez	2024	A+	TEX	19.3	271	29.9
Cameron Cauley	2024	A+	TEX	21.1	404	29.7
Wady Mendez	2024	A	TEX	19.4	262	29.7
Alejandro Osuna	2024	A+,AA	TEX	21.4	456	29.3
Josh Hatcher	2024	AA	TEX	25.5	454	29.2
Jose Barrero	2024	AAA	TEX	25.9	188	29.1
Marcus Smith	2024	AAA,A+,A	TEX	23.5	363	29.1

Overall Assessment

The model provides sufficient estimates for minor league sprint speed. Looking at the graphs of predicted vs true sprint speed, some of the extreme cases can be shrunk towards the mean due to a lack of data at the top and bottom. However, most of the players in the middle (~27-29 ft/sec) will be relatively accurate. All players will fall within the realm of baserunners with similar sprint speed.

MiLB Sprint Speed Model Workflow & Analysis

****Message to Code Manager****

The script for both the MiLB sprint speed model and predictions can be found in Databricks in `databricks_analytics/Lance/MiLB_Sprint_Speed`. You will need to set up a job to run each Sunday around 6 pm so the `sandbox.analytics.milb_sprint_speed` table can be updated regularly.