Methods for the Statistical Evaluation of Defensive Ability in Major League Baseball

A Survey

Ben Baumer\textsuperscript{1-2}  Andy Galdi\textsuperscript{1}  Rob Sebastian\textsuperscript{1}

\textsuperscript{1}New York Mets
Flushing, NY 11368

\textsuperscript{2}Department of Mathematics
The Graduate Center of the City University of New York
New York, NY 10016
bbaumer@gc.cuny.edu

Joint Statistical Meetings
Washington, D.C.
August 5th, 2009
Overview

Problem

Quantify the defensive ability of MLB players

1. History, Background, and Traditional Metrics
2. Discrete, Zone-Based Metrics
3. Continuous Spatial Model
4. Comparison, Summary, & the Future
Fundamental Problem

- Real and important *differences in fielding ability* among MLB players
- Not so obvious how to quantify those differences!
- Unlike the batter/pitcher confrontation, fielding is messy
  - Irregularly-shaped 2D playing surface
  - Movement of the ball
  - Responsibility for successes and failures
  - Disentangle the contributions of multiple players
Defensive Skills

- 4 largely independent defensive skills
  - **Sure-handedness**
    - Measured subjectively by Fielding Percentage
  - **Range**
    - Believed to have wide distribution among players
    - Believed to have a relatively large impact on team defensive performance
  - **Positioning**
    - Not currently measured. Stay tuned for HIT f/x....
  - **Throwing**
    - Recent paper from Carruth & Jensen ([1])
Errors and Fielding Percentage

- Chadwick (1860s): Invented the Error
- Errors: subjective determination by the official scorer that a play should have been made with "ordinary effort"
- Fielding Percentage:

\[ 1 - \frac{E}{TotalChances} = 1 - \Pr(E | Chance) \]

- Disadvantages:
  - Measures "sure-handedness" but not range
  - Largely a subjective measurement
  - Continues to dominate mainstream defensive assessment to this day!
Range Factor

- James (1976): Calculate the number of plays made per 9 innings

\[ RF = 9 \cdot \frac{\text{Assists} + \text{PutOuts}}{\text{InningsPlayed}} \]

- Idea: Making more plays is more important than making fewer errors

- Limitations:
  - Fixed Sum Problem: Always 27 outs in a game, but not always the same number of defensive chances
  - Highly correlated with the pitching staff’s BIP rate [e.g. - 0.62 for SS]
Defensive Efficiency

- James (1975): What percentage of balls put into play does a team convert into outs?

\[
DER = \frac{Outs}{BallsInPlay} \approx 1 - HPBP
\]

- Implicitly measures all defensive skills!
- Idea: Observed probability of recording an out on a ball in play
- Limitations:
  - Cannot be applied to specific fielders without play-by-play data
Overview

Problem

Quantify the defensive ability of MLB players

1 History, Background, and Traditional Metrics
2 Discrete, Zone-Based Metrics
3 Continuous Spatial Model
4 Comparison, Summary, & the Future
**High-Res Data Sets**

- Several high-resolution **play-by-play data** sets are now available.
- Mapping of location, trajectory, and fielder identity for every ball in play:
  - **Retrosheet**: Open, uses pre-defined 2D zones.
  - **STATS, Inc.**: Proprietary, uses polar and rectangular coordinates.
  - **BIS**: Proprietary, but less expensive. Uses rectangular coordinates.
- **Idea**: Discretize the playing surface and record defensive efficiency **for each fielder!**

(a) Retrosheet  
(b) STATS, Inc.  
(c) BIS
Zone Rating

- Dewan (1990): Percentage of plays made in a pre-defined zone of responsibility
- Extra credit given for plays made outside of one’s zone
- Advantages:
  - Objective measurement of range
- Limitations:
  - Zone definitions are somewhat arbitrary
  - Extra credit / lack of debit for plays not made outside zone can skew results
  - Ratings have no obvious benchmark and vary by position
  - Can over-value sure-handedness
A Probabilistic Model of Range

• Pinto (2003[5]): Actual number of plays made minus expected

• Methodology:
  1. Assign every ball in play to a unique bin
  2. For each bin, estimate the probability of an out Pr(Out \cap k) by each fielding position \( k \)
  3. For each ball in play, credit \( 1 - Pr(Out \cap k) \) to the fielder who made the play, and debit \( Pr(Out \cap k) \) to every other fielder
  4. Sum over all balls in play

• Result: For each player, an estimate of the number of plays made beyond what a league average fielder would have made.
PMR Bin Definition

- Bins defined by 6 parameters:
  1. Location: 1D Vectors (22 in STATS, Inc., 18 in BIS)
  2. Trajectory: Bunts, ground balls, line drives, flyballs, and pop-ups (BIS also includes “fliners”)
  3. Velocity: Soft, medium, or hard hit
  4. Ballpark
  5. Pitcher Handedness
  6. Batter Handedness

- Limitations:
  - Bins can be small without many years of data
    - For six years of data, over 38,000 bins with average of 20 balls in each
    - Only 23% of the balls in play land in bins containing at least 100 balls
  - 1D coordinates make sense for infielders, but for OF?
  - Ball-Hogging Dilemma
PMR Example

- Bin: Soft fly balls hit to left CF (by RHB off RHP at CitiField)
- Suppose
  - $\Pr(\text{Out} \cap \text{CF}) = 0.50$
  - $\Pr(\text{Out} \cap \text{LF}) = 0.49$
  - $\Pr(\neg \text{Out}) = 0.01$

- Each play is worth:

<table>
<thead>
<tr>
<th>Event</th>
<th>CF Catch</th>
<th>LF Catch</th>
<th>No Catch</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>$1 - 0.50 = 0.50$</td>
<td>$0 - 0.50 = -0.50$</td>
<td>$0 - 0.50 = -0.50$</td>
</tr>
<tr>
<td>LF</td>
<td>$0 - 0.49 = -0.49$</td>
<td>$1 - 0.49 = 0.51$</td>
<td>$0 - 0.49 = -0.49$</td>
</tr>
</tbody>
</table>

- Both credit and debit are position-dependent
- Each player’s rating is independent of the performance of his teammates!
Ball-Hogging Dilemma

- In the previous example, Pr(¬Out) = 0.01 is very low.
- Balls to this bin will almost always be caught, does it matter who catches them?
- Suppose on 100 balls to NYM, the CF always takes precedence in this discretionary bin.

<table>
<thead>
<tr>
<th></th>
<th>Pr(Out ∩ CF)</th>
<th>Pr(Out ∩ LF)</th>
<th>Pr(¬Out)</th>
</tr>
</thead>
<tbody>
<tr>
<td>League</td>
<td>0.50</td>
<td>0.49</td>
<td>0.01</td>
</tr>
<tr>
<td>NYM</td>
<td>0.70</td>
<td>0.29</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Result: NYM CF has a PMR of +20, while the LF has a PMR of −20.

Problem: Is this fair to the LF?
Ultimate Zone Rating

- Lichtman (2003[4]): Incorporate a "ball-hogging correction"

\[
UZR_k = \left( \text{PlaysMade} \right) \cdot \Pr(\neg\text{Out}) - \left( \text{PlaysNotMade} \right) \cdot \Pr(\text{Out} \cap k) = \left( \text{PMR} \right) + \text{(Ball-Hogging Correction)}
\]

- The Ball-Hogging Correction for player \( i \) at position \( k \) is

\[
\left( \text{PlaysMade} \right)_{i \text{ playing } k} \cdot \Pr(\text{Out}) \cdot \left[ s_k - \left( s_k \right)_{i \text{ playing } k} \right]
\]

where \( s_k = \text{percentage of outs made by position } k \)
**UZR Additional Improvements**

- Bins defined by 4 parameters:
  - Location (2D)
  - Trajectory
  - Velocity
  - Batter Handedness

- Adjustments for:
  - Ballpark
  - Base-Out configuration
  - Pitcher GB/FB ratio

- Larger bin sizes: 92% of balls in bins of size at least 100

- Multiply each play by $\hat{r}(z)$, the average run value for batted ball in bin $z$

- Final additions for OF Arms and Double Plays

- Result: The estimated number of runs saved
UZR Example

- Bin: Soft fly balls hit to left CF between 250 and 300 feet
- Suppose
  - $\Pr(Out \cap CF) = 0.50$
  - $\Pr(Out \cap LF) = 0.49$
  - $\Pr(\neg Out) = 0.01$
- Each play is worth:

<table>
<thead>
<tr>
<th>Event</th>
<th>CF Catch</th>
<th>LF Catch</th>
<th>No Catch</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>$0.01\hat{r}$</td>
<td>0</td>
<td>$-0.5\hat{r}$</td>
</tr>
<tr>
<td>LF</td>
<td>0</td>
<td>$0.01\hat{r}$</td>
<td>$-0.49\hat{r}$</td>
</tr>
</tbody>
</table>

- Credit is position-independent, debit is position-dependent
- No debit when a teammate makes a play, but performance of teammates can still affect rating!
Recall the Ball-Hogging Dilemma

- In the previous example, $\Pr(\neg Out) = 0.01$ is very low.
- Balls to this bin will almost always be caught, does it matter who catches them?
- Suppose on NYM, the CF always takes precedence this discretionary bin.

<table>
<thead>
<tr>
<th></th>
<th>$\Pr(Out \cap CF)$</th>
<th>$\Pr(Out \cap LF)$</th>
<th>$\Pr(\neg Out)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>League</td>
<td>0.50</td>
<td>0.49</td>
<td>0.01</td>
</tr>
<tr>
<td>NYM</td>
<td>0.70</td>
<td>0.29</td>
<td>0.01</td>
</tr>
<tr>
<td>PMR</td>
<td>+20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UZR</td>
<td>$+0.2\hat{r}(z)$</td>
<td>$-0.2\hat{r}(z)$</td>
<td></td>
</tr>
</tbody>
</table>

- Ball-hogging correction helps to mollify the problems with discretionary bins.
Consequences of the Ball-Hogging Correction

- Suppose out of 100 balls in a particular zone,

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Expected</th>
<th>PMR</th>
<th>UZR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mets CF</td>
<td>40</td>
<td>50</td>
<td>−10</td>
<td>−7.5̂</td>
</tr>
<tr>
<td>Mets LF</td>
<td>25</td>
<td>25</td>
<td>0</td>
<td>−2.5̂</td>
</tr>
<tr>
<td>Total</td>
<td>65</td>
<td>75</td>
<td>−10</td>
<td>−10̂</td>
</tr>
</tbody>
</table>

- Result: A player’s UZR can be affected by the performance of his teammates
- Is this fair? It could have been a dropped flyball by the CF that drags the LF’s UZR down!
- Underscores the difficulty in disentangling the contributions of individual fielders, due to their inherent interdependence as teammates
Plus/Minus System

- Dewan (2006[2]): 2\textsuperscript{nd} edition is very similar to UZR
- Can be shown that for a given bin \( z \):

\[
Plus/Minus = \left( \frac{1}{Pr(\neg Out) + Pr(Out \cap k)} \right) \cdot UZR
\]

- Differences from UZR (besides the above):
  - Uses (physically) smaller bins
  - Ballpark Factor: remove balls hit unreachably high off of wall
  - Controls for only some Base-Out states
  - No adjustments for Batter or Pitcher Handedness, Pitcher GB/FB rate
  - Includes ratings for Pitchers and Catchers
Overview

Problem

Quantify the defensive ability of MLB players

1. History, Background, and Traditional Metrics
2. Discrete, Zone-Based Metrics
3. Continuous Spatial Model
4. Comparison, Summary, & the Future
SAFE Motivation

- Jensen, et al. (2009[3]): Model the playing field as a continuous surface
- Estimate each player’s fielding ability as a smooth probability surface
  1. Use probit regression to estimate a shared normal prior
  2. Use hierarchical Bayesian techniques and Gibbs sampling to estimate each player’s surface
  3. Multiply by BIP frequency, estimated run value, and shared responsibility value
  4. Integrate using numerical techniques
- Result: Estimate of expected runs saved
- Big Idea: SAFE estimates true fielding ability, while discrete models estimate observed fielding performance
Step 1: Probit Regression Model

- Separate models based on trajectory:
  - Flyballs & Line Drives: 2D-Distance (D), Velocity (V), Direction (F)
  - Groundballs: 1D-Angle ($\theta$), Velocity (V), Direction (L)

- Probit regression model with interaction
Step 2: Player-Specific Probability Surfaces

- Repeat Step 1 for each player
- Shrink using hierarchical Bayesian techniques and Gibbs sampling

Result: Continuous surface $\hat{p}_i$ giving the probability of a successful fielding play
Step 3: Converting to Runs

- Estimate continuous **kernel density functions** with respect to location and trajectory:
  - $\hat{f}(x, y, v)$: the frequency of balls in play (or $\hat{f}(\theta, v)$)
  - $\hat{r}(x, y, v)$: the run consequence
  - $\hat{s}_k(x, y, v)$: the shared responsibility value

**Overall Density: Flyballs, Velocity=2**

**Run Consequence: Liner, Velocity=2**
Step 4: Putting it all Together

- Integrate the product of the preceding quantities over the playing surface and velocities

\[
SAFE^\text{fly}(i) = \int \hat{f}(x, y, v) \cdot \hat{r}(x, y, v) \cdot \hat{s}_k(x, y, v) \cdot [\hat{p}_i(x, y, v) - \hat{p}_+(x, y, v)] \, dx \, dy \, dv
\]

\[
SAFE^\text{grd}(i) = \int \hat{f}(\theta, v) \cdot \hat{r}(\theta, v) \cdot \hat{s}_k(\theta, v) \cdot [\hat{p}_i(\theta, v) - \hat{p}_+(\theta, v)] \, d\theta \, dv
\]

- Evaluate using numerical integration techniques
- Sum over trajectories

**Outfielders**: \( SAFE(i) = SAFE^\text{fly}(i) + SAFE^\text{liner}(i) \)

**Infielders**: \( SAFE(i) = SAFE^\text{fly}(i) + SAFE^\text{liner}(i) + SAFE^\text{grd}(i) \)

- Result: Expected runs saved relative to the average fielder at that position
SAFE Limitations

- Starting position of fielders is computed (centroid of successful plays)
  - Ignores situational position (e.g. - infielders holding runners, shifts, hit & run plays, etc.)
- Does not account for ballpark factors
- Includes infield pop-ups, which could be discretionary
- Does not account for other defensive skills:
  - Double plays
  - Outfield throwing ability
  - 1B ”scooping” ability
Overview

Problem

Quantify the defensive ability of MLB players

1. History, Background, and Traditional Metrics
2. Discrete, Zone-Based Metrics
3. Continuous Spatial Model
4. Comparison, Summary, & the Future
Ballpark Factors: 3 Approaches

- **PMR**: Use ballpark as a parameter in bin definition
  - Ideal, but requires large amounts of data
- **UZR**: Compute separate **ballpark factors** for LF, CF, RF, and IF
  - Complicated
- **Plus/Minus**: Throw out balls that hit **outfield walls** above 10 feet
  - Difficult to determine from the data

![Flyball Distribution: All Parks](image1)
![Freq](image2)
![Flyball Distribution: Fenway Park](image3)
![Freq](image4)
Ballpark Factors: 3 Approaches

- **PMR**: Use ballpark as a parameter in bin definition
  - Ideal, but requires large amounts of data
- **UZR**: Compute separate *ballpark factors* for LF, CF, RF, and IF
  - Complicated
- **Plus/Minus**: Throw out balls that hit *outfield walls* above 10 feet
  - Difficult to determine from the data
## Comparison

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>PMR</th>
<th>UZR</th>
<th>+/-</th>
<th>SAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
<td>Either</td>
<td>Either</td>
<td>BIS</td>
<td>BIS</td>
</tr>
<tr>
<td>Model</td>
<td>Discrete</td>
<td>Discrete</td>
<td>Discrete</td>
<td>Continuous</td>
</tr>
<tr>
<td>Infield LD &amp; Pop-Ups</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Location</td>
<td>1D</td>
<td>2D</td>
<td>2D</td>
<td>2D</td>
</tr>
<tr>
<td>Trajectory</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Velocity</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Base-Out</td>
<td>No</td>
<td>Yes</td>
<td>Some</td>
<td>No</td>
</tr>
<tr>
<td>Batter Handedness</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Pitcher Handedness</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ballpark</td>
<td>Bins</td>
<td>Regions</td>
<td>Walls</td>
<td>No</td>
</tr>
<tr>
<td>Result</td>
<td>Plays Made</td>
<td>Runs Saved</td>
<td>Runs Saved</td>
<td>Exp. Runs Saved</td>
</tr>
</tbody>
</table>

- SAFE features a more advanced statistical model
- UZR makes more comprehensive corrections
PMR, UZR Spread & Correlation

<table>
<thead>
<tr>
<th>Pos</th>
<th>N</th>
<th>$\sigma_{PMR}$</th>
<th>$\sigma_{UZR}$</th>
<th>$r_{PMR,UZR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B</td>
<td>100</td>
<td>28.5</td>
<td>14.5</td>
<td>0.635</td>
</tr>
<tr>
<td>2B</td>
<td>101</td>
<td>38.3</td>
<td>24.8</td>
<td>0.709</td>
</tr>
<tr>
<td>3B</td>
<td>98</td>
<td>36.4</td>
<td>24.7</td>
<td>0.706</td>
</tr>
<tr>
<td>SS</td>
<td>108</td>
<td>34.9</td>
<td>22.4</td>
<td>0.637</td>
</tr>
<tr>
<td>LF</td>
<td>166</td>
<td>15.0</td>
<td>13.1</td>
<td>0.668</td>
</tr>
<tr>
<td>CF</td>
<td>166</td>
<td>17.0</td>
<td>15.1</td>
<td>0.614</td>
</tr>
<tr>
<td>RF</td>
<td>161</td>
<td>15.7</td>
<td>15.4</td>
<td>0.569</td>
</tr>
<tr>
<td>Total</td>
<td>900</td>
<td>26.2</td>
<td>18.3</td>
<td>0.650</td>
</tr>
</tbody>
</table>

Table: PMR and UZR per 4000 BIP, 2004-2007 (min. 1000 BIP)

- UZR has a narrower spread, but they are closely correlated
Reliability

- Measured by *year-to-year correlation*
- PMR and UZR are per 4000 BIP, SAFE is raw

<table>
<thead>
<tr>
<th>Pos</th>
<th>$N$</th>
<th>PMR</th>
<th>UZR</th>
<th>SAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B</td>
<td>48</td>
<td>0.429</td>
<td>0.222</td>
<td>0.287</td>
</tr>
<tr>
<td>2B</td>
<td>53</td>
<td>0.431</td>
<td>0.546</td>
<td>0.051</td>
</tr>
<tr>
<td>3B</td>
<td>50</td>
<td>0.327</td>
<td>0.410</td>
<td>0.503</td>
</tr>
<tr>
<td>SS</td>
<td>59</td>
<td>0.453</td>
<td>0.438</td>
<td>-0.030</td>
</tr>
<tr>
<td>LF</td>
<td>66</td>
<td>0.343</td>
<td>0.568</td>
<td>0.594</td>
</tr>
<tr>
<td>CF</td>
<td>76</td>
<td>0.359</td>
<td>0.595</td>
<td>0.525</td>
</tr>
<tr>
<td>RF</td>
<td>68</td>
<td>0.113</td>
<td>0.326</td>
<td>0.444</td>
</tr>
<tr>
<td>Total</td>
<td>420</td>
<td>0.351</td>
<td>0.444</td>
<td>0.372</td>
</tr>
</tbody>
</table>

*Table: Year-to-Year Correlation, 2004-2007 (SAFE 2002-2005)*

- PMR and UZR reliability in the infield is similar
- UZR seems more reliable, but SAFE is *very competitive* in the OF
The Future: HIT f/x

- **HIT f/x**: Actual 2D locations and movement for the ball and fielders
- Continuous model is particularly adaptable to higher-resolution data
- **Fielder starting positions** can be exact for each ball in play, rather than a fixed centroid
  - Probability surfaces can be centered around actual starting position
  - Distance computations are more accurate
  - Ballpark & situational adjustments no longer necessary
- **Hang time**:
  - Replaces subjective, ordinal variable Velocity
  - Trajectory no longer necessary
Summary

- Traditional fielding metrics: severely limited by lack of high-resolution data & subjectivity
- Currently prominent metrics are based on very similar discrete models that estimate observed fielding performance
- SAFE: complex, continuous fielding metric that estimates true fielding ability, but lacks refinement
- Availability of ”true” continuous spatial data is forthcoming
  - Should fold seamlessly into SAFE, and dramatically improve accuracy
  - Not clear how much it would improve the discrete models
References

M. Carruth and S. Jensen.
Evaluating Throwing Ability in Baseball.

J. Dewan.
*The Fielding Bible.*

Bayesball: A Bayesian Hierarchical Model for Evaluating Fielding in Major League Baseball.

MGL.
Ultimate Zone Rating (UZR), Part 1.

D. Pinto.
Acknowledgements

- The New York Mets
- Mitchel Lichtman (MGL)
- David Pinto
- John Dewan and Ben Jedlovec (BIS)
- Shane Jensen
- Dave Allen (Fangraphs)
- Fence distance and height data provided by Greg Rybarczyk at Hit Tracker: http://www.hittrackeronline.com

THANK YOU!!!!!