

# AggPro: The Aggregate Projection System

Ross J. Gore, Cameron T. Snapp and Timothy Highley

**Abstract**— Currently there exist many different systems to predict the performance of Major League Baseball (MLB) players in a variety of statistical categories. We propose, AggPro, an aggregate projection system that forms a projection for a MLB player’s performance by weighting and aggregating the player’s projections from these systems. Using automated search methods each projection system is assigned a weight. The determined weight for a system is applied to all the projections from that system. Then, an AggPro projection is formed by summing the different weighted projections for a player across all the projection systems. The AggPro projections are more accurate than the projection systems when evaluated by average error, root mean square error (RMSE) and Pearson correlation coefficient from actual player performance for the 2008 and 2009 MLB seasons.

## I. INTRODUCTION

Many different methods for projecting the performance of Major League Baseball (MLB) players in a variety of statistical categories for an upcoming MLB season exist. These projection systems include: Brad Null [1], Bill James Handbook [2], CAIRO [3], CBS [4], CHONE [5], ESPN [6], Hardball Times [7], Hit Tracker [8], KFFL [9], Marcel [10], Oliver [11], PECOTA [12], RotoWorld [13], and ZiPS [14]. Despite the availability and prevalence of these systems there has been relatively little evaluation on the accuracy of the projections from these systems’. Furthermore, there has been no research that attempts to aggregate these projection systems together to create a single more accurate projection.

We propose, AggPro, an aggregate projection system that forms a projection for a MLB player’s performance by weighting the player’s projections from the existing projection systems. We refer to each existing projection system employed by AggPro as a constituent projection system. Using automated search methods each constituent projection system is assigned a weight. The determined weight for a constituent system is then applied to the projections from that constituent system for the upcoming year. Then, an AggPro projection is formed by summing the different weighted constituent projections for a player across all the projection systems.

We believe the aggregate projections contain the best parts of each projection system resulting in a system that is more accurate than any of the constituent systems in the AggPro projection. The AggPro projections are evaluated against all the constituent systems by measuring the average error, root mean square error (RMSE) and Pearson correlation coefficient of the projections from actual player performance for the 2008 and 2009 MLB seasons.

It is important to note that AggPro is not just another projection system. Instead it is a methodology for aggregating

effective projections from different systems into a single more accurate projection. Furthermore, Greg Rybarczyk [8] believes paradigm shifts that will improve the accuracy of projection systems are on the horizon. If paradigm shifting projection systems are developed, the AggPro methodology will be applicable and improve the projections from these systems as well.

In the next section we describe work related to AggPro. Then, AggPro is presented and evaluated. Finally we conclude the paper and present directions for future work with AggPro.

## II. RELATED WORK

Research efforts in the areas of baseball, computer science, and artificial intelligence have all contributed to AggPro. We review these related works here.

### A. BellKor and The Netflix Prize

The strategy of applying different weights to different predictions from effective projection systems has been used successfully by the winning solution for the Netflix prize [15], BellKor by AT&T labs [16]. In October, 2006 Netflix released a dataset of anonymous movie ratings and challenged researchers to develop systems that could beat the accuracy of its recommendation system, Cinematch. A grand prize, known as the Netflix Prize, of \$1,000,000 was awarded to the first system to beat Cinematch by 10%. The BellKor prediction system was part of the winning solution, with 10.05% improvement over Cinematch.

BellKor employs 107 different models of varying approaches to generate user ratings for a particular movie. Then BellKor applies a linear weight to each model’s prediction to create an aggregate prediction for the movie [16]. AggPro applies this prediction strategy to projecting the performance for MLB players by employing the different existing MLB projection systems.

### B. Nate Silver’s 2007 Evaluation of Projection Systems

In 2007 Nate Silver performed a quick and dirty evaluation of the on-base percentage plus slugging (OPS) statistic projection from eight 2007 MLB projection systems [17]. Silver’s work offers several evaluation metrics including average error, RMSE and Pearson’s correlation coefficient, which we employ to evaluate AggPro. However, Silver also offers a metric to determine which system provides the best information. The metric is based on performing a regression analysis on all the systems for the past year and identifying "which systems contribute the most to the projection bundle [17]." AggPro performs this same regression analysis using the projections of systems for the past year. Then AggPro applies each metric identified by the analysis as a weight to system’s projections for the upcoming year. This methodology identifies most accurate parts of each projection system and combines these parts in one aggregate projection produced by AggPro.

### III. AGGPRO

The AggPro projections are generated through a three part process. First, we collect the projections from five different systems for the years 2007, 2008 and 2009. Next, for each year we identify the players that were common among all five systems. We also identify the statistical categories that were common among all five projection systems.

For the upcoming year, we perform an automated search over all the combinations of possible weights for the projections of five systems from the previous year. The automated search identifies the weight set that minimizes the root mean squared error (RMSE) of the previous year's aggregate projections from the actual player performances for the previous year. Next, we apply the identified weight set to the projection from the five systems for the upcoming year. This process is discussed in more detail in the remainder of this section.

#### A. Projection and MLB Actual Data Collection

We collected projections from Bill James Handbook [2], CHONE [5], Marcel [10], PECOTA [12] and ZiPS [14] for the years 2007-2009. We collected the actual MLB performance data for 2007 and 2008 from Baseball Prospectus. These projection systems are a representative sample of the many different systems that exist. If AggPro can successfully create an aggregate projection from these systems that is more accurate than any of the constituent projection systems then the AggPro methodology will have been shown to be successful. Given a successful methodology the reader can apply AggPro to any combination of constituent projection systems s/he chooses.

#### B. Identification of Players and Statistics to Project

Recall, that each year AggPro only projects the performance of those players common to all five systems. The player list for each year is available at [18]. Also recall that AggPro can only project those statistical categories that are common to all five systems. The hitter categories common to the five systems are: At Bats, Hits, Runs, Doubles, Triples, Home Runs, RBIs, Stolen Bases, Walks, and Strikeouts. The pitcher categories common to the five systems are: Innings Pitched, Earned Runs, Strikeouts, Walks, and Hits. These sets of players and statistics represent the largest possible set that was common to all the systems.

#### C. Automated Search To Identify AggPro Weights

Given the five projection systems, the set of common statistics and common players the AggPro projections for an upcoming year are generated as follows:

1. The projections for the five systems for the previous year are gathered.
2. The actual MLB performance data for the previous year is gathered.
3. A brute force automated search is performed to identify the set of weights that when applied to the projections of the five systems for the previous year minimize the RMSE of the previous year's aggregate projections from the actual player performances for

the previous year. Within the automated search the aggregate projection is formed by applying each weight in the set to its respective projection system and summing the projections for a player together.

4. Once the search is completed, the identified weight set is applied to the projections of the five systems for the upcoming year. The AggPro projections for the upcoming year are formed by applying each weight in the set to its respective projection systems and summing the projections for a player together.

We generated AggPro projections for the year 2008 and 2009. For the 2008 AggPro projections, the weight set that minimizes the RMSE of the 2007 aggregate projections from the 2007 actual MLB player performance data is Bill James Handbook = 0.56, CHONE = 0.00, Marcel = 0.15, PECOTA = 0.29, and ZiPS = 0.00. Applying these weights to the projection systems for 2008 generates the 2008 AggPro projections. For the 2009 AggPro projections the weight set that minimizes the RMSE of the 2008 aggregate projections from the 2008 actual MLB player performance data is Bill James Handbook = 0.37, CHONE = 0.00, Marcel = 0.35, PECOTA = 0.28, and ZiPS = 0.00. Applying these weights to the projection systems for 2009 generates the 2009 AggPro projections.

In the next section we evaluate the accuracy of the AggPro projections for each year using average error, RMSE and Pearson's correlation coefficient as evaluation criteria.

### IV. EVALUATION

AggPro and the five constituent projection systems were evaluated by computing the average error, RMSE, and Pearson correlation coefficient for each year for each statistical category from the MLB actual data. All of this evaluation data is shown and discussed in the Appendix.

For each system, for each year we also computed the average of each evaluation criterion over all the statistical categories. Each year we identified the best constituent projection system (BCPS). The BCPS is the constituent system for a given year which had the best average evaluation criterion over all the statistical categories. Furthermore, we identified the best constituent projection in each statistical category for each evaluation criterion. Combining the best constituent projections of each category forms the theoretical projection system (TPS). The TPS amounts to given a fictional oracle function at the beginning of the season which could pick the most accurate projection from the five systems for each statistical category. Due to how it is constructed the TPS is guaranteed to be at least as accurate as the BCPS. We also computed the average of each evaluation criterion over all the statistical categories in the TPS. AggPro's percent improvement over the BCPS and the TPS for the average of each evaluation criterion for each year is shown in Table 1-3. The 2009 projections are evaluated through MLB games completed on September 20<sup>th</sup>, 2009.

#### Average Error

Year	Percent Improvement over BCPS	Percent Improvement over TPS
2008	5.7 (Bill James)	3.4

2009	4.2 (Bill James)	1.5
------	------------------	-----

**Table 1: The average error evaluation of AggPro.**

RMSE		
Year	Percent Improvement over BCPS	Percent Improvement over TPS
2008	7.2 (Bill James)	2.4
2009	6.5 (Marcel)	2.6

**Table 2: The RMSE evaluation of AggPro.**

Pearson Correlation Coefficient		
Year	Percent Improvement over BCPS	Percent Improvement over TPS
2008	2.3 (Bill James)	2.3
2009	0.7 (Bill James)	0.2

**Table 3: The Pearson correlation coefficient evaluation of AggPro.**

AggPro is an improvement over both the BCPS and TPS for each evaluation criterion. This result is surprising. Since the TPS is constructed to contain the best constituent projection for each statistical category we did not anticipate that AggPro would outperform it. Instead, we had anticipated that the TPS would be a baseline for the best theoretical improvement AggPro could achieve. However, it appears the weighting of the different projections creates an aggregate projection that is more than the sum of the best parts of the constituent projection systems. This bodes well for future work with the AggPro methodology.

#### V. CONCLUSION

There exist many different systems to predict the performance of Major League Baseball (MLB) players in a variety of statistical categories. We have shown that our methodology, AggPro, can aggregate these existing projection systems into a single aggregate projection that is more accurate than any of AggPro's constituent project systems. Furthermore, AggPro is more accurate than the TPS when measured by any of the three evaluation criteria for the years 2008 and 2009. In other words, even if a reader was given a fictional oracle function at the beginning of the season which could pick the most accurate projection from the five systems for each statistical category, AggPro's predictions would still be more accurate for the upcoming season.

In future work with AggPro we will explore using distinct weight sets for the constituent projection systems for hitting statistic categories and pitching statistic categories

#### APPENDIX

The evaluation of each system for each statistical category, for each evaluation criterion is listed in the following tables. AggPro is abbreviated with AP, Bill James Handbook is abbreviated with BJ, CHONE is abbreviated with CH, Marcel is abbreviated with M, PECOTA is abbreviated with P and ZiPS is abbreviated with Z. The system that performs the best for the given evaluation criterion for the given year is bolded.

Average error is the measure of the average absolute

(without regard to sign) error of the player projections. Average error is measured for each statistical category. The system with the smallest average error in each category in each year is bolded to indicate that it is the most accurate. The 2009 projections are evaluated through MLB games completed on September 20<sup>th</sup>, 2009.

#### Average Error: Hitter At Bats

Year	AP	BJ	CH	M	P	Z
2008	<b>111.0</b>	112.8	151.3	128.4	129.3	139.5
2009	104.3	<b>103.1</b>	134.2	118.2	115.4	124.5

#### Average Error: Hitter Hits

Year	AP	BJ	CH	M	P	Z
2008	<b>33.5</b>	34.6	42.5	37.1	37.4	40.5
2009	<b>31.8</b>	32.3	38.8	35.0	33.9	44.5

#### Average Error: Hitter Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>19.1</b>	19.9	23.5	20.7	20.6	21.8
2009	<b>17.1</b>	17.9	21.3	18.9	18.1	61.2

#### Average Error: Hitter Doubles

Year	AP	BJ	CH	M	P	Z
2008	<b>8.1</b>	8.5	9.6	8.6	8.7	9.3
2009	<b>7.7</b>	7.9	9.0	8.1	8.2	8.4

#### Average Error: Hitter Triples

Year	AP	BJ	CH	M	P	Z
2008	<b>1.3</b>	<b>1.3</b>	1.5	1.5	<b>1.3</b>	1.5
2009	<b>1.3</b>	<b>1.3</b>	1.4	1.5	<b>1.3</b>	1.4

#### Average Error: Hitter Home Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>5.0</b>	5.3	5.9	5.4	5.3	5.5
2009	<b>5.1</b>	5.3	5.7	5.5	5.4	5.4

#### Average Error: Hitter RBIs

Year	AP	BJ	CH	M	P	Z
2008	<b>18.3</b>	19.2	22.9	19.9	19.9	21.4
2009	<b>17.2</b>	17.7	20.9	18.9	18.3	19.8

#### Average Error: Hitter Stolen Bases

Year	AP	BJ	CH	M	P	Z
2008	<b>3.8</b>	<b>3.8</b>	4.4	4.1	4.4	4.1
2009	<b>3.8</b>	4.0	4.5	4.1	4.0	4.6

#### Average Error: Hitter Walks

Year	AP	BJ	CH	M	P	Z
2008	<b>12.3</b>	13.7	15.8	14.5	14.4	14.9
2009	<b>13.0</b>	13.5	15.6	13.9	13.6	13.9

#### Average Error: Hitter Strikeouts

Year	AP	BJ	CH	M	P	Z
2008	<b>21.7</b>	22.7	29.6	25.1	25.1	27.5
2009	21.2	<b>21.0</b>	27.7	23.6	24.1	25.2

#### Average Error: Innings Pitched

Year	AP	BJ	CH	M	P	Z
2008	34.7	37.3	40.6	35.8	<b>34.6</b>	41.5
2009	<b>29.5</b>	31.1	34.0	31.7	30.4	35.1

Average Error: Pitcher Earned Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>16.0</b>	17.2	19.5	16.9	16.4	21.1
2009	13.0	13.8	14.8	14.6	<b>12.9</b>	16.2

Average Error: Pitcher Strikeouts

Year	AP	BJ	CH	M	P	Z
2008	29.1	32.3	32.8	29.6	<b>28.8</b>	34.0
2009	26.4	29.1	29.0	27.7	<b>25.8</b>	31.0

Average Error: Pitcher Walks

Year	AP	BJ	CH	M	P	Z
2008	<b>12.3</b>	13.6	15.3	12.8	12.4	15.4
2009	10.9	12.3	12.2	11.4	<b>10.8</b>	13.0

Average Error: Pitcher Hits (Given up)

Year	AP	BJ	CH	M	P	Z
2008	34.7	37.1	41.0	36.3	<b>34.6</b>	42.7
2009	<b>27.5</b>	29.3	32.1	30.8	27.6	34.0

Root Mean Squared Error (RMSE) is the frequently-used measure of the differences between values predicted by a model or an estimator and the values actually observed from the phenomenon being modeled or estimated. RMSE is known as the best measure of accuracy for prediction models. RMSE is measured for each statistical category. The system with the smallest RMSE in each category in each year is bolded to indicate that it is the most accurate. The 2009 projections are evaluated through MLB games completed on September 20<sup>th</sup>, 2009.

RMSE: Hitter At Bats

Year	AP	BJ	CH	M	P	Z
2008	<b>144.1</b>	150.5	197.3	157.8	165.0	183.6
2009	<b>134.3</b>	139.2	174.5	148.3	151.9	166.4

RMSE: Hitter Hits

Year	AP	BJ	CH	M	P	Z
2008	<b>42.8</b>	45.1	54.3	45.8	46.9	51.7
2009	<b>39.9</b>	42.2	48.2	42.9	43.4	54.2

RMSE: Hitter Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>24.2</b>	25.9	29.3	25.2	25.9	27.3
2009	<b>21.6</b>	23.3	26.7	23.0	23.4	67.4

RMSE: Hitter Doubles

Year	AP	BJ	CH	M	P	Z
2008	<b>10.1</b>	10.6	12.0	10.5	10.8	11.6
2009	<b>9.5</b>	10.2	11.0	9.9	10.2	10.4

RMSE: Hitter Triples

Year	AP	BJ	CH	M	P	Z
2008	48.9	53.4	56.2	48.9	<b>48.5</b>	59.5
2009	39.2	42.3	46.6	42.9	<b>39.1</b>	48.5

2008	<b>1.7</b>	1.8	2.0	2.0	1.8	2.0
2009	<b>1.9</b>	<b>1.9</b>	2.0	<b>1.9</b>	<b>1.9</b>	2.0

RMSE: Hitter Home Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>6.9</b>	7.5	7.7	7.0	7.1	7.5
2009	<b>6.7</b>	7.1	7.3	7.1	7.1	7.2

RMSE: Hitter RBIs

Year	AP	BJ	CH	M	P	Z
2008	<b>23.7</b>	25.2	28.5	24.5	25.4	27.7
2009	<b>21.4</b>	23.2	25.5	23.0	23.1	24.8

RMSE: Hitter Stolen Bases

Year	AP	BJ	CH	M	P	Z
2008	<b>6.0</b>	6.5	7.1	6.5	6.8	6.9
2009	<b>6.1</b>	6.6	7.0	6.3	6.7	7.6

RMSE: Hitter Walks

Year	AP	BJ	CH	M	P	Z
2008	<b>17.0</b>	18.1	20.2	17.9	18.4	19.6
2009	<b>16.6</b>	17.8	20.1	17.7	17.7	18.0

RMSE: Hitter Strikeouts

Year	AP	BJ	CH	M	P	Z
2008	<b>29.2</b>	31.0	39.8	31.7	34.1	37.8
2009	<b>27.5</b>	28.5	27.7	29.8	33.3	34.1

RMSE: Innings Pitched

Year	AP	BJ	CH	M	P	Z
2008	48.7	53.5	54.4	48.6	<b>47.5</b>	57.2
2009	<b>41.2</b>	44.5	47.3	43.9	<b>41.2</b>	48.4

RMSE: Pitcher Earned Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>22.2</b>	24.0	26.3	22.8	22.7	28.7
2009	<b>17.9</b>	19.2	21.3	20.0	<b>17.9</b>	23.1

RMSE: Pitcher Strikeouts

Year	AP	BJ	CH	M	P	Z
2008	40.1	44.7	42.8	39.4	<b>38.8</b>	45.2
2009	35.5	39.4	39.1	37.2	<b>34.8</b>	40.8

RMSE: Pitcher Walks

Year	AP	BJ	CH	M	P	Z
2008	16.6	18.7	19.9	16.6	<b>16.3</b>	19.9
2009	14.8	16.9	16.8	15.6	<b>14.6</b>	17.9

RMSE: Pitcher Hits (Given up)

Year	AP	BJ	CH	M	P	Z
2008	48.9	53.4	56.2	48.9	<b>48.5</b>	59.5
2009	39.2	42.3	46.6	42.9	<b>39.1</b>	48.5

The Pearson correlation coefficient is a measure of the correlation (linear dependence) between two variables. The Pearson correlation coefficient is measured for each statistical category. The system with the highest Pearson correlation coefficient in each category in each year is bolded to indicate that it is the most accurate. The 2009 projections are evaluated through MLB games completed on September 20<sup>th</sup>, 2009.

2008	<b>.71</b>	.68	.54	.64	.61	.63
2009	<b>.71</b>	<b>.71</b>	.50	.64	.57	.58

## Pearson Correlation Coefficient: Innings Pitched

Year	AP	BJ	CH	M	P	Z
2008	<b>.72</b>	.70	.67	.68	.69	.67
2009	<b>.76</b>	.75	.66	.72	.75	.68

## Pearson Correlation Coefficient: Hitter At Bats

Year	AP	BJ	CH	M	P	Z
2008	<b>.68</b>	.66	.47	.59	.55	.53
2009	<b>.70</b>	<b>.70</b>	.55	.59	.56	.54

## Pearson Correlation Coefficient: Pitcher Earned Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>.73</b>	.72	.68	.70	.69	.66
2009	<b>.78</b>	.76	.67	.71	.77	.67

## Pearson Correlation Coefficient: Hitter Hits

Year	AP	BJ	CH	M	P	Z
2008	<b>.69</b>	.68	.55	.63	.59	.58
2009	<b>.70</b>	<b>.70</b>	.63	.63	.60	.57

## Pearson Correlation Coefficient: Pitcher Strikeouts

Year	AP	BJ	CH	M	P	Z
2008	<b>.69</b>	.68	.65	.66	.67	.63
2009	<b>.73</b>	.71	.66	.70	<b>.73</b>	.65

## Pearson Correlation Coefficient: Hitter Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>.70</b>	.67	.61	.64	.63	.63
2009	<b>.71</b>	<b>.71</b>	.63	.64	.63	.60

## Pearson Correlation Coefficient: Pitcher Walks

Year	AP	BJ	CH	M	P	Z
2008	<b>.71</b>	.69	.62	.66	.67	.60
2009	<b>.71</b>	.68	.61	.67	<b>.71</b>	.60

## Pearson Correlation Coefficient: Hitter Doubles

Year	AP	BJ	CH	M	P	Z
2008	<b>.66</b>	.65	.53	.60	.57	.56
2009	.65	<b>.66</b>	.57	.59	.55	.58

## Pearson Correlation Coefficient: Pitcher Hits (Given up)

Year	AP	BJ	CH	M	P	Z
2008	<b>.73</b>	.72	.69	.71	.70	.69
2009	<b>.79</b>	.78	.68	.73	.78	.70

## Pearson Correlation Coefficient: Hitter Triples

Year	AP	BJ	CH	M	P	Z
2008	<b>.64</b>	.62	.55	.55	.58	.56
2009	<b>.52</b>	<b>.52</b>	.49	.46	.47	.48

## Pearson Correlation Coefficient: Hitter Home Runs

Year	AP	BJ	CH	M	P	Z
2008	<b>.77</b>	.75	.72	.74	.73	.73
2009	.73	<b>.74</b>	.70	.69	.69	.70

## Pearson Correlation Coefficient: Hitter RBIs

Year	AP	BJ	CH	M	P	Z
2008	<b>.72</b>	.71	.63	.69	.66	.65
2009	.72	<b>.73</b>	.65	.67	.66	.66

## Pearson Correlation Coefficient: Hitter Stolen Bases

Year	AP	BJ	CH	M	P	Z
2008	<b>.79</b>	<b>.79</b>	.74	.74	.73	.74
2009	<b>.74</b>	<b>.74</b>	.69	.71	.69	.67

## Pearson Correlation Coefficient: Hitter Walks

Year	AP	BJ	CH	M	P	Z
2008	<b>.76</b>	.74	.70	.71	.70	.69
2009	<b>.73</b>	.72	.65	.68	.68	.68

## Pearson Correlation Coefficient: Hitter Strikeouts

Year	AP	BJ	CH	M	P	Z
------	----	----	----	---	---	---

## ACKNOWLEDGMENT

Ross J. Gore would like to thank Michael Spiegel for helping to hone this idea and referring us to the BellKor literature despite his "healthy distaste" for sports. We would also like to thank Chone Smith for his prompt reply to our query about the availability of the CHONE projections.

## REFERENCES

- [1] <http://www.bradnull.blogspot.com/> accessed 12 March 2009.
- [2] <http://bis-store.stores.yahoo.net/bijahapr203.html> accessed 12 March 2009.
- [3] [http://www.replacementlevel.com/index.php/RLYW/comments/cairo\\_projections\\_v01](http://www.replacementlevel.com/index.php/RLYW/comments/cairo_projections_v01) accessed 12 March 2009.
- [4] <http://fantasynews.cbssports.com/fantasybaseball/stats/sortable/points/1B/standard/projections> accessed 12 March 2009.
- [5] <http://www.baseballprojection.com/> accessed 12 March 2009.
- [6] <http://games.espn.go.com/flb/tools/projections> accessed 12 March 2009.
- [7] <http://www.actasports.com/detail.html?id=019> accessed 12 March 2009.
- [8] [http://baseballanalysts.com/archives/2009/02/2009\\_projection.php](http://baseballanalysts.com/archives/2009/02/2009_projection.php) accessed 12 March 2009.
- [9] <http://www.kffl.com/fantasy-baseball/2009-baseball-draft-guide.php> accessed 12 March 2009.
- [10] <http://www.tangotiger.net/marcel/> accessed 12 March 2009.
- [11] <http://statspeak.net/2008/11/2009-batter-projections.html> accessed 12 March 2009.
- [12] <http://www.baseballprospectus.com/pecota/> accessed 12 March 2009.
- [13] [http://www.rotoworld.com/premium/draftguide/baseball/main\\_page.aspx](http://www.rotoworld.com/premium/draftguide/baseball/main_page.aspx) accessed 12 March 2009.
- [14] <http://www.baseballthinkfactory.org/> accessed 12 March 2009.
- [15] J. Bennet and S. Lanning, "The Netflix Prize", KDD Cup and Workshop, 2007.
- [16] R. Bell, Y. Koren, and C. Volinsky, "Chasing \$1,000,000: How we won the netflix progress prize", ASA Statistical and Computing Graphics Newsletter 18(2):4-12, 2007.

- [17] <http://www.baseballprospectus.com/unfiltered/?p=564> accessed 28 July 2009.
- [18] <http://www.cs.virginia.edu/~rjg7v/aggpro/> accessed 28 July 2009.