



SMART FANTASY BASEBALL

USING THE PROJECTION AGGREGATOR

WHAT IS THE PROJECTION AGGREGATOR?

I admit the name of this thing is extremely boring, but an Excel file that can quickly and easily calculate the average of up to three separate projection systems into one is pretty exciting to me.

The Projection Aggregator can aggregate several projections using equal or customized weighting (to give priority to your favorite projections). You can include playing time estimates from one or all of the systems and you can even enter your own playing time estimate if you disagree.

I created this for two reasons:

1. Combined projections have proven to be more accurate than individual projections (or forecasts). Here's a [formal research paper](#) on the topic and here's a [post describing the findings at SFBB](#).
2. I normally prefer to teach others how to do something, but the act of combining different systems together [proved to be extremely difficult](#). Manually calculating an average of three separate systems in one Excel file could take hours to do. It seemed much more practical to build a flexible Excel file that could be easily used by many people than to write a very complex instruction manual that would be far less usable.

WHAT CAN THE PROJECTION AGGREGATOR DO?

Here's a list of the specific things the Projection Aggregator can do:

- Combine up to three separate hitter and three separate pitcher projections into one with minimal setup involved.
- Combine any projection system provided it uses player IDs from Fangraphs, MLB, CBS, Retrosheet, Baseball-Reference, NFBC, ESPN, Davenport, or Baseball Prospectus.
- Calculate a combined projection using equal weighting (a simple average of all three) or using your own custom weighting preferences (maybe you want PECOTA to be weighted twice as heavy as Steamer).

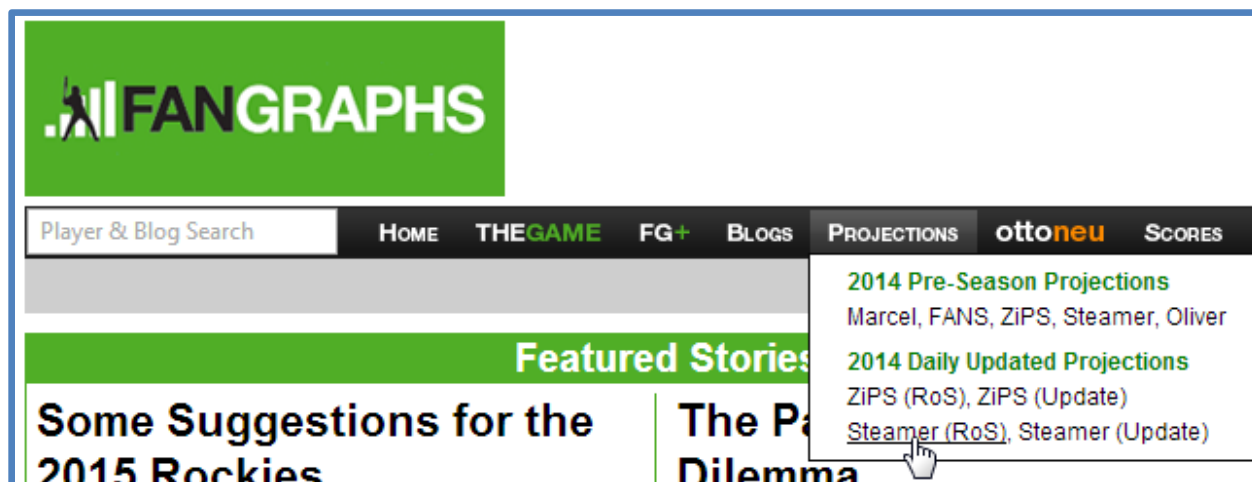


- Convert all projection information into a per plate appearance or per inning pitched basis (playing-time-neutral).
- Extend all of the playing-time-neutral projections out using playing time estimates from the projection system(s) of your choice (maybe you want Steamer and ZiPS playing time projections but not Oliver) or your own personal estimate of playing time for the player.


WHAT TYPE OF PROJECTION DOWNLOAD DO YOU NEED?

The ideal projection download is an Excel file or delimited text file (e.g. CSV) that can be easily inserted as a new tab or worksheet in Excel. The download must have a column containing Player ID information.

The projections offered freely at Fangraphs are in this exact format. If you have access to premium services like Baseball Prospectus PECOTA projections, they should also be available in the format described above.



INSTRUCTIONS

Step	Instructions
1.	Before beginning, I would suggest making a copy of the Projection Aggregator and keeping an original version. You can use this many times into the future, so keeping the original unaltered file is a good idea.
2.	Download projection information from your favorite sources. This can be preseason or rest of season projection information. The ideal download will be in Excel or CSV format (which can easily be opened by Excel).
3.	Copy projection information into the Projection Aggregator.
4.	Look at your downloaded projection information to make sure Player ID is in the first column of the data. If it's not, cut the column and move it so it appears in column A.
5.	Repeat steps 2 – 4 for each set of projection data you want to include.
6.	After you have brought in all of your projection information, go to the “Settings” tab. 
7.	In the “Tab Name” column, type in the EXACT name of the tab/worksheet containing the first projection system. Type this information into the “Hitter Projection #1” row. Return to and look at your projection information for “Hitter Projection #1”. Determine the last column containing information. In this example below, the information ranges to column “X”:



R	S	T	U	V	W	X	Y
OBP	SLG	OPS	wOBA	Fld	BsR	WAR	
0.401	0.552	0.954	0.411	-0.5	1.4	3	
0.392	0.562	0.954	0.406	-0.2	-0.2	2	
0.391	0.549	0.94	0.405	2.2	-0.1	1.9	
0.394	0.521	0.915	0.396	-0.1	0	1.5	

Return to the “Settings” tab and enter the “Column Range” you just identified. For the example directly above, I would have entered “A:X”.

	A	B	C
1			
2		Tab Name	Column Range
3	Hitter Projection #1	ZiPS Hitter ROS 8.2.14	A:X

8. Repeat step 7 for each projection system you are using. If you are only using two sets of projections, leave “Hitter Projection #3” blank.

	A	B	C
1			
2		Tab Name	Column Range
3	Hitter Projection #1	ZiPS Hitter ROS 8.2.14	A:X
4	Hitter Projection #2	Steamer Hitter ROS 8.2.14	A:W
5	Hitter Projection #3		

9. After you have named your tabs and given the column range, look under the “PROJECTION COLUMN NUMBER” section (to the right of the yellow shaded cells).



Check that the Projection Aggregator is properly locating the columns from your downloaded projection information.

You want to see this:

PROJECTION COLUMN NUMBER									
PA	AB	H	2B	3B	HR	R	RBI	BB	SO
4	5	6	7	8	9	10	11	12	13
3	4	5	6	7	8	9	10	11	12

You don't want to see "Not Found":

IP	W	L	SV	H	BB	ER	HR	SO
8	3	4	Not Found	9	13	10	11	12
9	3	4	8	10	14	11	12	13
Not Found	Not Found	Not Found	Not Found	Not Found	Not Found	Not Found	Not Found	Not Found

Any number populating in this area means the Projection Aggregator has located that statistic in your download. For example, Doubles (2B) are located in column 7 of the "Hitter Projection #1" system and column 6 of the "Hitter Projection #2" system.

Any "Not Found" entry indicates that the statistic cannot be found in your projection information. For example, one of the pitcher projections above cannot locate Saves (SV). This may be because the projection system simply does not project saves (this is not unusual), or it may be that your projection data uses a different abbreviation.

The "Projection Column Number" is looking for the exact abbreviations listed. So if your projection system uses the abbreviation "K" for pitcher strikeouts, you will need to go back to your projection data and change the column header to be "SO" instead.

It is OK if your projection system does not include all the statistics listed. If a statistic cannot be located, it will



simply not be used in the calculation of the averages. For example, if you bring in three sets of pitching projections and only one of the systems projects saves, the combined projection will really just be that one system's saves projections.

10. After you have verified the statistics are being located in the “PROJECTION COLUMN NUMBERS” section, decide on how you wish to weight the projections. Type a numeric value in the “Weighting” column for each projection system used.

If you want equal weighting, just make sure you enter equal values for all projection systems. For example, entering a “1”, “1”, and “1” for three systems will give equal weighting. Entering “1”, “2”, “1” will give one system double weighting.

If a name is not given for a projection system, it will not be given any weight, even if you provide a value in the weighting column.

	A	B	C	D
1				
2		Tab Name	Column Range	Weighting
3	Hitter Projection #1	ZiPS Hitter ROS 8.2.14	A:X	1.00
4	Hitter Projection #2	Steamer Hitter ROS 8.2.14	A:W	1.00
5	Hitter Projection #3			1.00
6				3.00



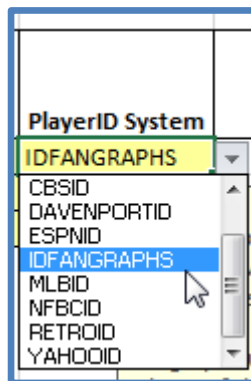
11. Specify if a projection system should be included in the plate appearance or innings pitched estimate of playing time.

Some projection systems make no effort to project playing time, so you should choose “No”. Some systems make a very significant effort at projecting playing time.

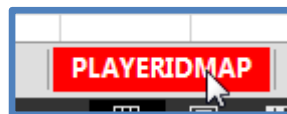
You must choose at least one system to be included in the PA or IP playing time projection.

Include PA in AVG?
No
No
No
Select at least one system to be used for PA average.

12. Finally, choose the player ID system used for each projection system.



If you're unsure of the system being used, locate the PLAYERIDMAP tab of the spreadsheet. Attempt to trace player IDs from your download file into the map. The columns in the PLAYERIDMAP correspond to the drop down menu on the “Settings” tab.

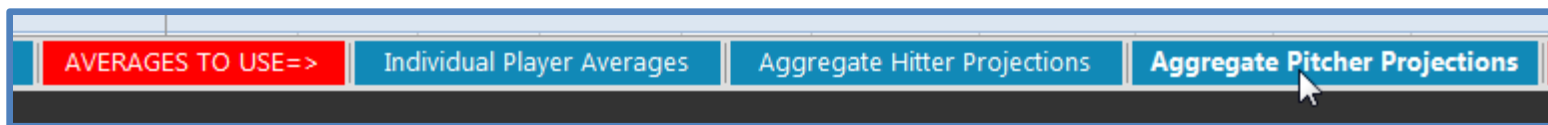


13. That's it!



VIEWING THE RESULTS

After completing the steps above, you can view the final aggregated projections on the “Aggregate Hitter Projections” or “Aggregate Pitcher Projections” tabs.



If you don't like the playing time estimate, you can enter your own IP (for pitchers) or PA (for hitters) estimates in the “OVERRIDE” column. Your playing time estimate will be multiplied by the aggregated playing-time-neutral projections.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	PLAYERID	LNAM	FNAM	TEAM	POS	IDFAN	IP	OVERRIDE	IP_US	W	L	SV	H	ER	HR	SO	BB	ERA	WHIP
2	darviyu01	Darvish	Yu	TEX	P	13074	63.00		63.00	4.89	2.45	-	49.34	21.50	6.34	77.59	22.45	3.07	1.14
3	scherma01	Scherzer	Max	DET	P	3137	66.00		66.00	4.86	2.91	-	56.76	24.26	7.30	74.19	18.93	3.31	1.15
4	salech01	Sale	Chris	CHW	P	10603	62.00		62.00	4.03	3.02	-	52.93	20.66	6.05	68.06	14.61	3.00	1.09
5	strasst01	Strasburg	Stephen	WAS	P	10131	62.00		62.00	4.57	2.53	-	51.32	20.32	5.08	67.58	16.25	2.95	1.09
6	kershcl01	Kershaw	Clayton	LAD	P	2036	62.00		62.00	5.24	1.96	-	47.79	15.79	4.32	68.24	13.38	2.29	0.99

You can also view all of the projections for an individual player on the “Individual Player Averages” tab. Simply choose the desired player using the yellow drop down menu cell.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2		Select a Hitter:	Jose Abreu																		
3		Player ID:	abreujo02																		
4																					
5																					
6																					
7																					
8																					
9																					
10																					



ADDING PLAYERS

For the Projection Aggregator to work properly, each player must be listed in each of the following tabs:

Hitters	Pitchers
PLAYERIDMAP	PLAYERIDMAP
Hitter Averaging	Pitcher Averaging
Aggregate Hitter Projections	Aggregate Pitcher Projections

If you determine that a player is missing from the Projection Aggregator, be sure to add their player ID in each of the three locations. When adding a player to the PLAYERIDMAP you should add information to these columns of the tab:

- IDPLAYER
- PLAYERNAME
- TEAM
- POS
- And the player ID information for the projection systems you will be using (for example, if you'll be using the Steamer projections from Fangraphs, be sure to complete the IDFANGRAPHS column).

