# Some Solutions for Baseball Manager's Problems: Choosing a Set of Starters in Their Fielding Positions

**Marina V. Polyashuk**
Northeastern Illinois University
Chicago, IL 60625, USA
m-polyashuk@neiu.edu

Department of Mathematics
Northeastern Illinois University
Chicago, IL 60625

**Technical Report # 15-1206**

December 2015

# Some Solutions for Baseball Manager's Problems: Choosing a Set of Starters in Their Fielding Positions

**Marina V. Polyashuk**
Northeastern Illinois University
Chicago, IL 60625, USA
m-polyashuk@neiu.edu

Summary: *We will discuss various decision making situations facing a Major League Baseball (MLB) team manager in his everyday work such as selection of the 9-player set for a batting lineup, deciding on the batting order, and selecting a substitute player while a game is in progress. More specifically, the paper focuses on the task of selecting a set of starters for a given game as a bi-criteria problem.*

## 1. Introduction

If you, by chance, decided to use the key-word "baseball" in your search of a scientific database, you would be buried under thousands upon thousands of titles. These titles cover all possible aspects of this exhilarating game, from the absorption of moisture by baseball jerseys to the trajectories of fly balls, and to the comparative importance of pitching and hitting in winning a baseball game. And then, of course, there is sabermetrics (derived from SABR, Society for American Baseball Research), which became known to the general audience of recent movie-goers through "Moneyball". Most mathematicians, let alone statisticians, are aware that baseball is an ultimate breeding ground for statistical studies and analyses. In short, baseball is not just an "America's favorite pastime", but also an inspiration for numerous research and business ideas. These ideas allow many people to have fun not only by watching the game, but by analyzing, exploring, and predicting its outcomes.

Clearly, for professionals who are directly involved in baseball, it is not only "fun", but also hard work and for some this work is not physical but mental, requiring full concentration and an ability to make timely decisions that are yielding good results. Yes, the subject of this paper is focused on the Major League Baseball (MLB) team managers, people who are usually not talked about when things are going well for their baseball club. All of us who follow baseball have some anecdotal evidence of how good or bad managers appear to be at what they do: a player who excelled after being moved to a different field position or a "miraculous" game-winning home run by a pinch-hitter or a team going from first to last place after a manager had been replaced. I am sure that each of the fans has an opinion on the manager of their favorite team. There are also publications that attempt to evaluate managers' performance, such as annual editions of the Baseball Prospectus, The Bill James Handbook, and various internet analyses. It is difficult to separate the manager's contribution from his players' and as a result the managers' statistics are very much intertwined with those of the team players. We will not attempt to propose criteria for managers' evaluation; instead, we will try to analyze managers' tasks that define his performance.

A Major League Team manager (whom we will be referring to as manager from now on) does not make most important personnel decisions. Manager does not hire players, does not solely decide how many pitchers versus field players will be on their 25-man roster, does not put players on the trading block or pursue trades for specific players from other teams, nor sends players to the Minor League. In short, a manager does not form a team that he manages. Therefore, he, at least explicitly, does not participate in the most important part of decision making, which would ultimately affect the team's performance in any given season. However, it would be wrong to conclude that manager is entirely out of the decision making process and that he does not make a difference. After all, team manager decides on the following issues prior and during every of the 162 games of the season:

1. Before the game, the manager decides which players will start the game in each of the 9 field positions and as the designated hitter (the latter, in the American League only).
2. Before the game, the manager decides on the order in which 9 selected players will bat, which is referred to as lineup or batting order.
3. During the game, the manager decides on player substitutions in pitching (calls for the bullpen), running (pinch-runners), and hitting (pinch-hitters).

Note that usually the first two tasks are viewed as the single problem of forming a lineup (as in, for example, Sugrue P.K. and Mehrota A., 2007). However, as we will be able to see further, splitting this problem into two tasks stated above allows for more efficient decision making models. It is also important to mention other, lesser decisions that managers make every day: making a call for a sacrifice bunt or an intentional walk, signaling to steal a base or hit-and-run, etc., however manager's contribution to the team's success or failure is mostly affected by his performance while fulfilling the above three tasks. In this paper, the author, who is also a baseball fan, will make a humble attempt of suggesting enhanced solutions for the first of the three most important everyday problems listed above.

After introducing necessary terms and definitions, we will go over some examples of managerial decisions as well as samples of research related to the discussed types of decisions. In this paper, we will focus on selection of the 9-player set of starters on any given game day. This problem will be formulated as a bi criteria portfolio selection problem with such criteria as defensive quality of the set and offensive potential of the set. We will then attempt to compare results of the model applications with the results of actual managerial decisions. And because this paper is about baseball, writing it is a lot of fun for the author – all the way through the last inning.

## 2. Common baseball terms and notations

At this point, we will assume that a reader has basic knowledge of the baseball rules and is familiar with the game in general. Overwhelming majority of the 25 terms (we did not include any terms reflecting on the pitchers' performance) listed below is widely accepted in baseball related literature. There are slight differences in how some of the terms are defined by different authors; in those cases, we provide a reference to the source. All of the terms below refer to an individual player's performance.

1. *AB*: the number of at-bats;
2. *BB:* the number of walks;
3. 1*B*: the number of singles;
4. 2*B*: the number of doubles;
5. 3*B*: the number of triples;
6. *HR*: the number of home runs;
7. *SH*: the number of sacrifice hits;
8. *PA*: the number of plate appearances, $PA = AB + BB + HBP + SF$ ;
9. *H*: the total number of hits, $H = 1B + 2B + 3B + HR$;
10. *RBI:* runs batted in;
11. *SB*: the number of stolen bases;
12. *CS*: the number of times caught stealing;
13. *BA*: batting average, $BA = \frac{H}{AB}$;
14. *TB:* total bases, $TB = H + 2B + 2 \cdot 3B + 3 \cdot HR = 1 \cdot 1B + 2 \cdot 2B + 3 \cdot 3B + 4 \cdot HR$;
15. *SLG:* slugging average, $SLG = \frac{TB}{AB}$ ;
16. *OBP:* on-base percentage, $OBP = \frac{H + BB + HBP}{PA}$;
17. *OPS:* on-base percentage plus slugging average, $OPS = OBP + SLG$;
18. *RC:* the number of runs created, $RC = \frac{(H + BB) \cdot TB}{AB + BB}$ (Albert and Bennett, 2003);
19. *RC/G*: the number of runs created by a player per game,
    $RC / G = \frac{RC \cdot 27}{AB - H + SH + SF + CS + GDP}$ (Albert and Bennett, 2003);
20. *PO:* the number of defensive outs (putouts);
21. *A:* the number of defensive plays assisting an out (assists);
22. *DP*: the number of double plays turned;
23. *E:* the number of fielding errors;
24. *FP*: fielding percentage, $FP = \frac{PO + A}{PO + A + E}$ ;
25. *DRS:* the number of defensive runs saved (Baseball Info Solutions (BIS) and Bill James, 2011).

The above is an incomplete list of baseball measurements, which are used to evaluate defensive and offensive performance of batters and teams and which are carefully accumulated, computed, and analyzed, by MLB, SABR, and many other organizations and individuals. Clearly, the items on the list can be categorized as offensive or defensive measurements. It is important to keep in mind that there are many other, more refined, measurements that are seriously considered when comparing/evaluating players and teams. For example, *BA* and *OBP* can be considered based on right- and left-handedness, on grass vs. turf, with or without runners in scoring position, at home and on the road; defensive statistics are considered for each position, which any given player has taken on the field, etc.

## 3. Player selection as a multiple criteria problem

### 3.1. Preliminaries

Before every game, the manager has to select 10 players (in American League, AL) or 9 players (in National League, NL) to start the game.  Since the examples in this paper will be based on the teams in AL, we will focus on the 10-player selection.  In addition, the question of the starting pitcher is usually a separate issue of pitcher rotation and, short of injuries or emergency situations, is determined in advance.  Therefore, we will consider the player selection problem as the task of choosing 9 payers from *N* non-pitching members of the team's active roster that will be in the starting lineup.


**Figure 1** (Wikipedia)

Basically, the manager is supposed to select a permutation of 9 players: one DH (who replaces the pitcher in the batting lineup), one catcher, one IB, and so on, in the order of the standard defensive position numbers chart (see Fig. 1).  These 9 players have to be selected from the set of *n* players who are non-pitching staff members on the team.  In 2012 among 14 AL teams, seven teams have 12 such players while another seven teams have 13. This means that the total number of selections is between $P(12,9) = 79,833,600$ and $P(13,9) = 259,459,200$, which would be a difficult choice.  In reality, however, the number of alternatives is not even close.  The fact is, an overwhelming majority of players can play three or fewer defensive field positions, and, to understand this, one only has to glance at the depth chart of any given team.  For example, based on the depth chart for the Chicago White Sox (CWS) as of August 31, 2012 (see Fig. 2), we can determine in how many ways it is possible to assign infielders and outfielders to their starting positions.  As it follows from the depth chart and from the tree diagrams shown in Figures 3 and 4, positions 2 – 9 and the DH can be chosen in the following number of ways: catchers, 2 ways, infielders, 9 ways, outfielders, 10 ways, and DH, 4 ways (in spite of the official depth chart, any one of 13 roster players who is not assigned to play a defensive position, can be appointed as the DH). As a result, the number of CWS manager's choices is limited by the following number: $N = 2 \cdot 9 \cdot 10 \cdot 4 = 720$.

| CATCHER | 1ST BASE | 2ND BASE | SHORTSTOP |
|---|---|---|---|
| 1. Pierzynski | 1. Konerko | 1. Beckham | 1. Ramirez |
| 2. Flowers | 2. Dunn | 2. Olmedo | 2. Olmedo |
| | 3. Youkilis | | |

| 3RD BASE | LEFT FIELD | CENTER FIELD | RIGHT FIELD |
|---|---|---|---|
| 1. Youkilis | 1. Viciedo | 1. Wise | 1. Rios |
| 2. Olmedo | 2. Wise | 2. Danks | 2. Wise |
| | 3. Danks | 3. Rios | 3. Danks |

**Figure 2**

Still, assigning the starting players is not an easy choice, and this is why a manager usually goes with the first player on the list for each position. These are the first choice players who are considered regular starters and play every game, unless the manager decides that they need rest. Most of the time, we perceive the choice as a good one. However, sometimes the fans wonder whether the predetermined order on the depth chart is always the best option and whether the manager makes a mistake by following this rule.

Let us go over some examples of Chicago White Sox managers' choices. In 2011, CWS manager Ozzie Gillen used Alex Rios as the center fielder (he was number one on the depth chart for that position). In 2012, the current manager Robin Ventura moved Rios to the right field, and the team's both defense and offense benefited from this player's success. Does this mean that 2011 depth chart was faulty, which led to a continual misuse of a valuable player? Is it entirely possible that among hundreds of possibilities there are several that can compete or be better than the general rule of following the depth chart? In order to find out the answers to these questions, it is necessary to formulate the problem of choosing the set of starters as a multiple criteria problem.

### 3.2. Problem formulation

When formulating a multiple criteria problem, we have to first determine the decision space and the criteria space. A 9-player starting assignment can be presented as a $9 \times p$ matrix where a row represents a selected starter player with his characteristics, and the length of a row, $p$, is the number of the characteristics of a single player that fully describe him as a team member. Such a row has fields that are labeled by the player's name, weight, height, age, salary, batting/throwing sides, the number of games started and innings played in each position, as well as his offensive and defensive numbers in the positions played. This approach is consistent with the formulation of portfolio selection problem with several criteria in Polyashuk (2005). Let us assume that the first row corresponds to the selected starting catcher (C), the second – to the first baseman (1B), and so on, according to the standard defensive position numbers chart, with the last row allotted to the designated hitter (DH). The set of all possible starting selections, which forms decision space, is the set of all possible $9 \times p$ matrices.

In order to form criteria space for the starting assignment problem, let us consider what the manager is trying to achieve when solving it. The ultimate goal, of course, is to win the game, and to reach this goal the manager has to put together a set of players that has the potential to score most runs while allowing least runs to the opposite team. At the same time, since a regular baseball season lasts 6 months and consists of 162 games, the manager has to use his team's resources wisely, which means maintaining the team's overall physical and psychological health.

While we are clearly coming up with three criteria: the offensive potential of the set of selected players, the defensive quality of the set, and the overall physical and psychological condition of the set, we will assume that the first two criteria, which are quantitative, are more important than the third, qualitative criterion. (We will view relative importance of criteria groups in the sense of Definition 1.2 in Polyashuk, 2005). There are other factors affecting manager's decisions. One of such factors is the players' handedness, which is an important consideration for so called "platoon advantage" (Baseball Info Solutions (BIS) and Bill James, 2011): more righties against left-handed pitchers and more lefties against right-handers.

The goal of this research effort, however, is to produce a set of optimal starting assignments with respect to the first two criteria only, while assuming that the team manager is capable of making the final selection based on the overall physical/psychological condition of the set, players' handedness, etc.

The next step is to use various characteristics of individual players in order to form two criteria, which would adequately represent offensive potential and defensive quality of a set of players. We will rely on a composite feature that has been extensively tested by baseball statisticians, which allows using offensive and defensive team numbers to approximate winning percentage of a team and, ultimately, predicting the outcome of the regular season very accurately. This characteristic is Winning Percentage and is calculated based on the so called Pythagorean Formula: $\frac{RS^2}{RS^2 + RA^2} = \frac{1}{1 + \left(RA/RS\right)^2}$.

More recently, the exponent used in the formula was replaced by 1.83, which gives a better fit with the empirical data: $\frac{1}{1 + \left(RA/RS\right)^{1.83}}$. As it is clear from the formula, the winning percentage depends on the ratio of *RA* (runs allowed by the team) and *RC* (runs created by the team). The formula above is a key to our method of forming offensive and defensive criteria for a set of starters in any given game, which is effectively based on composite characteristics of individual players reflecting their offensive potential and defensive value. More specifically, since winning is mostly about runs, let us use runs created by a player in order to quantify the player's overall offensive performance, and runs saved by a player in order to quantify the player's defensive performance in a given fielding position.

Let us assume that offensive potential of a single player in a lineup is defined by the number of runs created by a player per game, $RC/G = \dfrac{RC \cdot 27}{AB - H + SH + SF + CS + GDP}$.
Further, let us assume that the defensive quality of single player in each fielding position is characterized by defensive runs saved above average, *DRS*, as defined by Baseball Info Solutions (BIS). (We will not describe here the algorithm for computing the *DRS* but will simply use the data provided by BIS.)

**Definition 3.1** Let the set $T = \{p_1, p_2, \ldots, p_{25}\}$ be the team roster. Suppose a permutation of players $S = \left(p_{j_1}, p_{j_2}, \ldots p_{j_9}\right)$ is a starting assignment of players consisting of a catcher, a first baseman, a second baseman, a shortstop, a third baseman, a left fielder, a center fielder, a right fielder, and a designated hitter, in this order. Suppose the $i-\text{th}$ player in the lineup *S*, $p_{j_i}$, has offensive potential $r_i$ and defensive quality $d_i$, $i = 1, \ldots, 9$. Then the overall offensive quality of the 9-player starting assignment *S* is defined as $R_S = \frac{1}{9}\sum_{i=1}^{9} r_i$. The overall defensive quality of the starting assignment *S* is defined as $D_S = \sum_{i=1}^{9} d_i$.

Basically, the above definition provides criteria mapping, which assigns to every alternative in the decision space (in our case, the set of feasible starting assignments) the corresponding vector of criteria values in the criteria space. Further, we will consider a starting player assignment optimal if it provides maximum simultaneously for $R_S$ and $D_S$. The concept of simultaneous maximization of two criteria values will be interpreted in this paper in the context of the classical definition of Pareto binary relation, which defines a preference binary relation on the set of possible starting assignments.

**Definition 3.2** Starting assignment $S_A$ is preferred to starting assignment $S_B$, which is expressed as $S_A \prec S_B$, if and only if either $R_{S_A} \le R_{S_B}$ and $D_{S_A} < D_{S_B}$, or $R_{S_A} < R_{S_B}$ and $D_{S_A} \le D_{S_B}$.

**Definition 3.3** Let *A* be the set of all feasible 9-player starting assignments and $X = \{S_1, S_2, \ldots S_N\}$, $X \subseteq A$, is the set of the feasible starting assignments for a given game. We define the set of optimal assignments with respect to defensive potential and defensive quality criteria, $MAX(X)$ as the set of assignments that are non-dominated with respect to the binary relation $\prec$:

$$S^* \in MAX(X) \Leftrightarrow \neg\exists S \in X \text{ such that } S^* \prec S.$$

In other words, we are using a simple bi-criteria Pareto model (Sen, 1970) in order to choose the set of optimal starting player sets. Isolating the set of Pareto-optimal points should substantially reduce the total number of possible alternatives which should allow the manager make the final choice confidently. In any case, the greatest benefit from applying this model would be the elimination of dominated selections, which could become quite an eye opener. Let us now follow up with the example of starting 9-player selection for the 2012 Chicago White Sox.

### 3.3. Finding optimal starting player assignments: 2012 Chicago White Sox

Figures below show the decision charts for selecting the starters for infield positions (1B-2B-SS-3B) (Figure 3) and outfield positions (LF-CF-RF) (Figure 4). These charts confirm that CWS can the infield starters 9 different ways and outfield starters – 10 ways.

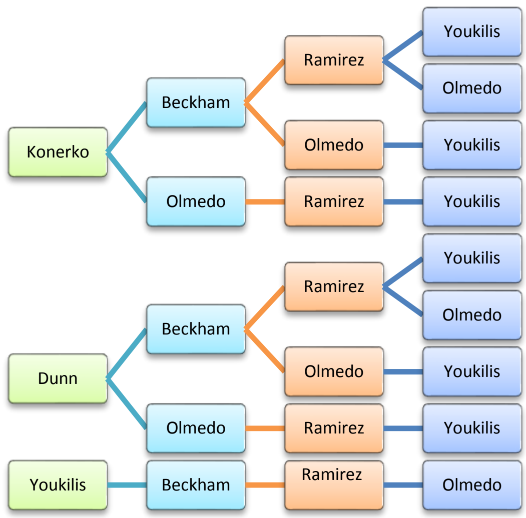



**Figure 3**

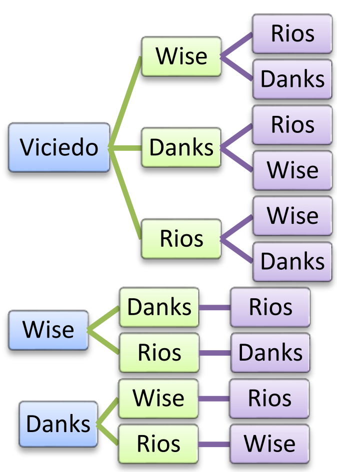



**Figure 4**

Next, we can evaluate both offensive potential and defensive quality for each position player, each catcher as well as for each infield and outfield combination of starting

players. The following tables show these numbers respectively. We are not providing a table for the catchers since their stats are covered by Table 1.
Notice that the tables include players' handedness, which is an important consideration for so called "platoon advantage" (Baseball Info Solutions (BIS) and Bill James, 2011): more righties against left-handed pitchers and more lefties against right-handers.

| # | Players | Position | Offense: *RC/G* | Bats | Defense: *DRS* |
|---|---|---|---|---|---|
| **1** | Beckham | 2B | 3.3 | Right | -4 |
| **2** | Danks | OF | 3.0 | Left | LF 1, CF 1, RF 0 |
| **3** | Dunn | 1B | 5.8 | Left | -4 |
| **4** | Flowers | C | 4.0 | Right | 2 |
| **5** | Konerko | 1B | 6.9 | Right | -6 |
| **6** | Olmedo | INF | 2.1 | Switch | 2B 1, SS 0, 3B 0 |
| **7** | Pierzynski | C | 6.3 | Left | -9 |
| **8** | Ramirez | SS | 3.5 | Right | 10 |
| **9** | Rios | OF | 5.6 | Right | RF 5, CF -9 |
| **10** | Viciedo | LF | 3.8 | Right | -1 |
| **11** | Wise | OF | 5.3 | Left | LF -1, CF -2 |
| **12** | Youkilis | INF | 6.0 | Right | 1B 0, 3B 2 |

**Table 1: Individual players' stats**

| # | Infield player assignment 1B, 2B, SS, 3B | Offense: average *RC/G* | Bats R/L/S | Defense: total *DRS* |
|---|---|---|---|---|
| **1** | Konerko, Beckham, Ramirez, Youkilis | 5.45 | 4/0/0 | 2 |
| **2** | Konerko, Beckham, Ramirez, Olmedo | 4.48 | 3/0/1 | 0 |
| **3** | Konerko, Beckham, Olmedo, Youkilis | 4.58 | 3/0/1 | -8 |
| **4** | Konerko, Olmedo, Ramirez, Youkilis | 5.15 | 3/0/1 | 7 |
| **5** | Dunn, Beckham, Ramirez, Youkilis | 5.18 | 3/1/0 | 4 |
| **6** | Dunn, Beckham, Ramirez, Olmedo | 4.20 | 2/1/1 | 2 |
| **7** | Dunn, Beckham, Olmedo, Youkilis | 4.43 | 2/1/1 | -6 |
| **8** | Dunn, Olmedo, Ramirez, Youkilis | 4.88 | 2/1/1 | 9 |
| **9** | Youkilis, Beckham, Ramirez, Olmedo | 4.25 | 3/0/1 | 6 |

**Table 2: Stats for all possible infield assignments**

| # | Outfield player assignment LF, CF, RF | Offense: average *RC/G* | Bats R/L/S | Defense: total *DRS* |
|---|---|---|---|---|
| **1** | Viciedo, Wise, Rios | 4.90 | 2/1/0 | 5 |
| **2** | Viciedo, Wise, Danks | 3.03 | 1/2/0 | 0 |
| **3** | Viciedo, Danks, Rios | 4.13 | 2/1/0 | 8 |
| **4** | Viciedo, Danks, Wise | 4.03 | 1/2/0 | 5 |
| **5** | Viciedo, Rios, Wise | 4.90 | 2/1/0 | -5 |
| **6** | Viciedo, Rios, Danks | 4.13 | 2/1/0 | -7 |
| **7** | Wise, Danks, Rios | 4.63 | 1/2/0 | 5 |
| **8** | Wise, Rios, Danks | 4.63 | 1/2/0 | -10 |
| **9** | Danks, Wise, Rios | 4.63 | 1/2/0 | 4 |
| **10** | Danks, Rios, Wise | 4.63 | 1/2/0 | -10 |

**Table 3: Stats for all possible outfield assignments**

After the statistical data have been organized, we can proceed with analyzing bi-criteria evaluations for all 720 possible starting player assignments. Figure 5 below demonstrates the scatter diagram of these evaluations in criteria space.

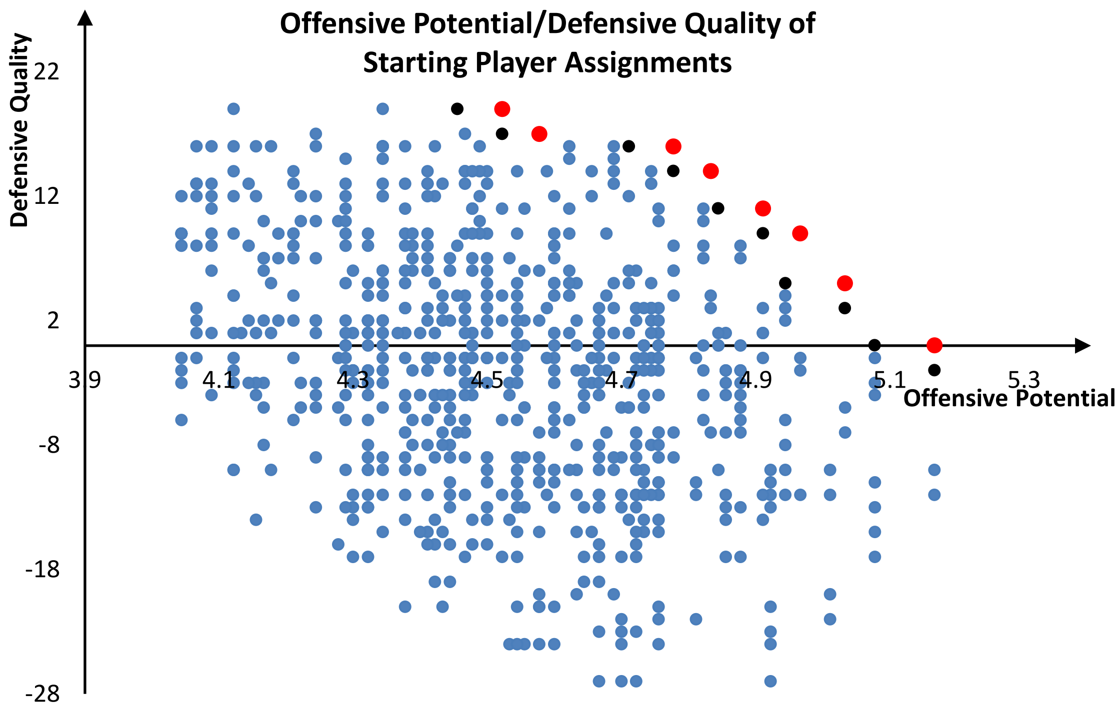



**Figure 5**

It is clear that, in our example, there are eight Pareto-optimal points in the bi-criteria space (shown in red color), and it is easy to identify the corresponding starting player assignments. Black dots correspond to the "next best" assignments: the points, which become optimal if the set of the original optimal points is deleted. For example, if the optimal solutions are unavailable, we can take a look at the next-best solutions. It appears that the optimal points and the corresponding assignments are (left to right) as follows: Optimal Solution 1 (OS1), $(4.52,\ 19)$: infield assignment #8, outfield assignment #3, catcher Flowers, and DH Konerko; OS 2, $(4.58,\ 17)$: infield assignment #4, outfield assignment #3, catcher Flowers, and DH Pierzynski; OS3, $(4.78,\ 16)$: infield assignment #8, outfield assignment #1, catcher Flowers, and DH Konerko; OS 4, $(4.83,\ 14)$: infield assignment #4, outfield assignment #1, catcher Flowers, and DH Pierzynski; OS 5, $(4.91,\ 11)$: infield assignment #5, outfield assignment #1, catcher Flowers, and DH Konerko; OS 6, $(4.97,\ 9)$: infield assignment #1, outfield assignment #1, catcher Flowers, and DH Pierzynski; OS 7, $(5.03,\ 5)$: infield assignment #8, outfield assignment #1, catcher Pierzynski, and DH Konerko; and OS8, $(5.17,\ 0)$: infield assignment #5, outfield assignment #1, catcher Pierzynski, and DH Konerko.

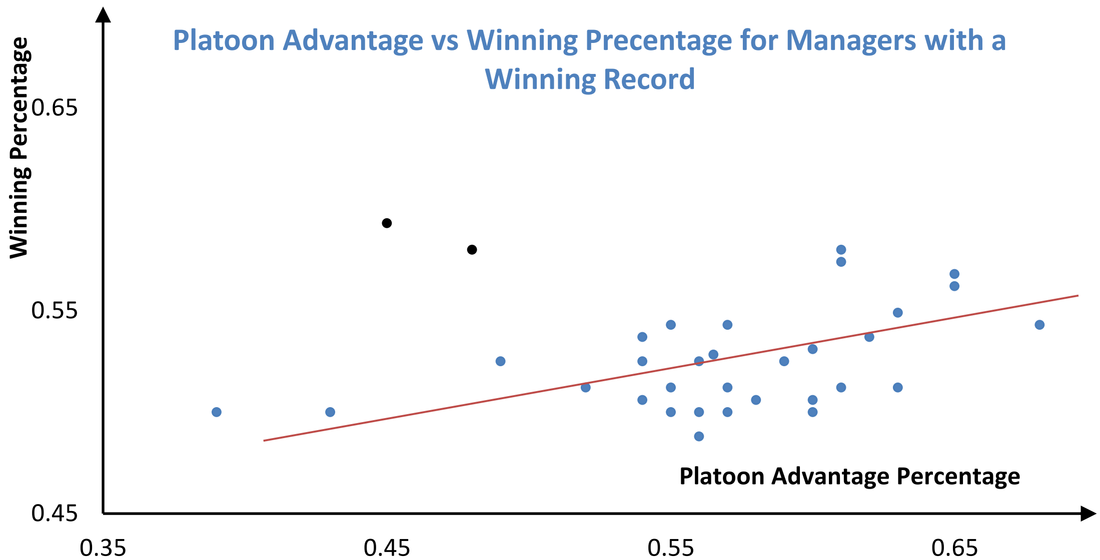
Let us consider how often these starting assignments were used by Robin Ventura, the CWS manager. According to the Basebal-Reference.com website, by the end of August 2012 (after 131 games of the season), the only optimal sets of starters that were played were Optimal Solution 5 (5 games) and Optimal Solution 8 (4 games). When calculating these numbers, we assumed that Dewayne Wise and Alejandro De Aza are interchangeable due to very similar offensive and defensive characteristics. (De Aza was not used in the starting assignments outlined in the tables above because he was on the disabled list and unavailable at the time of data gathering for this paper.)

Notably, there exists one starting assignment that was used by the CWS quite often (22 games), which displays the same offensive number as our OS8 but a lower defensive number: $(5.17,\ -2)$ versus $(5.17,\ 0)$. This assignment is identical to our OS8, except that two players have traded their positions: Adam Dunn plays DH while Paul Konerko plays first base, rather than the opposite assignment in our optimal solution. The corresponding point is one if the "next best" solutions and is among the points shown in black color on the scatter plot. Although this point is located pretty close to the Pareto set, it is clear that from the defense perspective Dunn playing first base should be preferable to Konerko. There are other notable differences between our findings and the CWS practice in 2012. For instance, five of eight of our optimal points correspond to starting assignments in which Ray Olmedo, rather than the regular starter Beckam, plays the second base; in reality, Olmedo started just one game in this capacity. Also, six out of eight optimal solutions have Tyler Flowers in the catcher's position rather than the regular starter A. J. Pierzynski. In spite of the differences, however, we can observe a lot of features of our optimal solutions that are consistent with the team practice in 2012. For example, all of our solutions as well as most frequent CWS manager's selections use Alexei Ramirez as SS, Kevin Youkilis as 3B, Dayan Viciedo as LF, and Rios as RF.

In support of the manager, we can note that the active roster keeps changing due to the players' injuries and the players' statistics change from week to week, which makes decisions more difficult. However, without making far-going conclusions about effectiveness of any manager, imagine how much more informed his decisions on the starting player assignments could be had he known what the optimal solutions are before any given game.

### 3.4. Applying players' handedness when choosing an optimal starting assignment

Platoon advantage percentage (Baseball Info Solutions (BIS) and Bill James, 2011) is the percentage of players in the lineup who can either bat right against a left-handed starting pitcher of the opposing team or bat left against a right-handed pitcher. Ideally, a manager should try to maximize platoon advantage percentage. Our analysis of a sample of thirty-three 2011 baseball managers with career winning records shows a small positive correlation between the platoon advantage percentage and the winning for these managers, with correlation coefficient of 0.163 (see the scatter plot below). Eliminating a couple of outliers (shown on the plot in black color), leaves the sample of 31 managers, which demonstrates a much stronger positive correlation between platoon advantage percentage and the winning percentage (correlation coefficient of 0.476.)




In order to apply the concept of maximizing platoon advantage to the CWS team, let us consider 28 starting player assignments maximizing the number of players that can bat right, which was eight for the 2012 CWS, and 130 starting assignments maximizing the number of players batting left (in the case of the 2012 CWS, five). Two figures below display scatter plots with points corresponding to such assignments (Figures 6 and 7). Applying Pareto preference relation allows identifying two optimal solutions in each case. (The corresponding points on the scatter plots are marked red.) Black dots correspond to the "next best" assignments.

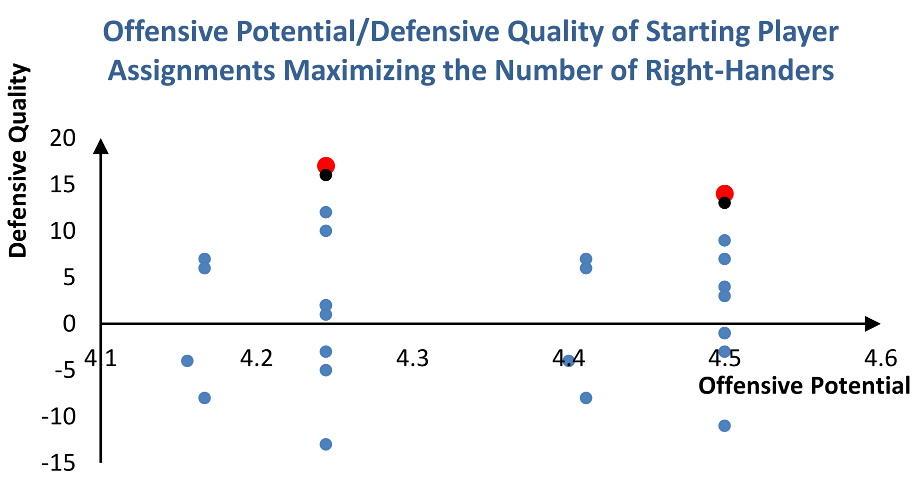



**Figure 6**

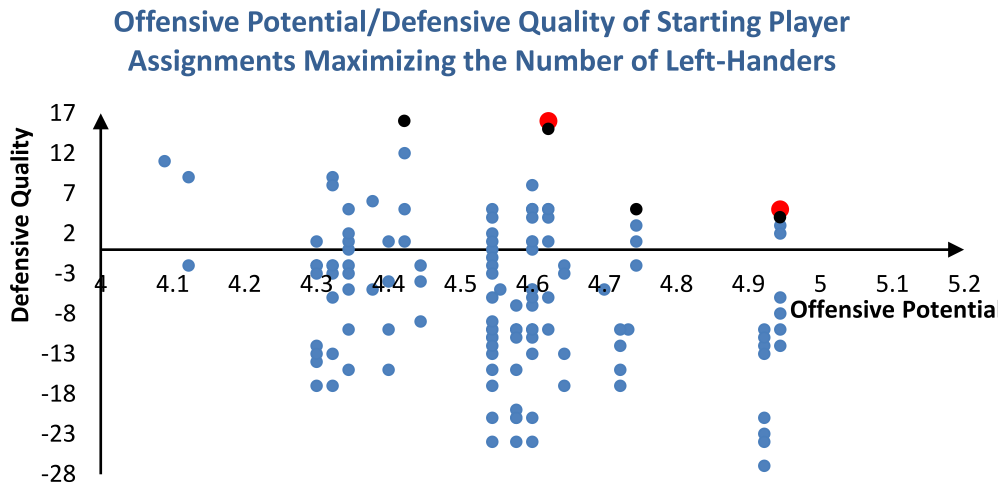



**Figure 7**

If we compare the optimal solutions with selections of the CWS, the results are mixed: in some respects, our model validates decisions made by the CWS manager but in other respects model identifies surprisingly different solutions from the actual practices. Let us go over several striking similarities first. According to our model, Alexei Ramirez should play shortstop in every one of the four (combined for both right- and left-handed starting selections) optimal solutions, and in reality he is a regular shortstop starter on the CWS team. Adam Dunn, according both to the model and to the real life, should have played in all games against right-handed pitchers, and catcher A. J. Pierzynski should have always played against right-handers; Dayan Viciedo should have started in the left field against left-handers, Alex Rios should have always started in the right field, all starting assignments should have used Kevin Youkilis, and Paul Konerko should play almost always (in three out of four optimal selections).

On the other hand, our model differs from typical practice of the CWS starting selections. For example, infielder Ray Olmedo appears to be a member of every one of the four optimal starting assignments as the second baseman. This makes sense because he is the only switch hitter on the team. However, in reality Olmedo served as a bench player and did not get a lot of starts for the team. Another observation: the team's right-handed catcher Tyler Flowers, according to our model's choices, should be used in every game against left-handed pitchers, and second baseman Gordon Beckam should not be played against right-handers. In reality, Flowers sat on the bench in many such games, and Beckam played second base in all but 9 out of first 131 games of the 2012 season.

Is it possible to conclude that Robin Ventura, the Chicago White Sox manager, has made some unjustified decisions in selecting starters in 2012? In the opinion of one who believes in strictly following the "law of large numbers" of statistical data and who is able to shake off prejudices of conventional wisdom, the answer is "yes".

## 5. Discussion and Conclusions

Let us summarize achievements of our approach to solving some of the baseball manager's problems, which have been presented in this paper, as well as some deficiencies and limitations of the current research.

We may list the following achievements and advances:

- The lineup selection problem is decomposed into two easier problems: selecting a starting position assignment and choosing a batting order;
- It becomes possible to compare starting assignments based on two major quantitative criteria: RC and DRS;
- A new formula for the expected number of runs scored by a single batting order rotation taking into account probabilities that players score a run if followed by other players has been developed;
- Finding an optimal batting order is a tractable integer programming problem that can be solved by enumeration.

In the list of limitations it is possible to outline the following:

- The third criterion, physical/psychological condition, was not used; constant changes in the roster must be considered;
- RC and DRS criteria are not perfect;
- Suggested formula for the expected value is an underestimate and can be improved.

Directions for further research include incorporating a third criterion in solving the problem of choosing an optimal starting assignment; formulating and solving various substitution problems, finding formulas providing a better estimate for the expected number of runs scored by a single batting order rotation.

## 6. References


Albert J. and Bennett J. 2003, *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*, revised edition, Copernicus Books: New York

Baseball Info Solutions and James B. 2011, *The Bill James Handbook*, ACTA Sports: Chicago

BaseballReference.com

Kaufman K. and Tan CM, ed. 2012, *Baseball Prospectus 2012.* John Wiley: Hoboken, New Jersey

Polyashuk M. 2005, A Formulation of Portfolio Selection Problem with Multiple Criteria, *Journal of Multi-Criteria Decision Analysis* **13**: 135-145

Sen AK. 1970, *Collective Choice and Social Welfare*, Holden-Day: San Francisco

Sugrue P.K. and Mehrota A. 2007, An Optimization Model to Determine Batting Order in Baseball, *International Journal of Operational Research* 2: 39-46