Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

GRANULARITY OF WEIGHTED AVERAGES AND USE OF RATE STATISTICS IN AGGPRO

Timothy Highley

La Salle University
1900 West Olney Ave.
Philadelphia, PA 19141

Ross Gore

The University of Virginia
151 Engineer's Way, P.O. Box 400740
Charlottesville, VA 20052

Cameron Snapp

CapTech Ventures, Inc
1419 West Main Street
Richmond, VA 23220

ABSTRACT

AggPro predicts baseball statistics by utilizing a weighted average of predictions provided by several other statistics projection systems. The aggregate projection that is generated is more accurate than any of the constituent systems individually. We explored the granularity at which weights should be assigned by considering four possibilities: a single weight for each projection system, one weight per category per system, one weight per player per system, and one weight per player per category per system. We found that assigning one weight per category per system provides better results than the other options. Additionally, we projected raw statistics directly and compared the results to projecting rate statistics scaled by predicted player usage. We found that predicting rate statistics and scaling by predicted player usage produces better results. We also discuss implementation challenges that we faced in producing the AggPro projections.

1 INTRODUCTION

Statistics projection is an important problem for sports teams, especially when selecting players in a draft or deciding which players to offer a contract. The team's success is directly tied to the future performance of its athletes. Statistics projection is also important to the millions of people who play fantasy sports, which is a game for spectators where the set of real athletes in a sport is partitioned among the fantasy sports participants (typically through a draft or auction). The winner in fantasy sports is the person who selected athletes that accumulate the best statistics during the following year (according to some predetermined criteria) (Ballard 2001).

AggPro is an aggregate projection system. There are many different approaches to statistics projection, but AggPro does not project statistics directly. Instead, it utilizes existing projections and combines them in a way that produces a projection that is more accurate than any of the constituent projections. Previous work has demonstrated that such aggregation can provide an improvement (Gore, Snapp, and Highley 2009). In this paper, we ask whether greater improvement can be gained by assigning separate weights on a per-player basis and/or a per-category basis. The motivation for exploring the granularity of the weights is the idea that some projection systems may be better predictors for a particular type of play-

er or a particular statistic. We found that assigning weights on a per-category (but not per-player) basis performed best.

Some statistics are almost entirely a function of how the player is used and how healthy the player is (e.g. innings pitched and plate appearances in baseball). These are qualitatively different from statistics that indicate the skill of a player. In addition to projecting raw statistics directly, we used a second approach where we treated usage statistics and skill-based statistics differently; skill-based statistics were treated as rate statistics. We projected the rate statistics and scaled by the projected player usage to produce raw statistics. We compare this result to the results of projecting the raw statistics directly. Our evaluation shows that projecting rate statistics and scaling by projected player usage, used along with per-category weight sets, provides a significant improvement over the other projections we evaluated: the constituent projection systems and AggPro projections formed exploring the other weight set granularities (including our prior work).

We also discuss optimization efforts that we used to improve the speed of the program. The optimization used to find the best weighted averages is similar to finding the best values for flexible points in coercible simulations (Waziruddin, Reynolds, and Brogan 2003).

2 BACKGROUND

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

2.1 Projection Systems

Many different methods exist for projecting the performance of Major League Baseball (MLB) players in a variety of statistical categories for an upcoming MLB season. These projection systems include: Brad Null (Baseball Calculus 2010), Bill James Handbook (Baseball Info Solution 2010), CAIRO (Replacement Level Yankees Weblog 2010), CBS (CBS Sports 2010), CHONE (Baseball Projection.com 2010), Marcel (Marcel 2010), PECOTA (Baseball Prospectus 2010), and ZiPS (Baseball Think Factory 2010). Although there are a large number of systems and they are widely available, there has been relatively little evaluation on the accuracy of the predictions these systems produce. In prior work, it has been demonstrated that AggPro can successfully improve on existing projections by identifying a beneficial weighting of constituent systems (Gore, Snapp, and Highley 2009).

2.2 Silver's 2007 Evaluation of Projection Systems

In 2007 Nate Silver performed an evaluation of the on-base percentage plus slugging (OPS) statistic projection from eight 2007 MLB projection systems (Baseball Prospectus 2007). Silver's work uses several evaluation metrics including average error, RMSE and Pearson's correlation coefficient. We employ each of these to evaluate AggPro. However, Silver also offers a metric based on performing a regression analysis on all the systems for the past year. The metric ranks the quality of the information provided by the systems. This metric is similar to the weights AggPro calculates to form projections for an upcoming year.

2.3 Combining Data

The problem of combining data to produce new insights or provide better accuracy is common. Such data fusion is an important problem in wireless sensor networks (Nakamura, Loureiro, and Frery 2007), information retrieval (Efron 2009), and determining the degree of membership in a fuzzy database (Cunningham 2006), among others.

The strategy of applying different weights to different predictions from effective projection systems has been used successfully by the designers of the winning solution for the NetFlix prize, BellKor Pragmatic Chaos by AT&T labs (Bell, Korin, and Volinsky 2007). In October 2006, Netflix released a dataset

of anonymous movie ratings and issued a challenge to researchers: develop a system 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 Pragmatic Chaos prediction system was 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 (Bell, Korin, and Volinsky 2007). AggPro applies this strategy to projecting the performance for MLB players.

Horowitz explored weights for college football rankings. He argued that predisposition to a single set of weights for aggregated ratings is a bad idea (Horowitz 2004), and that it is appropriate to calculate a new set of weights whenever the experts receive new information.

2.4 COERCE

When constructing a model, abstractions inevitably must be selected in order to reduce complexity, improve performance, or provide estimations for unknown information. When developing *coercible* simulations a subject matter expert (SME) identifies a set of *abstraction opportunities and alternatives* for each model abstraction. A flexible point of a simulation reflects one model abstraction opportunity and the corresponding bindings for the flexible point reflect abstraction alternatives. The developer and the SME use optimization (automatic function minimization) and/or manual modification to find new bindings for the selected flexible points. The SME may interrupt this step if it becomes apparent that a satisfactory set of bindings will not be found. This process repeats until the new requirement is met (Waziruddin, Reynolds, and Brogan 2003). The goal of AggPro is not to adapt existing models to new requirements. However, the use of optimization to identify the set of weights that minimize error from actual player performance statistics is similar to the process of finding the best values for flexible points in COERCE, which is a simulation technology designed to aid in the development of coercible simulations. The authors' prior experience with COERCE was a major part of the motivation for AggPro.

3 EXPERIMENTAL METHODOLOGY

First, we collected the projections from five different systems for the years 2007, 2008 and 2009 and converted them to a common format. The projection systems that we used were Bill James (B), Chone (CH), Marcel (M), Pecota (P), and Zips (Z). Next, for each year we identified the players that were common among all five systems. The player list for each year is available at AggPro (2010). Then we identified the statistical categories that were common among all five projection systems. The hitter categories common to the five systems are: At Bats, Plate Appearances, 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.

AggPro generated projections for this set of common players and statistics by using weighted sums of the projections from the five constituent systems. The primary purpose of this study was to identify the best granularity to use when assigning weights. A weight set consists of one weight for each constituent projection system, where each weight x_i is constrained such that $0 \le x_i \le 1$ and $\sum x_i = 1$. We examined four different approaches to granularity.

1. One weight set is used for all projections, regardless of the player or statistical category. This is the approach used in (Gore, Snapp, and Highley 2009).
2. A different weight set is used for each statistical category. For each statistical category, the corresponding weight set is used to project that statistic for all players. We expected this approach to work well if some constituent projection systems were consistently good at predicting certain statistics while other projection systems were better at predicting other statistics.

3. A different weight set is used for each player. Each player's weight set is used to project all statistics for that player. We expected this approach to work well if some constituent projection systems were consistently good at predicting statistics for certain types of players while other projection systems were better for different types of players.
4. For each combination of player and statistical category, a different weight set is generated and used. This is a combination of the approaches in numbers 2 and 3, and we expected that it would work under a combination of the circumstances described in numbers 2 and 3.

In each case, we computed weight sets that minimized the root mean square error for the statistics in question during one year, and used those weight sets to predict statistics for the following year. (We had converted statistics for years 2007, 2008, and 2009 to our standard format, and we generated projections for 2008 and 2009.) We generated statistics using each of the four approaches above and compared the resulting aggregate projections with the constituent projection systems.

3.1 Raw Statistics versus Rate Statistics

All of the statistics identified above are raw accumulation statistics. They do not include computed statistics such as batting average or earned run average. Because all of the statistics are raw accumulation statistics, it is possible that some constituent projection systems may be heavily weighted not because they do a good job of predicting how well a player will perform, but only because they predict player usage well (e.g. whether the player will be in the starting lineup, whether the player will be injured, or whether a pitcher will be a starter or a reliever).

In order to separate the projection of player usage from the projection of player quality, we introduced a second approach. We defined rate statistics corresponding to each of the raw statistics identified above. For the hitters, we defined the rate statistics by dividing each of the raw statistical categories by Plate Appearances, where Plate Appearances is computed as At-Bats plus Walks. For the pitchers, we defined the rate statistics by dividing each of the raw categories by Innings Pitched. AggPro generated weight sets using each of the four approaches to granularity described above, but minimized the root mean square error over the rate statistics instead of the raw statistics. To generate the AggPro projections using rate statistics, we used the weight sets to predict the rate statistics, and then multiplied each predicted rate statistic by the predicted value for Plate Appearances or Innings Pitched, as appropriate. The resulting raw statistics were compared with the other AggPro projections and the constituent projection systems.

In total, AggPro generated eight sets of projections, based on the four approaches to weight set granularity. Each approach was used first to project raw statistics directly, as well as to project rate statistics which were then scaled.

4 OPTIMIZATION

Recall that previous research with AggPro only identified one weight set that was applied to all common statistical categories among the five projection systems (Gore, Snapp, and Highley 2009). The weight set minimized the RMSE of the aggregate projections from the previous year's statistics when applied to the projections of the five systems for the previous year. The weight set was identified through a brute force search across all weight sets, up to two decimal places. (Recall that a weight set consists of weights, where each weight x_i is constrained such that $0 \le x_i \le 1$ and $\sum x_i = 1$.) This approach requires exponential time in terms of the number of constituent projection systems, making it computationally inefficient (Gore, Snapp, and Highley 2009). Our new approach remedies this inefficiency. The approach views each weight, x_i, as a variable in a linear equation and each projection of a constituent system in each statistical category as the coefficient of the respective variable. For each player and each statistical category a linear equation is formed using two matrices, *x* and *C*. The matrix *x* contains the variables of the linear equation. Each of these variables is the weight for each of the x_i constituent projection systems where $i = B, CH, M, P$ or Z. The matrix *C* contains the coefficients of each variable. Each of the coeffi-

cients, C_{iq}, is the constituent projection for the player/category q in projection system i. Each linear equation formed from C and x is set equal to the actual performance data, d. The actual performance data for player/category q is d_q. Figure 1 illustrates the structure of each of these matrices.

These three matrices (C, x and d) form a system of linear equations, where the values within the matrices depends on which of the four options for weight set granularity is used. The system of linear equations is then solved using least squares linear regression, which only requires polynomial time in terms of the number of constituent projection systems (m) and the number of players within the projection systems (n), specifically, $O(\max(n,m)^3)$ (Matlab 2009). We ensure the same constraints, $0 \le x_i \le 1$ and $\sum x_i = 1$, as in our previous approach (Gore, Snapp, and Highley 2009).

$$\begin{bmatrix} C_{B1}, C_{CH1}, C_{M1}, C_{P1}, C_{Z1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ C_{Bq}, C_{CHq}, C_{Mq}, C_{Pq}, C_{Zq} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ C_{Bn}, C_{CHn}, C_{Mn}, C_{Pn}, C_{Zn} \end{bmatrix} * \begin{bmatrix} x_B, x_{CH}, x_M, x_P, x_Z \end{bmatrix} = \begin{bmatrix} d_1 \\ \ldots \\ d_q \\ \ldots \\ d_n \end{bmatrix}$$

Figure 1: The matrices required to solve the linear system of equations (Matlab 2009).

For Option 1 of AggPro, where one set of weights is applied to all statistical categories for all players in the constituent projection systems, one linear system of equations was solved. C contained an entry for all players in all categories from the constituent projection system. Due to the required symmetry, d contained an entry for all players in all statistical categories from the actual performance data.

For Option 2 of AggPro, where a different weight set is applied to each statistical category (but is used for all players in the constituent projection systems), $|L|$ linear systems of equations were solved, where L is the set of statistical categories common to the constituent projection systems, and $|L|$ is the number of statistical categories common to the constituent projection systems. Given a statistical category L_k, C contained an entry for all players for L_k. Due to the required symmetry, each d contained an entry for all players for statistical category L_k from the actual performance data.

For Option 3 of AggPro, where a different weight set is applied to each player (but is used for all statistical categories in the constituent projection systems), $|G|$ linear systems of equations were solved, where G is the set of players common in the constituent projection systems, and $|G|$ is the number of players common to all constituent projection systems. Given a player G_j, C contained an entry for the projections for player G_j from all the constituent projection systems for all statistical categories. Due to the required symmetry, each d contained an entry for all statistical categories for player G_j from the actual performance data.

For Option 4 of AggPro, where a different weight set is applied to each player for each statistical category in the constituent projection systems, $|G|$ * $|L|$ linear systems of equations were solved. Given a player G_j and a statistical category L_k, C contained an entry for the projections for player G_j for the statistical category L_k from all constituent projection systems. Due to the required symmetry, each d contained an entry for player G_j for the statistical category L_k from the actual performance data.

5 EVALUATION

AggPro generated eight sets of projections. There are four approaches to weight set granularity, and each approach was used to first predict raw statistics directly. These four projections are listed as aggPro_raw1 through aggPro_raw4 in the tables below. Next, each weight set granularity was used to generate a projection by predicting rate statistics and scaling by predicted player usage. These four projections are listed as aggPro_rate1 through aggPro_rate4 in the tables below. When these eight sets of projections are evaluated along with the five constituent projection systems, there are thirteen projections in total.

We compared all of the projections using the following criteria: average error, the Pearson correlation coefficient, and RMSE. We separated the statistical categories into two groups: fifteen raw statistics and thirteen rate statistics. There are two fewer rate statistics because we do not include rate statistics for Plate Appearances or Innings Pitched, since those two are the basis for computing rate statistics. (It is not useful to compute Plate Appearances per Plate Appearance or Innings Pitched per Inning Pitched.)

For each statistical category, we ranked the thirteen projections according to each criterion. Using each criterion in turn, we found the average ranking for each projection across all of the raw statistical categories ("Average Raw" below) and across all of the rate statistical categories ("Average Rate" below). For "Average Rate" and "Average Raw," lower numbers are better with 1 being the best possible and 13 being the worst possible since there are thirteen projection systems.

For each evaluation criterion, we counted the number of statistical categories for which each projection was identified as the best predictor. "Raw First" is the number of raw statistics (out of the 15) for which the projector was the best predictor. "Rate First" is the number of rate statistics (out of the 13) for which the projector was the best predictor. In the case of ties for first place, the tied projections were both given credit.

Finally, we counted the number of statistical categories for which each projection ranked in the top half of the thirteen projections ("Raw Top Half" and "Rate Top Half" for the raw statistics and rate statistics, respectively). The maximum value for "Raw Top Half" is 15 categories, and the maximum value for "Rate Top Half" is 13 categories.

We examined projections for the 28 statistical categories for the years 2008 and 2009, comparing the projections using three evaluation criteria, thus yielding a total of 168 evaluations of the projections. (The 168 evaluations are not independent.) Overall, we found that aggPro_rate2 performed the best. The aggPro_rate2 projection uses a separate weight set for each statistical category, and makes its predictions by predicting rate statistics and scaling them by predicted player usage. Both the predicted rate statistics and the predicted player usage are based on the constituent projection systems. The aggPro_rate2 projection produced the best results for 43 out of 90 comparisons for raw statistics, and for 31 out of 78 comparisons for rate statistics. The aggPro_rate2 projections performed well more consistently than any of the other projection systems: it ranked in the top half of projections for 88 of the 90 comparisons for raw statistics, and for 75 out of 78 comparisons for rate statistics.

AggPro options 3 and 4, where weight sets are derived for each player individually, consistently perform worse than AggPro options 1 and 2, where all players use the same weight sets. Depending on the year and evaluation criteria, options 3 and 4 sometimes rank worse than some of the constituent projection systems. It is reasonable to expect that a weight set targeted to a specific player may identify constituent projection systems that are particularly adept at predicting statistics for that specific player. It is also reasonable to expect such a weight set to perform poorly since far less information is taken into consideration. Our results indicate that the latter factor is more important than the former. This is at least true for our approach, which only looks at the previous year to determine how to apply the weight sets. Looking at more history may make per-player weights more useful.

By separating raw statistics and rate statistics, it becomes apparent that the billJames projections do a very good job at projecting player usage. The billJames projection was determined to be the best of the five constituent projection systems in (Gore, Snapp and Highley 2009). However, the chone and pecota projections do just as well at predicting rate statistics, and were given correspondingly higher weights

when AggPro was projecting rate statistics. The weights for player usage use a heavy weight for the bill-James projections, similar to our original findings.

Although we find aggPro_rate2 to be the best projector in our experiments, it is not the winner by a large margin. In (Gore, Snapp, and Highley 2009), the aggPro_raw1 projection was shown to be an improvement over the best constituent projection system by 0.7% to 7.2%, depending on the year and evaluation criterion. According to most of the evaluation criteria we employed, the aggPro_rate2 projection presented here is a further improvement. For the 2009 projections, aggPro_rate2 improved the average error by 0.7%, the RMSE by 0.5%, and the Pearson correlation coefficient by 0.1% over aggPro_raw1. For the 2008 projections, aggPro_rate2 improved the average error by 1.9% and the RMSE by 1.2% over aggPro_raw1. However, for 2008 the Pearson correlation coefficient for aggPro_rate2 was 0.5% worse than that for aggPro_raw1.

5.1 Average Error

Table 1: Average Error Evaluation of the thirteen projections for 2008.

	Average Error: 2008					
Projection	**Average Raw**	**Raw First**	**Raw Top Half**	**Average Rate**	**Rate First**	**Rate Top Half**
billJames	8.3	0	3	9.1	1	3
Chone	12.6	0	0	5.3	1	8
Marcel	9.1	0	5	9.6	1	3
Pecota	8.1	0	6	4.4	0	11
Zips	12.1	0	0	7.4	2	4
aggPro_rate1	3.7	2	14	4.8	0	9
aggPro_rate2	*1.8*	*9*	*15*	*3.0*	*6*	*12*
aggPro_rate3	7.9	0	3	8.3	0	3
aggPro_rate4	6.2	1	9	6.5	0	8
aggPro_raw1	3.5	0	14	5.3	1	8
aggPro_raw2	2.4	5	*15*	6.2	1	7
aggPro_raw3	7.7	0	3	8.4	0	2
aggPro_raw4	7.1	0	4	12.8	0	0

Table 2: Average Error Evaluation of the thirteen projections for 2009.

	Average Error: 2009					
Projection	**Average Raw**	**Raw First**	**Raw Top Half**	**Average Rate**	**Rate First**	**Rate Top Half**
billJames	8.4	0	5	9.7	0	2
chone	12.3	0	0	5.5	2	9
marcel	10.5	0	1	10.0	0	2
pecota	8.0	0	5	5.7	0	10
zips	12.3	0	0	9.2	0	2
aggPro_rate1	2.9	3	14	3.9	0	11
aggPro_rate2	*2.6*	*7*	14	*2.1*	*7*	*13*
aggPro_rate3	7.0	1	4	6.6	0	5
aggPro_rate4	6.9	0	7	8.8	0	1
aggPro_raw1	2.8	3	14	4.3	0	11
aggPro_raw2	2.9	3	*15*	5.9	4	5

aggPro_raw3	7.2	0	4	6.5	0	7
aggPro_raw4	6.7	0	8	12.7	0	0

5.2 Correlation Coefficient

Table 3: Correlation Coefficient Evaluation of the thirteen projections for 2008.

	Correlation Coefficient: 2008					
Projection	**Average Raw**	**Raw First**	**Raw Top Half**	**Average Rate**	**Rate First**	**Rate Top Half**
billJames	5.7	0	11	5.8	0	9
chone	12.3	0	0	12.3	0	0
marcel	10.5	0	0	10.5	0	0
pecota	10.6	0	1	10.6	0	1
zips	12.2	0	0	12.2	0	0
aggPro_rate1	2.4	*5*	*15*	2.6	3	*13*
aggPro_rate2	2.8	4	*15*	2.8	*4*	*13*
aggPro_rate3	7.8	0	3	7.8	0	3
aggPro_rate4	6.5	0	9	6.6	0	7
aggPro_raw1	*2.1*	*5*	*15*	*2.3*	3	*13*
aggPro_raw2	2.5	3	*15*	2.4	3	*13*
aggPro_raw3	7.5	0	3	7.5	0	3
aggPro_raw4	7.4	0	5	7.6	0	3

Table 4: Correlation Coefficient Evaluation of the thirteen projections for 2009.

	Correlation Coefficient: 2009					
Projection	**Average Raw**	**Raw First**	**Raw Top Half**	**Average Rate**	**Rate First**	**Rate Top Half**
billJames	4.9	2	13	4.9	2	11
chone	10.9	1	1	10.7	1	1
marcel	10.9	0	0	11.1	0	0
pecota	9.1	0	5	9.3	0	4
zips	11.4	0	0	11.2	0	0
aggPro_rate1	3.3	2	14	3.4	1	12
aggPro_rate2	*2.6*	*7*	*15*	*2.7*	*6*	*13*
aggPro_rate3	7.6	0	1	7.6	0	1
aggPro_rate4	8.1	0	6	8.1	0	5
aggPro_raw1	2.9	1	*15*	3.0	0	*13*
aggPro_raw2	3.1	4	*15*	3.2	3	*13*
aggPro_raw3	7.5	0	2	7.5	0	2
aggPro_raw4	8.2	0	4	8.2	0	3

5.3 RMSE

Table 5: RMSE Evaluation of the thirteen projections for 2008.

	RMSE: 2008					
Projection	**Average Raw**	**Raw First**	**Raw Top Half**	**Average Rate**	**Rate First**	**Rate Top Half**
billJames	9.8	0	1	10.2	0	1
chone	12.5	0	0	5.9	2	8
marcel	7.7	0	5	10.0	0	1
pecota	8.1	0	6	*3.1*	*4*	*12*
zips	12.4	0	0	7.8	1	3
aggPro_rate1	3.5	3	*15*	4.8	2	10
aggPro_rate2	*1.9*	*8*	*15*	3.2	2	11
aggPro_rate3	7.5	0	2	7.6	0	5
aggPro_rate4	6.5	0	8	6.5	1	6
aggPro_raw1	3.5	1	*15*	5.1	0	10
aggPro_raw2	2.3	5	*15*	6.5	1	6
aggPro_raw3	7.6	0	2	7.5	0	5
aggPro_raw4	7.2	0	6	12.8	0	0

Table 6: RMSE Evaluation of the thirteen projections for 2009.

	RMSE: 2009					
Projection	**Average Raw**	**Raw First**	**Raw Top Half**	**Average Rate**	**Rate First**	**Rate Top Half**
billJames	10.0	0	1	10.3	0	0
chone	12.1	0	0	5.0	2	9
marcel	9.1	0	1	10.2	0	2
pecota	8.3	0	5	4.7	0	9
zips	12.5	0	0	8.8	0	3
aggPro_rate1	3.1	2	*15*	4.1	0	11
aggPro_rate2	2.5	*8*	14	*2.4*	*6*	*13*
aggPro_rate3	6.7	0	7	7.0	0	5
aggPro_rate4	7.4	0	7	8.2	0	3
aggPro_raw1	2.6	3	*15*	4.4	2	11
aggPro_raw2	*2.1*	4	*15*	6.2	3	7
aggPro_raw3	6.7	0	7	6.8	0	5
aggPro_raw4	7.5	0	5	12.8	0	0

5.4 Improvement Over Previous Projections

Table 7: Percent Improvement of aggPro_rate2 vs. aggProraw for 2008 and 2009.

Percent Improvement: aggPro_rate2 vs. aggPro_raw1

Category	Average Error		RMSE		Correlation Coefficient	
	2008	2009	2008	2009	2008	2009
AB	0.3%	0.9%	-0.4%	0.2%	-0.2%	0.7%
H(H)	2.1%	1.6%	1.5%	1.4%	0.1%	0.8%
R	0.8%	0.7%	0.8%	-0.2%	0.3%	0.4%
D	2.5%	1.7%	2.5%	1.6%	-0.1%	0.3%
T	-2.8%	1.4%	-4.5%	0.5%	-4.1%	0.6%
HR	0.6%	0.5%	1.8%	-0.7%	0.3%	-0.4%
RBI	0.7%	0.9%	1.7%	0.6%	0.6%	0.6%
SB	4.7%	-0.6%	3.0%	-2.3%	0.0%	-0.7%
BB(H)	1.3%	-0.2%	1.0%	0.2%	1.0%	0.7%
K(H)	-1.2%	-2.8%	-2.1%	-3.4%	-0.8%	-1.8%
IP	1.5%	-2.2%	4.3%	0.0%	-0.6%	-0.2%
ER	0.0%	2.3%	1.4%	1.4%	-1.3%	0.7%
K(P)	3.3%	0.2%	5.4%	-0.2%	-0.5%	-0.5%
BB(P)	1.8%	1.9%	4.9%	2.0%	-0.9%	0.5%
H(P)	1.4%	0.0%	2.7%	1.0%	-1.3%	0.2%
H(H)/PA	4.5%	3.0%	2.5%	3.3%	0.1%	0.8%
R/PA	-0.1%	0.2%	0.1%	0.8%	0.3%	0.4%
D/PA	3.2%	1.5%	2.0%	1.7%	-0.1%	0.3%
T/PA	-1.6%	2.5%	-0.7%	0.5%	-4.1%	0.6%
HR/PA	0.5%	0.0%	0.1%	-0.6%	0.3%	-0.4%
RBI/PA	0.7%	1.9%	1.6%	1.4%	0.6%	0.6%
SB/PA	1.9%	1.1%	-1.1%	-2.8%	0.0%	-0.7%
BB(H)/PA	1.2%	1.4%	0.5%	1.3%	1.0%	0.7%
K(H)/PA	3.4%	-0.6%	3.1%	2.1%	-0.8%	-1.8%
ER/IP	1.7%	1.3%	1.5%	2.6%	-1.3%	0.7%
K(P)/IP	2.0%	0.7%	1.0%	-1.1%	-0.5%	-0.5%
BB(P)/IP	-0.3%	-0.1%	0.2%	0.7%	-0.9%	0.5%
H(P)/IP	-0.8%	-0.1%	-0.1%	1.9%	-1.3%	0.2%
AVERAGE	**1.2%**	**0.7%**	**1.2%**	**0.5%**	**-0.5%**	**0.1%**

6 CONCLUSION AND FUTURE WORK

We have presented an algorithm for aggregate projections of sports statistics that provides results that are better than any of the constituent projections. We explored the question of weight set granularity and determined that examining each statistical category individually can yield improved results, but examining each player individually can cause a loss of accuracy, since a far smaller dataset is used in generating the weight set. Because we generated many different weight sets, it became necessary to find a more effi-

cient process to identify the weight sets. We have described the optimization technique we employed, and have enabled AggPro to work with a large number of constituent projection systems.

In future work, we plan to explore utilizing the AggPro projections for simulated fantasy baseball drafts in combination with empirical Average Draft Position (ADP) data to create an overall draft value metric for each projected player.

ACKNOWLEDGMENTS

Ross J. Gore would like to thank Michael Spiegel for helping to hone this idea and referring us to the BellKor Pragmatic Chaos literature despite his "healthy distaste" for sports.

REFERENCES

AggPro 2010. Additional AggPro Information. Available via <http://www.cs.virginia.edu/~rjg7v/aggpro/> [accessed April 6, 2010].

Ballard, C. 2004. Fantasy world. *Sports Illustrated* 100(27): 64-70.

Baseball Calculus 2010. MLB 2010 Season Predictions. Available via <http://www.baseballcalculus.com/> [accessed April 6, 2010].

Baseball Prospectus 2007. 2007 Hitter Roundup. Available via <http://www.baseballprospectus.com/unfiltered/?p=564> [accessed April 6, 2010].

Baseball Info Solution 2010. Bill James Handbook: 2010 Projections. Available via <http://www.baseballprospectus.com/unfiltered/?p=564> [accessed April 6, 2010].

Baseball Projection.com 2010. 2010 Team Pages. Available via <http://www.baseballprojection.com> [accessed April 6, 2010].

Baseball Prospectus 2010. Baseball Prospectus 2010 PECOTA Player Cards. Available via <http://www.baseballprospectus.com/card/index.php?mode=pecota> [accessed April 6, 2010].

Baseball Think Factory 2010. ZiPS 2010 Projections. Available via <http://www.baseballthinkfactory.org> [accessed April 6, 2010].

Bell, R., Y. Koren, and C. Volinsky. 2007. Chasing $1,000,000: How We Won the Netflix Progress Prize. *ASA Statistical and Computing Graphics Newsletter* 18(2):4–12.

CBS Sports 2010. Fantasy Baseball Stats: 2010. Available via <http://fantasynews.cbssports.com/fantasybaseball/stats/sortable/points/1b/standard/projections> [accessed April 6, 2010].

Cunningham, J. 2006. Determining an Optimal Membership Function Based on Community Consensus in a Fuzzy Database System. In *Proceedings of the 44th Annual Southeast Regional Conference* (Melbourne, Florida, March 10 - 12, 2006). ACMSE 44.

Efron, M. 2009. Generative Model-Based Metasearch for Data Fusion in Information Retrieval. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries* (Austin, TX, USA, June 15 - 19, 2009). JCDL '09.

Gore, R., C. Snapp, and T. Highley. 2009. AggPro: The Aggregate Projection System. *Baseball Research Journal* 38(2): 20-25.

Horowitz, I. 2004. Aggregating Expert Ratings Using Preference-Neutral Weights: The Case of the College Football Polls. *Interfaces* 34(4):314-320.

Marcel 2010. Marcel 2010 Forecasts. Available at <http://www.tangotiger.net/marcel/> [accessed April 6, 2010].

Matlab 2009. Solve linear least-squares problems – MATLAB. Available via <http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/lsqlin.html> [accessed April 6, 2010].

Nakamura, E., A. Loureiro, and A. Frery. 2007. Information Fusion for Wireless Sensor Networks: Methods, Models, and Classifications. *ACM Computing Surveys* 39(3).

Replacement Level Yankees Weblog 2010. 2010 CAIRO Projections v1.0. Available at <http://www.rlyw.net/stuff/sg/cairo2010.htm> [accessed April 6, 2010].

Waziruddin, S., P.F. Reynolds and D.C. Brogan. 2003. The Process for Coercing Simulations. *Proceedings of the Fall 2003 Simulation Interoperability Workshop*, Simulation Interoperability Standards Organization. Orlando, FL.

AUTHOR BIOGRAPHIES

TIMOTHY HIGHLEY is an Assistant Professor of Computer Science at La Salle University. His email address is <highley@lasalle.edu>.

ROSS GORE is a PhD candidate in Computer Science at The University of Virginia. His email address is <rjg7v@virginia.edu>.

CAMERON SNAPP is a Senior Consultant at CapTech Ventures, Inc in Richmond, VA. His email address is <cameron.snapp@gmail.com>.