Statistical Models for the Evaluation of Fielding in Baseball

Shane T. Jensen Kenny Shirley Abraham Wyner

Department of Statistics, The Wharton School, University of Pennsylvania

2007 New England Symposium on Statistics in Sports

Quantifying Fielding Performance

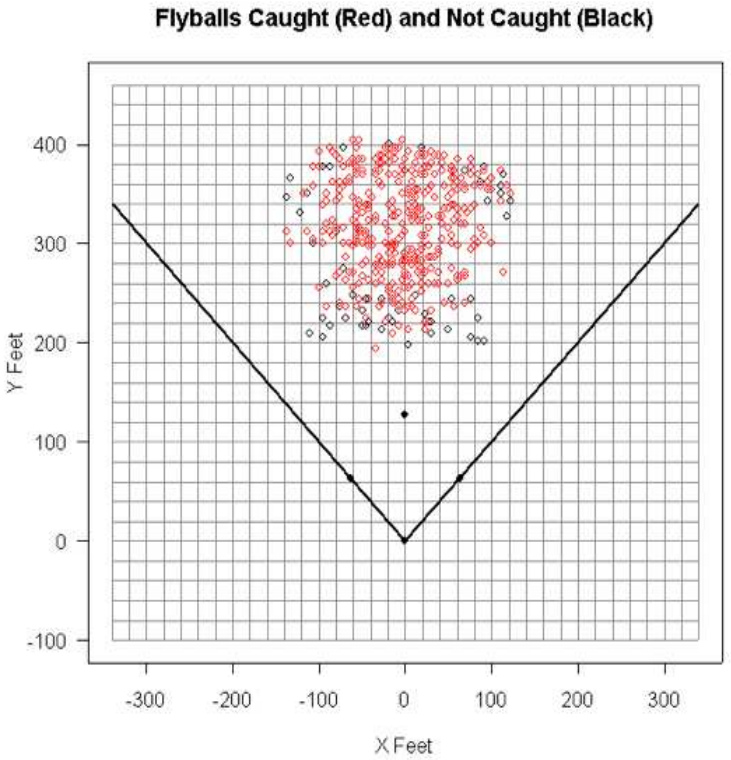
- **Overall goal**: accurate evaluation of the fielding performance of each major league baseball player
- Many aspects of game (eg. hitting, pitching) are easy to quantify and tabulate
  - finite number of outcomes, baserunner configurations
- Fielding is a more **continuous** aspect of the game
  - presents a greater data and modeling challenge

Popular Fielding Evaluation Methods

- **Ultimate Zone Rating** (Mitchel Lichtman):
  - divides field into large zones and tabulates of successful vs. unsuccessful plays for each fielder within zones
- **Probabilistic Model of Range** (David Pinto):
  - Field is cut into 18 pie slices (every 5 degrees) on either side of second base
  - replacement for UZR (which now has limited availability)
- Both methods have similar weakness: separate zones used when field is actually a **single continuous surface**
- Each zone/slice is large which limits ability to detect small differences between fielders
- Need **higher-resolution data** for continuous models

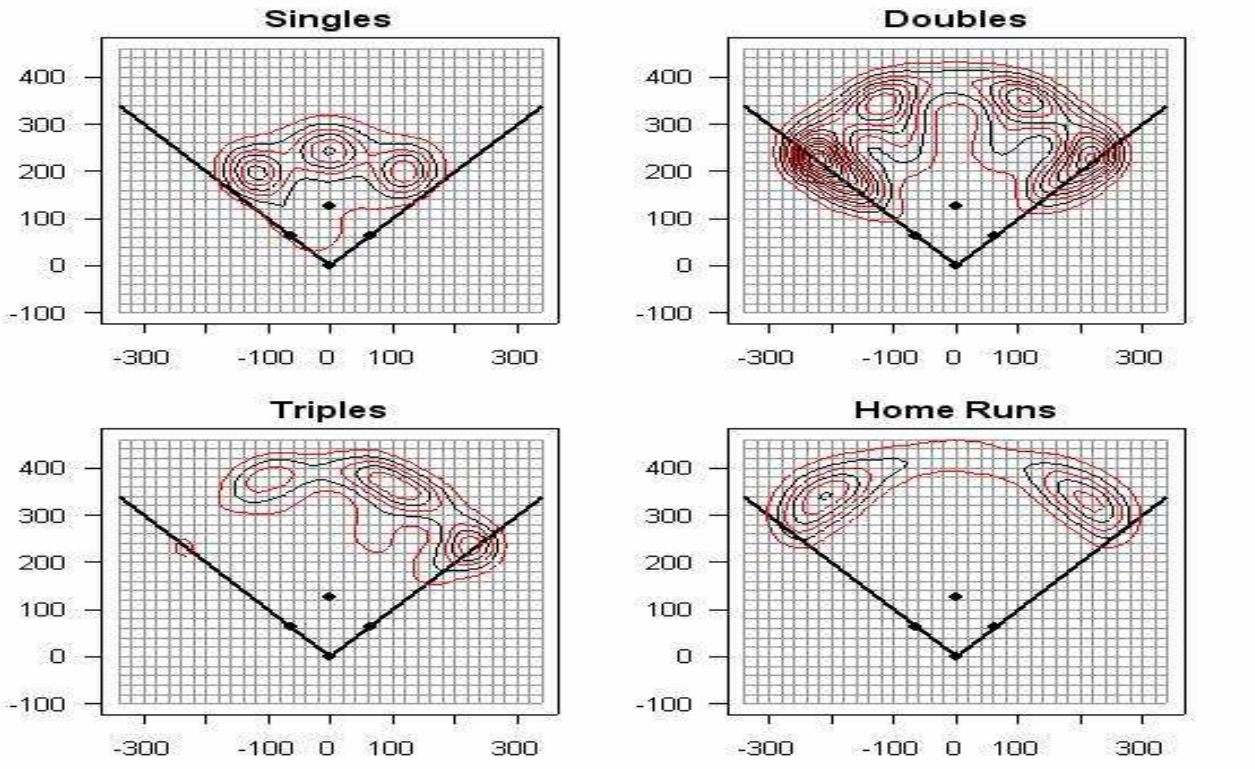
Baseball Info Solutions (BIS) Data

- **High-resolution BIS data** available via ESPN grant
- 4 years (02-06) with 120000 balls-in-play (BIP) per year
  - 42% grounders
  - 33% flys
  - 25% liners
- Each BIP is mapped to a **much smaller area** (4×4 feet) than the UZR zones
- **Velocity** information also but only as category



Smooth Fielding Curves

- High-resolution data allows us to fit **smooth curves** to the **continuous playing field**



- **Plus-Minus System** (John Dewan) also based on BIS data, but does not use smooth curves

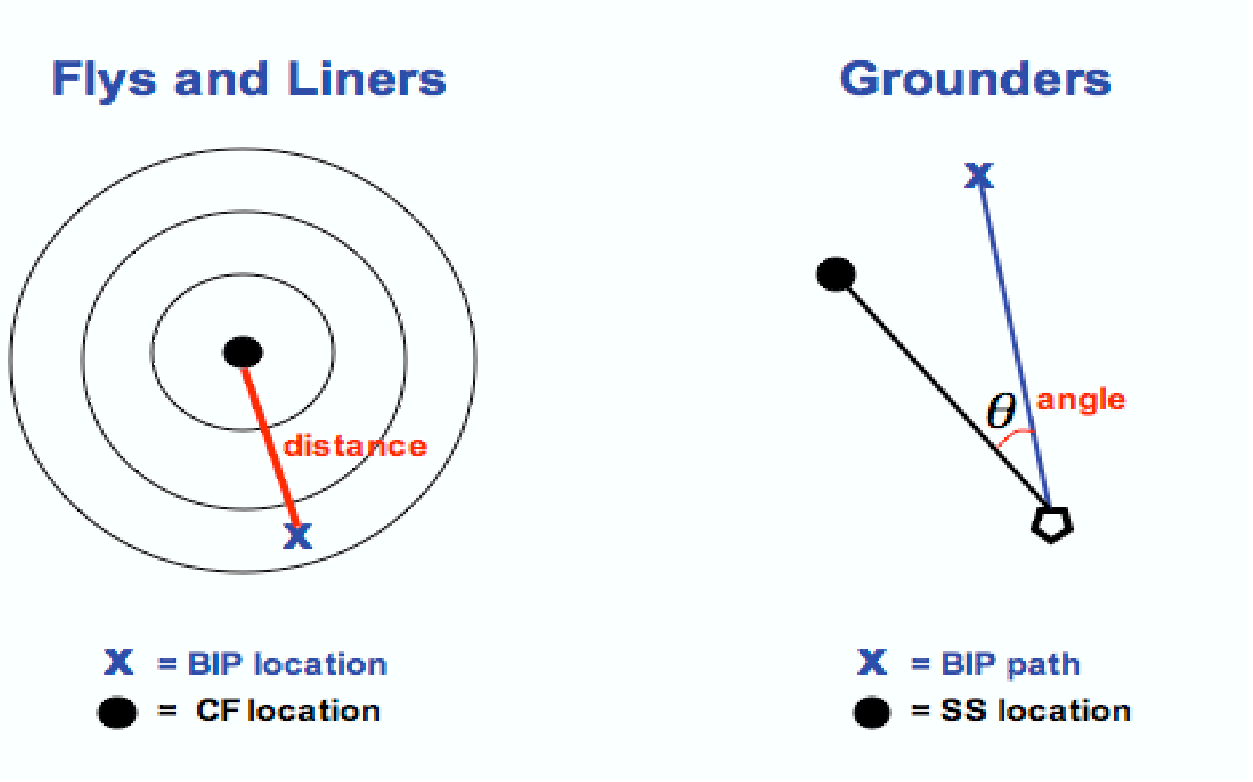
SAFE: Spatial Aggregate Fielding Evaluation

1. Fit smooth curve for **average fielder** in each position:
   - Using all players, estimate probability of success on a BIP as function of distance, direction and velocity
2. Fit separate smooth curve for each **individual fielder**
3. Calculate **difference** at each point between average curve and each individual curve
4. Weight difference at each point by **BIP frequency**
5. Weight difference at each point by **run consequence**
6. **Aggregate** runs saved/cost over all points for each fielder
   - Numerical integration over a fine grid used for aggregation

SAFE = (Individual - Average) $\times$ BIP Freq. $\times$ Run conseq.

Different Ball-In-Play Types

- **Two-dimensional curves** needed for **fly-balls/liners**: success depends on distance and direction to BIP
- **One-dimensional curves** needed for **grounders**: success depends on direction and angle between fielder and BIP



Logistic function for each smooth curve

- Logistic functions used to model curves for **probability** P of a successful fielding play
- **Logistic function for grounders:**

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \cdot \text{Angle} + \beta_2 \cdot \text{Velocity}$$

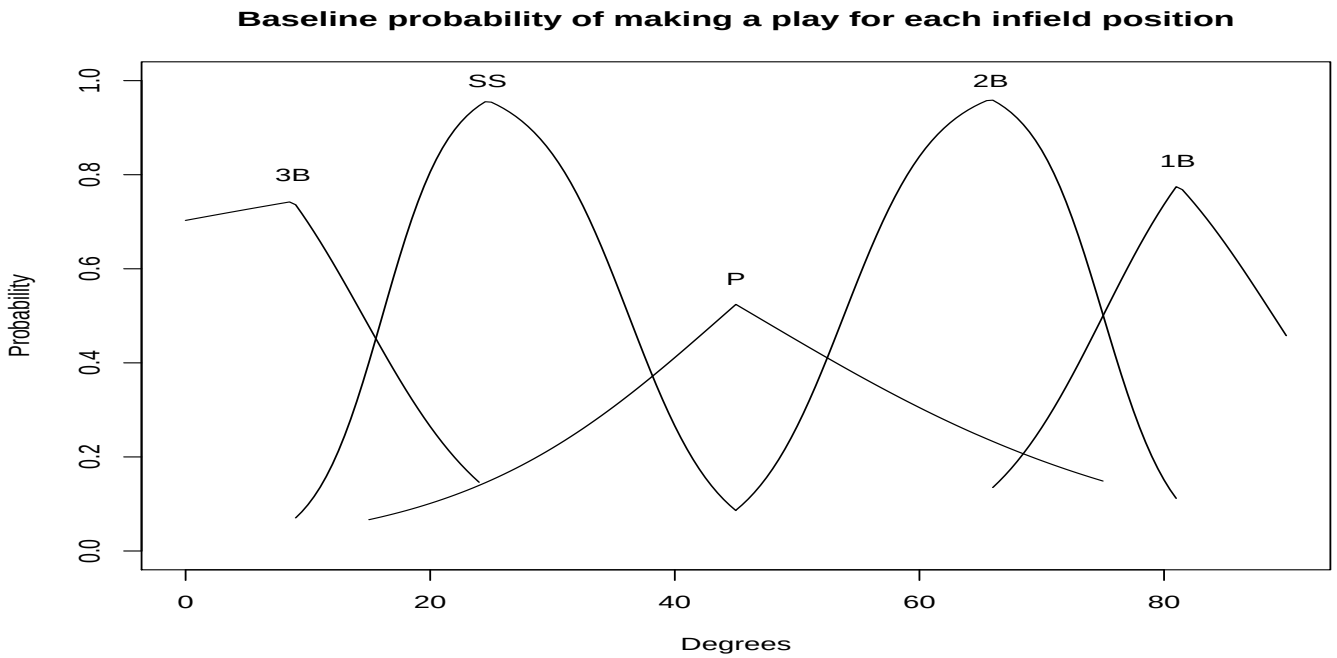
  - Different β_1 used for moving **left vs. right**
- **Logistic function for fly-balls/liners:**

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \cdot \text{Distance} + \beta_2 \cdot \text{Velocity}$$

  - Different β_1 used for moving **forward vs. back**

Average model for each position

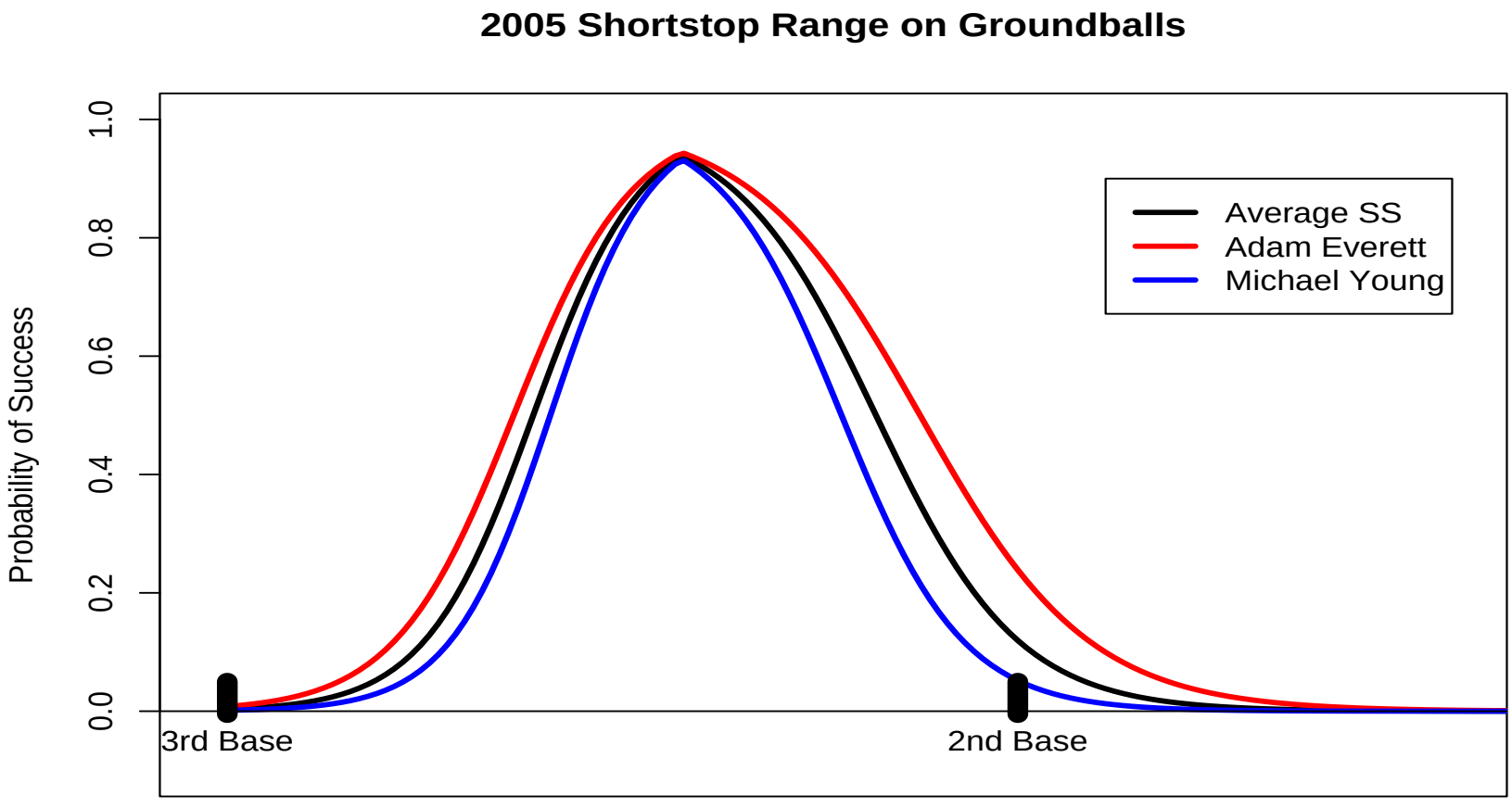
- Average model estimated by using **all players at position**.



- Curves centered at point with **highest success prob**.
  - Each distance is an estimate since we don't know exactly where fielder was standing at start of each play
- Note the different curves for moving to the left vs. right

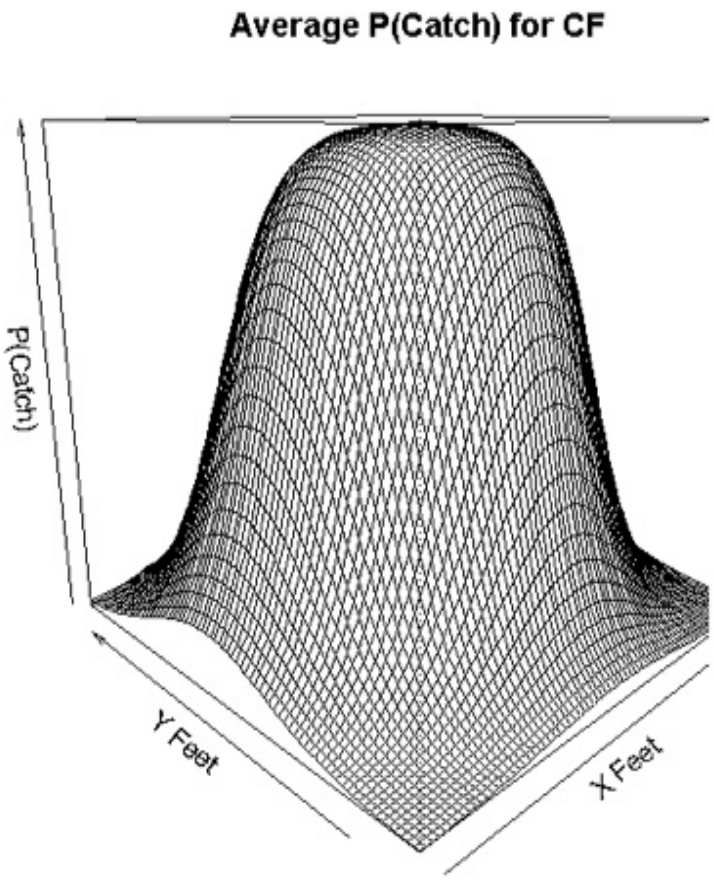
Individual models for Grounders

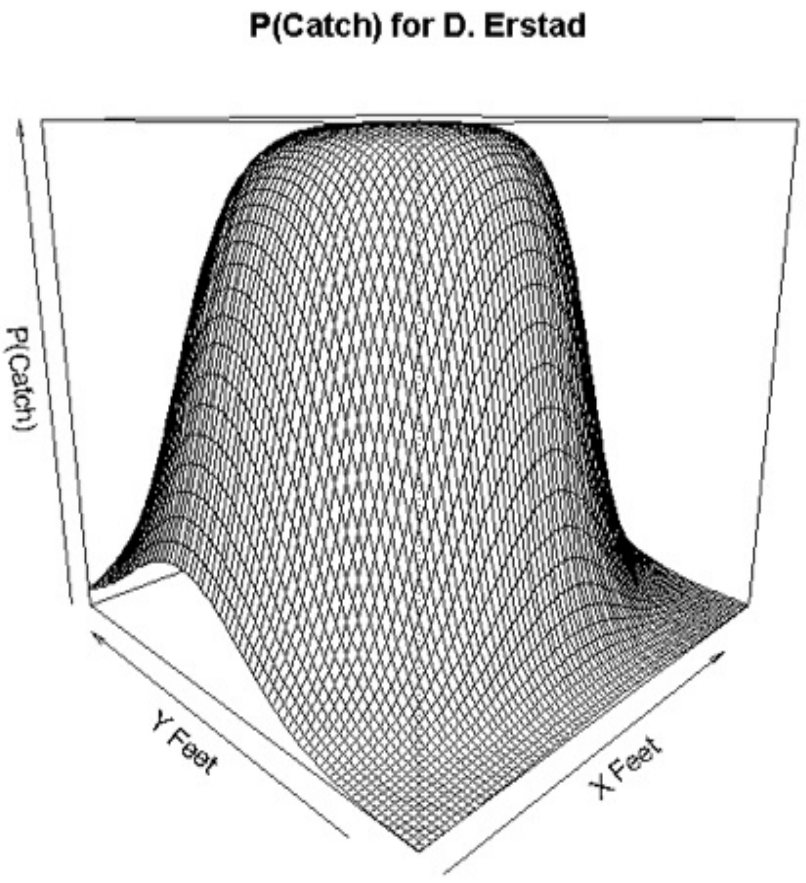
- Fit different 1-D logistic curves for each **individual fielder**.



Individual models for Fly Balls

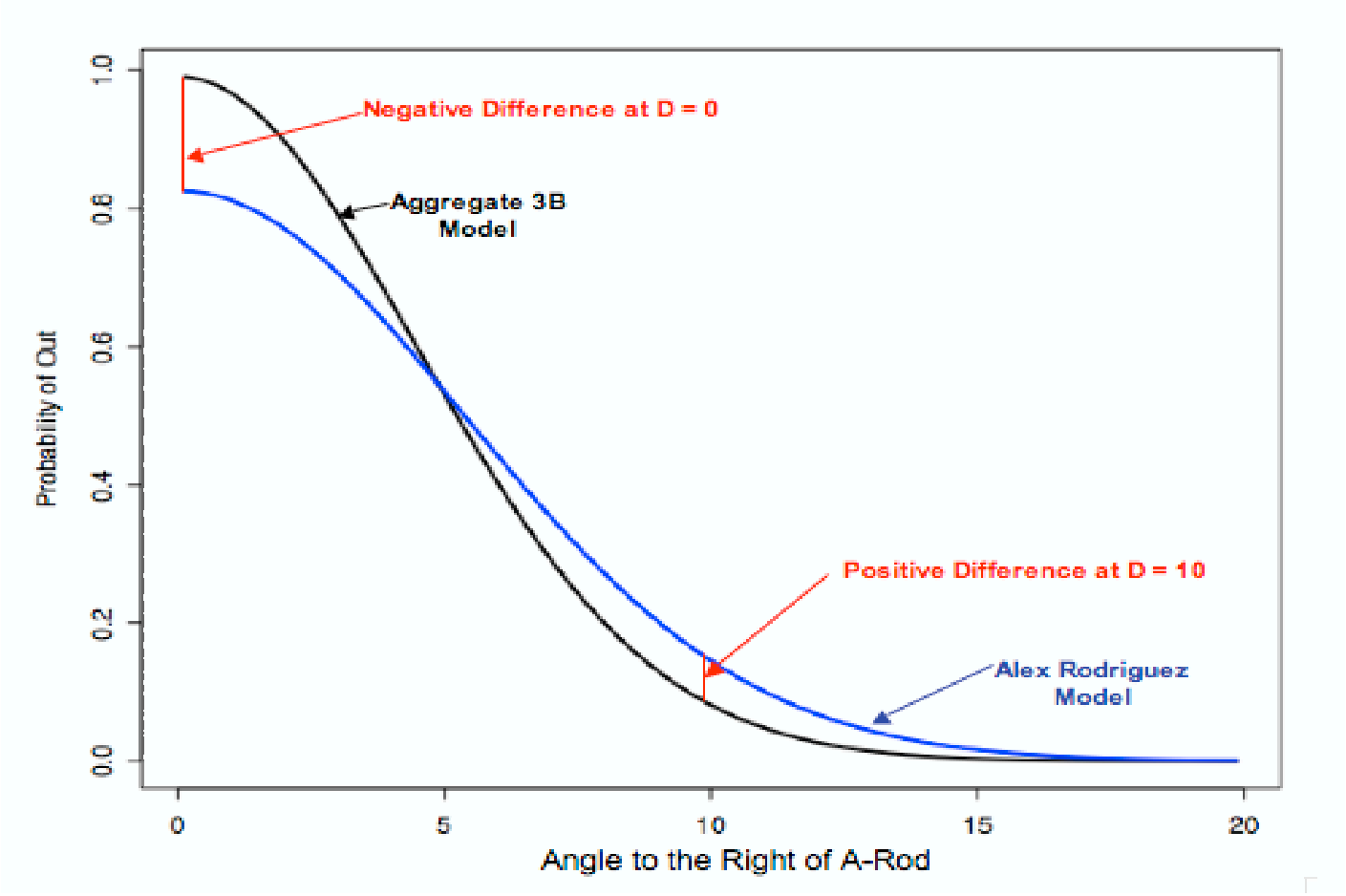
- Fit different 2-D logistic curves for each **individual fielder**.





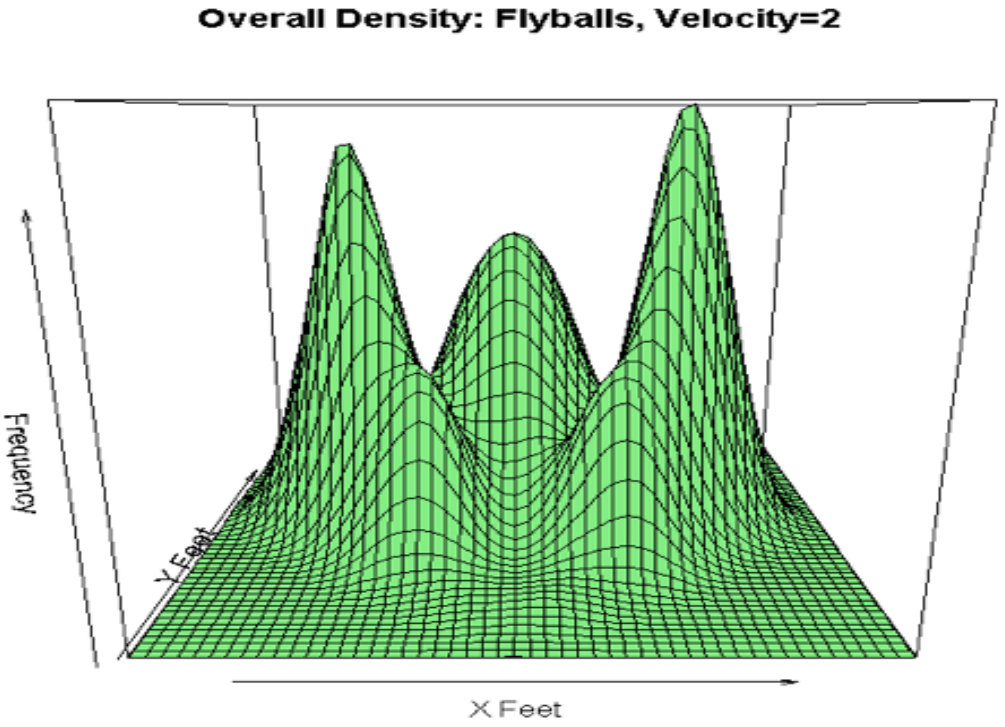
Curve Differences

- Calculate **point-by-point differences** between individual fielder curves and average curves at the position



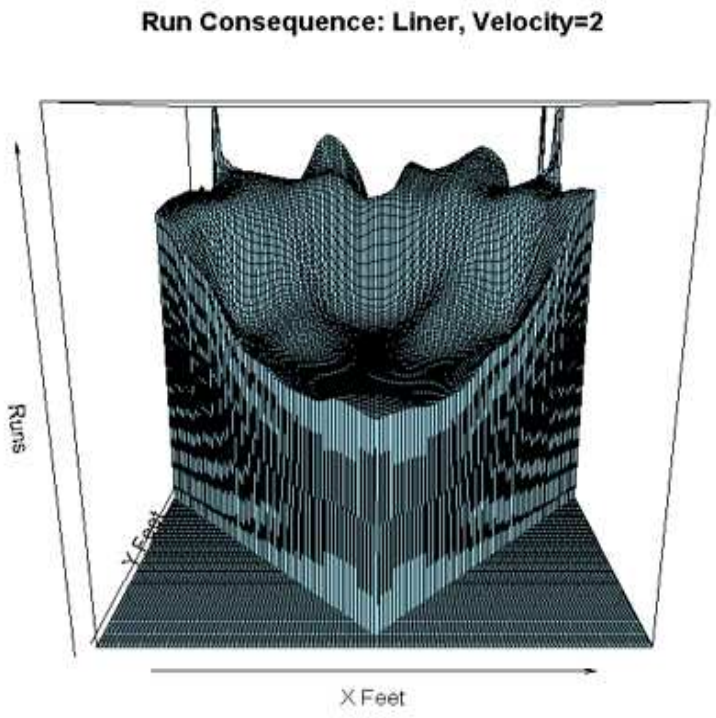
Weighting by BIP Frequency

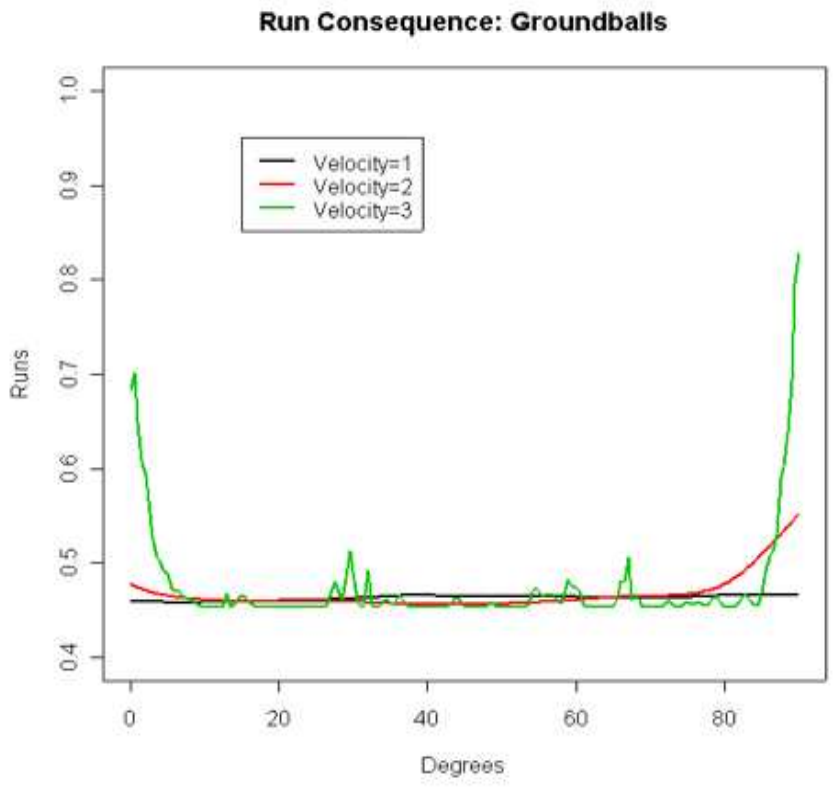
- Could add up curve differences (individual - aggregate) over all points, but not all points have **same frequency**
- Need to **weight this tabulation** so that more frequent distances or angles are more important



Weighting by Run Consequence

- Also calculate the **run consequence** of a unsuccessful play using frequencies of each hit type at the point
- Weight each point by run consequence to put differences in terms of **runs saved/cost**





Putting it all together with an example

- Carl Crawford has a 0.95 **probability of making a catch** on BIPs to a particular point in CF
  - The average CF has a 0.85 probability, giving Carl a **positive difference** of 0.10
- **BIP frequency** for this point is 15 balls per season, so Carl catches an extra $15 \times 0.1 = 1.5$ BIP to that point
- How many **runs** are these extra 1.5 catches worth?
  - Frequency of singles, doubles and triples to this point used to calculate average **run consequence** of missed catch which is 0.65 runs per BIP for this point
- So Carl has saved $1.5 \times 0.65 = 0.975$ runs at that point
- Aggregating Carl's run values across all points in CF gives the **total runs saved/cost** for Carl Crawford

Results for Infielders: Top 10 (average run value across 02-05)

	First Baseman		Second Baseman	
Best	1B Doug Mientkiewicz	7.30	2B Orlando Hudson	9.69
	1B Mark Teixeira	5.96	2B Nick Punto	7.71
	1B Chad Tracy	5.42	2B Mark Ellis	6.95
	1B Albert Pujols	4.93	2B Craig Counsell	6.31
	1B Ryan Howard	4.56	2B Chase Utley	5.81
	1B Darin Erstad	3.95	2B Junior Spivey	4.50
	1B Lance Niekro	3.81	2B Brian Roberts	3.94
	1B Kevin Millar	2.79	2B Adam Kennedy	3.80
	1B Tony Clark	0.91	2B Marcus Giles	2.81
Worst	1B Derrek Lee	-2.64	2B Ray Durham	-4.62
	1B Christopher Shelton	-3.01	2B Rich Aurilia	-4.72
	1B Richie Sexson	-3.35	2B Ruben Gotay	-5.13
	1B Shea Hillenbrand	-3.38	2B Todd Walker	-6.02
	1B Matt Stairs	-4.53	2B Rickie Weeks	-6.88
	1B Lance Berkman	-4.72	2B Miguel Cairo	-7.36
	1B Carlos Delgado	-4.77	2B Jose Vidro	-9.18
	1B Rafael Palmeiro	-5.73	2B Robinson Cano	-9.65
	1B Adam LaRoche	-6.38	2B Bret Boone	-9.67
	1B Jason Giambi	-7.28		

Results for Infielders: Top 10 (average run value across 02-05)

	Third Baseman		Shortstop	
Best	3B Scott Rolen	9.93	SS Adam Everett	12.32
	3B Adrian Beltre	8.65	SS Clint Barmes	8.96
	3B Sean Burroughs	6.00	SS Jack Wilson	6.79
	3B Corey Koskie	5.32	SS Cesar Izturis	5.86
	3B David Bell	5.09	SS Jason Bartlett	5.39
	3B Pedro Feliz	5.01	SS Neifi Perez	3.94
	3B Joe Crede	2.64	SS Juan Castro	3.68
	3B Bill Mueller	2.60	SS Omar Vizquel	3.46
	3B Morgan Ensberg	2.11	SS Julio Lugo	3.33
	3B Eric Chavez	2.10	SS Carlos Guillen	2.39
Worst	3B Joe Randa	-1.31	SS Miguel Tejada	-1.88
	3B Melvin Mora	-1.78	SS Marcos Scutaro	-2.06
	3B Brandon Inge	-2.58	SS Khalil Greene	-2.23
	3B Aramis Ramirez	-2.67	SS Cristian Guzman	-2.50
	3B Michael Cuddyer	-2.86	SS Jhonny Peralta	-2.71
	3B Alex Gonzalez	-4.08	SS Felipe Lopez	-5.81
	3B Mark Teahen	-5.61	SS Russ Adams	-8.06
	3B Mike Lowell	-5.71	SS Angel Berroa	-8.11
	3B Edgardo Alfonzo	-7.41	SS Derek Jeter	-9.14
	3B Troy Glaus	-8.78	SS Michael Young	-10.78

Results for Outfielders: Top 10 (average run value across 02-05)

Center Fielder			Left Fielder			Right Fielder		
CF	Aaron Rowand	20.56	LF	Covelli Crisp	18.93	RF	Trot Nixon	17.07
CF	Exavier Logan	20.32	LF	Carl Crawford	15.24	RF	Jeff Francoeur	13.95
CF	Laynce Nix	17.81	LF	Reed Johnson	10.14	RF	Casey Blake	10.75
CF	Jeremy Reed	15.87	LF	Randy Winn	8.57	RF	David Drew	8.46
CF	Torii Hunter	10.01	LF	Rondell White	8.46	RF	Ichiro Suzuki	8.36
CF	Andruw Jones	9.48	LF	Terrence Long	7.30	RF	Richard Hidalgo	8.12
CF	Grady Sizemore	9.24	LF	Craig Monroe	7.24	RF	Jose Cruz	6.94
CF	Willy Taveras	9.20	LF	Christopher Burke	5.54	RF	Mike Cameron	6.34
CF	Joey Gathright	8.77	LF	Frank Catalanotto	4.65	RF	Jeromy Burnitz	4.72
CF	Corey Patterson	7.36	LF	Raul Ibanez	4.60	RF	Emil Brown	4.68
CF	Mark Kotsay	-2.87	LF	Hideki Matsui	-4.77	RF	Sammy Sosa	-8.00
CF	Kenny Lofton	-4.34	LF	Eric Byrnes	-6.16	RF	Victor Diaz	-9.28
CF	Johnny Damon	-4.73	LF	Pat Burrell	-7.34	RF	Jason Lane	-9.60
CF	Dave Roberts	-7.53	LF	Ryan Klesko	-7.90	RF	Craig Monroe	-10.28
CF	Preston Wilson	-7.65	LF	Todd Hollandsworth	-8.35	RF	Bobby Abreu	-11.64
CF	Brad Wilkerson	-8.62	LF	Pedro Feliz	-8.57	RF	Jacque Jones	-12.11
CF	Cory Sullivan	-9.42	LF	Cliff Floyd	-8.95	RF	Michael Tucker	-12.65
CF	Steve Finley	-11.89	LF	Adam Dunn	-10.24	RF	Gary Sheffield	-14.59
CF	Bernie Williams	-19.23	LF	Miguel Cabrera	-16.86	RF	Wily Pena	-16.32
CF	Ken Griffey Jr.	-21.83	LF	Manny Ramirez	-22.06	RF	Larry Walker	-18.94

Comparison of Results

- Decent **overall agreement** between SAFE and UZR
  - Overall correlation between SAFE and UZR around 0.5
- No gold standard for comparison, but can examine **correlation between years**

Position	UZR 03 vs 04	SAFE 03 vs 04
1B	**0.29**	0.22
2B	0.07	**0.35**
3B	0.56	**0.69**
SS	0.04	**0.43**
CF	**0.72**	0.54
LF	**0.77**	0.73
RF	0.12	**0.41**
ALL	0.44	**0.49**

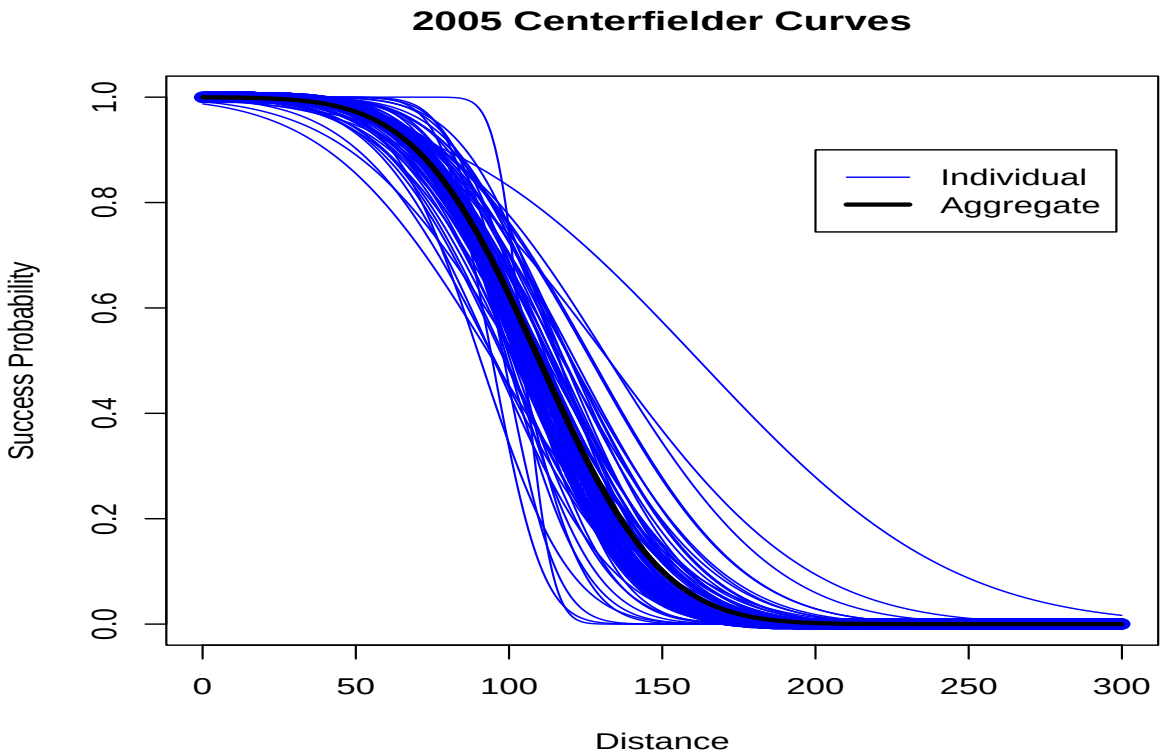
- 1B seems to be biggest problem for SAFE (even worse performance in other year-by-year comparisons)

Summary

- Higher resolution BIP data allows more detailed examination of differences between players
- Model-based approach: **smooth probability function** with estimated parameters for each player
  - Smoothing reduces variance of results by sharing information between all points near to a fielder
  - In contrast, UZR tabulates each zone independently
- **SAFE run value** aggregates individual differences while weighting for BIP frequency and run consequence
- **Year-to-year correlation** compares favorably with UZR but still has problems with some positions (eg. 1B)

Small Sample Issues

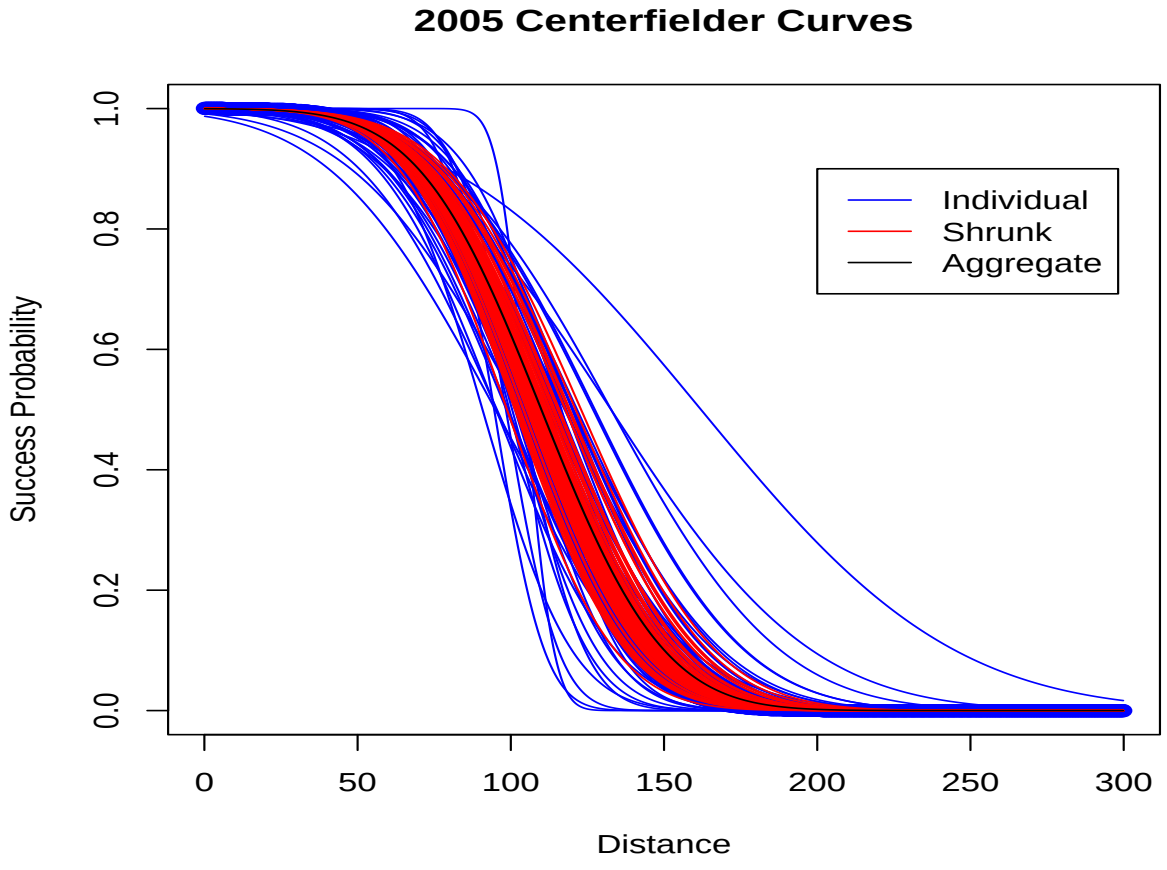
- Small samples for some players leads to **highly variable estimates** of their smooth probability curves



- Can use **hierarchical model** instead of estimating each player's curve separately

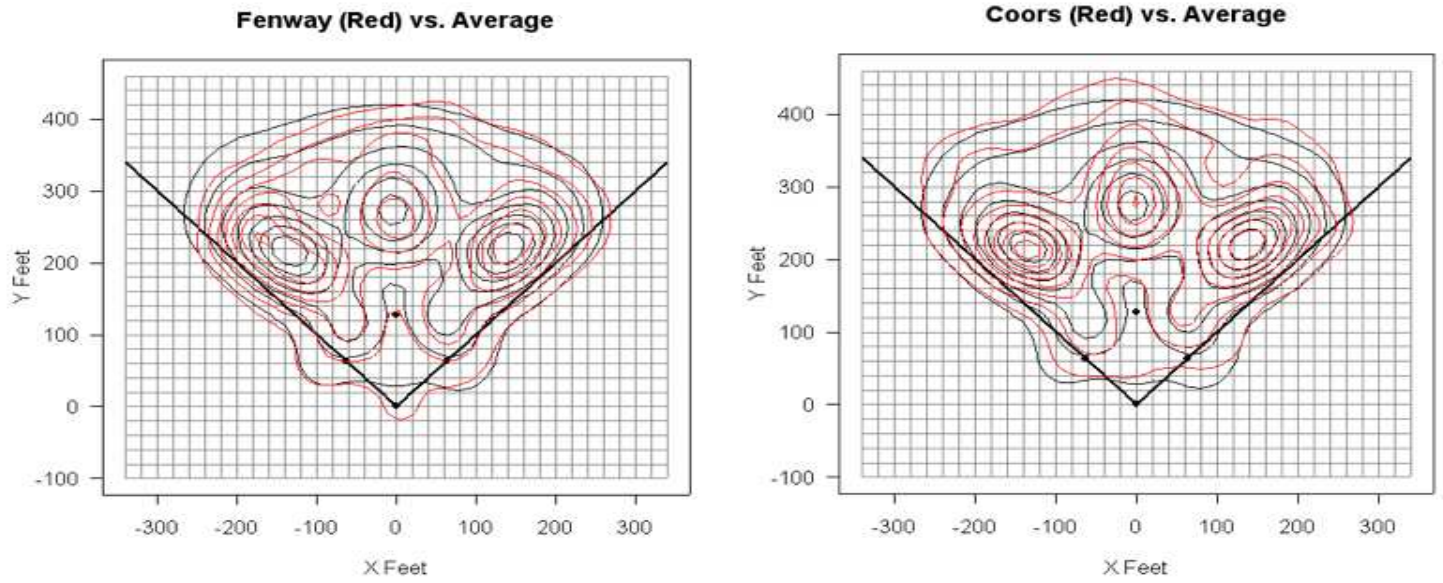
Hierarchical Shrinkage Model

- **Shares information** between parameters for each player
- Result is player curves are **shrunk** towards aggregate
- Players with small samples have curves shrunk the most



Differences between Ballparks

- Current analysis does not take into account **differences in the playing field** for different parks
- Could impact both evaluation of infielders (turf vs. grass) and outfielders (different outfield shapes)



- **Park-specific BIP densities** will account for differences in shape but will have higher variance (less data)

Thank you!

http://stat.wharton.upenn.edu/∼stjensen/research/safe.html

Google search: shane jensen safe