Understanding Underlying Dynamical Systems in Sports Psychology

Keith Heyde, Shane Ross

***Abstract*- Analyzing chaotic data can provide insight into the underlying dynamics of a nonlinear system. Here, baseball hitting average performance, an inherently chaotic signal, is coupled with a dynamical systems approach to analyzing streaky performance. Analysis on signal patterns related to critical slowing down are then applied in order to discern early warning signals for streaky behavior. Finally, underlying dynamical structures are explored through chaotic signal processing techniques in order to provide greater insight into the shape and causes of baseball streaks.**

I. Introduction

Human performance in any venture is the amalgamation of numerous input factors, both external and internal. Athletics provide an interesting and data rich field to analyze how these inputs impact a performance signal. In particular, the tendency for baseball hitters to display complex dynamical traits, such as streaks (bistability), periodicity, and chaotic jumps, make them ideal candidates for dynamic system analysis. Although this paper focuses on baseball hitting, it should be noted that the chaotic signal processing approach can be applied to variety of signals ranging from market indicators to health metrics.

Historically, the phenomenon of hitting streaks in baseball has been attributed to a number of different factors. Originally, streaks were consider the byproduct of random probabilistic factors, akin to rolling a die. However, numerous statistical approaches have rendered this understanding moot, particularly at the professional athletic level.

In Albert [2], statistical approaches to streaks are reviewed by parsing apart baseball hitting as a series of 0's (outs) and 1's (hits). From this data presentation, classical statistics such as length of longest hitting streaks, or runs above a certain average be easily calculated. Figure 1 gives examples of these statistical tools and a brief interpretation.

Figure 1

Toy Data Hits/Out Sequence:*11100000011000110001000010*

Average	0.36
Variance	0.24
Longest Hit Streak	3
Longest Out Streak	6
Average Hit Streak	1.8
Average Out Streak	3.2

	Previous At Bat		
		1	0
Current at bat	1	4	4
	0	5	11

It should be noted that by analyzing previous performance as an influencing factor for current at bat, as shown to the right of Figure 1, there is an inherent understanding of the dynamical nature of performance.

Data sets may also be presented for hitting data on a game-by-game basis. This method presents a player's performance in the form of "x-for-y" where x is the number of hits and y is the number of at bats for a given game. From such a data set, a moving average may be calculated, allowing for a signal based approach.

In this paper, this game based data approach was chosen for player psychology reasons, as a given player is more likely to reflect on previous performance after a game. Furthermore, by presenting game by game performance as rolling averages, one can turn discrete data into a more informative time series or signal. It should be noted that picking appropriate windows for calculating signals is important for a balance of smoothness and granularity, as demonstrated in Figure 2. The data for Figure 2 was taken from the professional player Mike Piazza's 2000 season performance. It will be used as sample data throughout the paper.

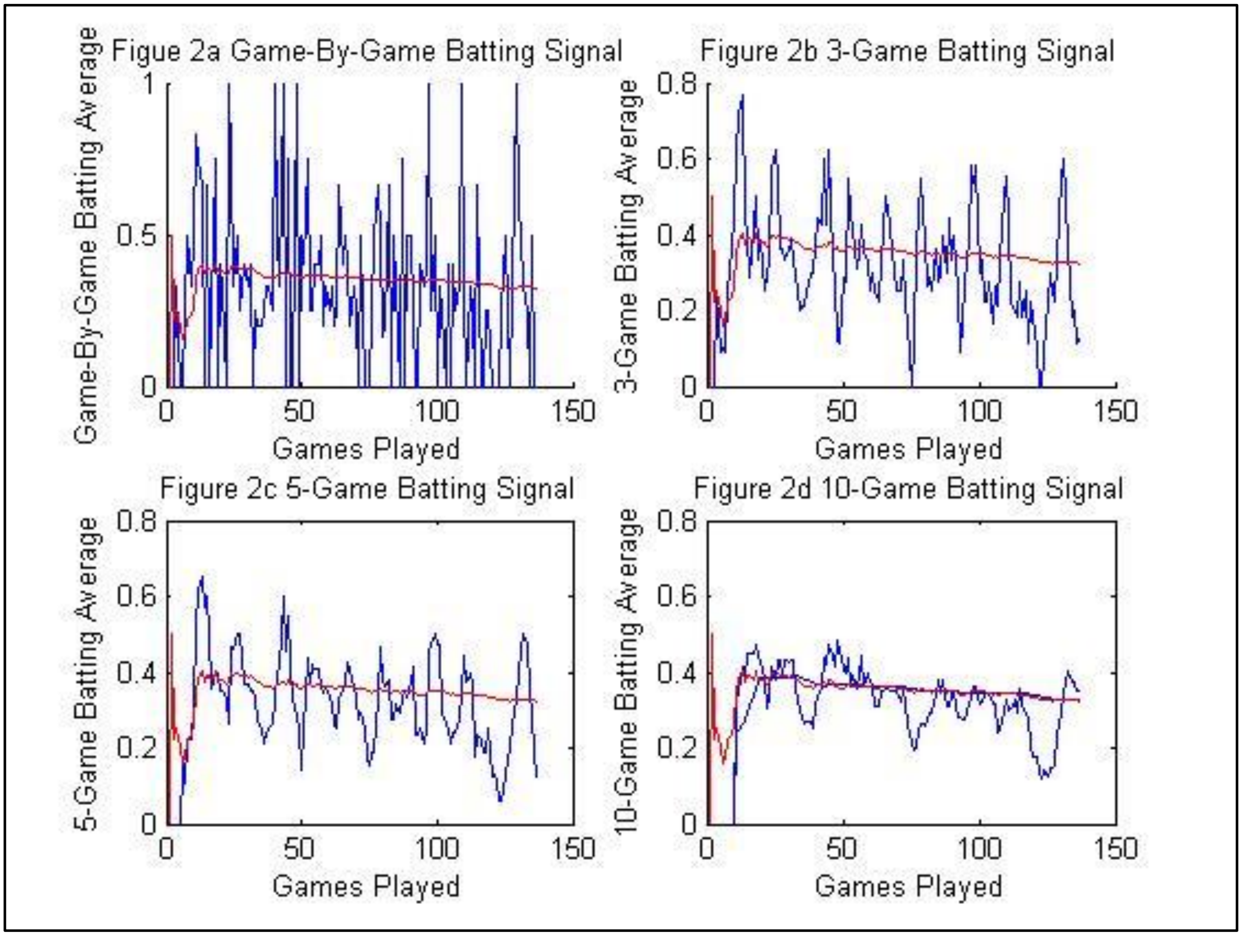


Figure 2a takes a real player's batting performance on a game-by-game basis, and then computes rolling averages on a three game (**Figure 2b**), five game (**Figure 2c**), and ten game segments **(Figure 2d).** In each case, the blue line represents the signal of rolling averages whereas the red line is a year-to-date season average. It is intuitive that different smoothing factors reveal different structures for analysis.

II. A Dynamical Systems Approach

Although statistical approaches yield certain insight, they are limited in their ability to understand and predict shifts in player performance. However, by interpreting the signal produced from rolling average as the amalgamation of underlying parameters/factors, one can approach batting performance as a dynamical systems. This approach is already used in ecology whereby population signals represent underlying dynamic parameters governed by pollution, fecundity, climate loss, and countless other inputs [11],[12],[14].

Previous studies [17] have demonstrated how the psychological phenomenon of manic-depression may be understood as two a dynamical phase space consisting of two, distinct and separate regimes, with a chaotic transition between the two. This creates a bistable structure whereby presence within one of the regimes is self-reinforced.

Similarly, this paper characterizes a given streaky player (Player A) as having hitting patterns that fall into one of two regimes. One, is a state of constant success, whereby the previous (successful) outings influence the present. The other is a regime of constant slump, whereby the previous (non-successful) outing influence the present, making the chance for a turnaround difficult. Let us call these two regimes R_m and R_d respectively. In contrast, a non-streaky player (Player B) could have an underlying dynamic structure with much less dramatic wells, perhaps hovering much around an average regime, $R_{a.}$ Figure 3 contrasts these two different types of players.

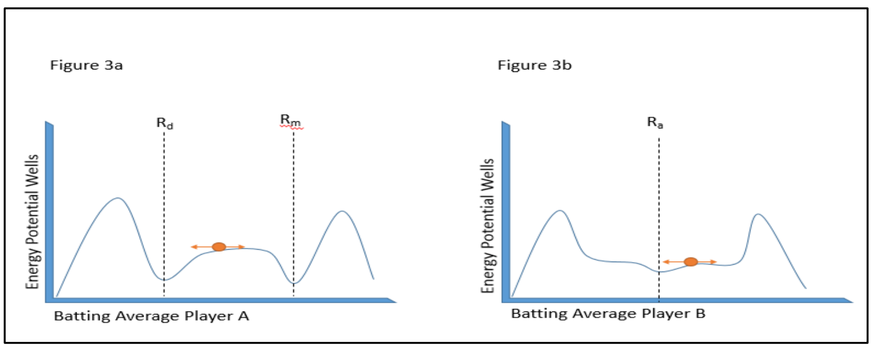


Figure 3a presents a potential dynamic well for Player A with both hot and cold streak wells. **Figure 3b** presents the potential dynamic well for Player B characterized by a gentler average well.

III. Data Preparation

In order to analyze if the bistable structure in Figure 3a exists for streaky players, it is crucial to examine real performance data. Three players were selected for this analysis: Mike Piazza (hereafter Player A) in 2000 as the classic streaky player, Derek Jeter (Player B) in 2003 as a non-streaky or consistent player and Andruw Jones (Player C) in 2000 as a player without any preconceived notions. These first two players were chosen based from a qualitative recommendations from a baseball analyst on streaky vs consistent players [10].

The game-by-game averages were then processed into a smoothed signal by taking a five game average. This game measurement was chosen due to a qualitative understanding of how long streaks generally last (~7-10 games) as well as how long a given series is in baseball (~3 games).

IV. **Data Analysis**

a. **Predicting Critical Transitions**

Figure 4, below, presents a plot the rolling batting averages for Player's A, B, and C in the first row. These plot is coupled with lines for both average batting average (red) and a single standard of deviation away from the average (yellow). From a qualitative perspective once can note that Player A seems to spend most of the time outside the standard of deviations whereas player B (excluding the extreme low around game 25 and the high around game 50) tends to reside within the standard of deviation range. This assessment is reinforced quantitatively as player A goes beyond the standard of deviate 15 unique times whereas player B only leaves 5 separate times. This gives tacit evidence for the dynamics presented in Figure 3.

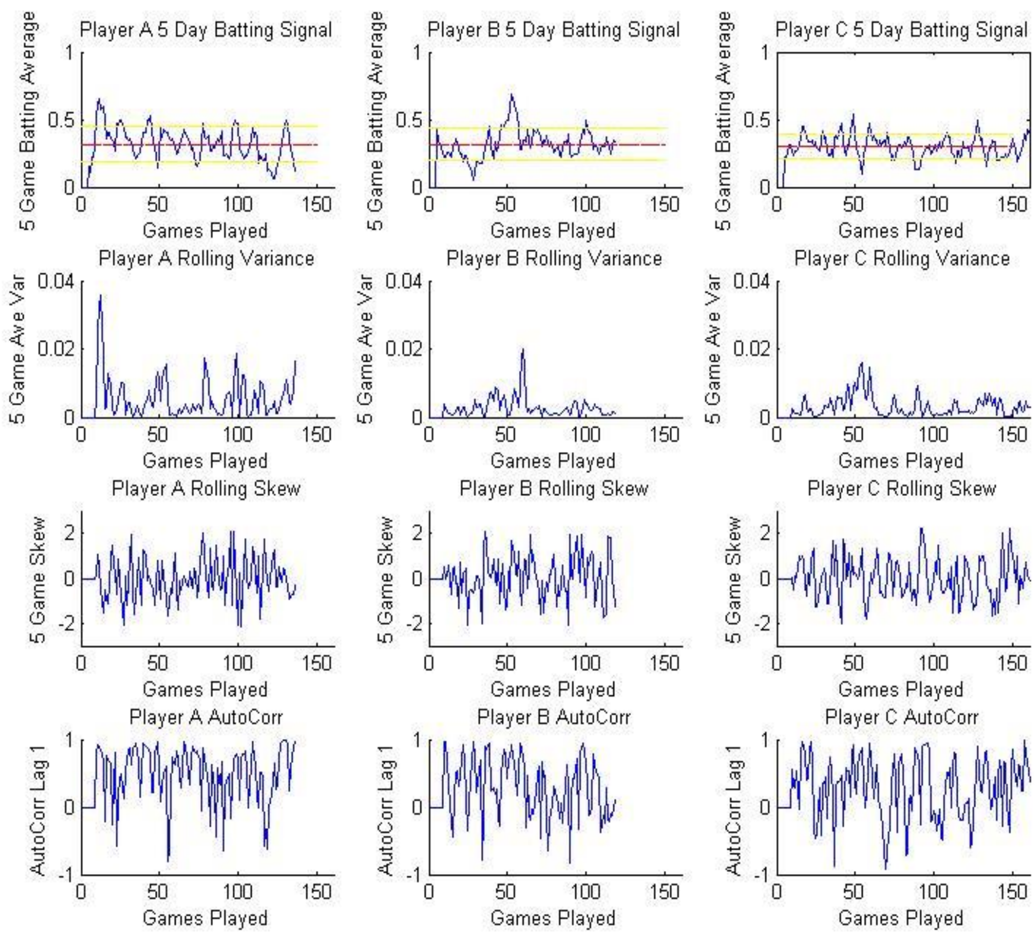


Figure 4 gives a graphical visual for analyzing the signals of Players A,B, and C. The first row gives a rolling 5 game batting average in blue, with a red line designating the mean batting average and the yellow lines representing +/- one standard of deviation away from the mean. Row two gives the 5 game period variance for each of the players and row three gives the 5 game period skewness. The final row plots the autocorrelation for a 5 game period with a lag-1 in effect.

Furthermore, there are a series of established techniques for predicting and characterizing tipping points, or regime shifts, based off of a general understandings of bistable dynamical systems theorized in Figure 3. These techniques revolve around the phenomena of critical slowing down in the signal response of a system as it approaches a critical transition. Critical slowing down can be best identified from its symptoms: slowing return to stable state from noise, an increased autocorrelation near a transition point, a one sidedness to the skew near a tipping point, and increased variance near a transition point [3],[4],[5],[6],[8],[9],[14]. The dynamical justification for this generic behavior is elaborated in Scheffer [14].

In this data, there is some evidence for critical slowing down serving as a relevant tool for streak prediction. For instance, for Player A, there are spikes in the variance and slight upward movements in autocorrelation lag 1 around games 15, 27 and 92. In each of these cases, two games later corresponding with a sharp decline in batting average. However, it is also apparent that there are many spikes in the variance for all of the players *without* accompanied average regime shifts and vice versa. Please see Appendix B for a graphical presentation of this shortcoming. Therefore one cannot definitively claim increase of variance as a sufficient signaling factor. Further study will look into the correlation between the one sidedness of the skew and the increase in the variance as a combined indicator for batting transitions.

b. Detecting Underlying Structure

Assuming batting data is the resulting one-dimensional signal from a variety of input factors, it is logical to question what the underlying dynamics are of the phase space. Is the bistable solution presented in Figure 3 a fair assumption or not? If not, how can one derive a sense of the phase space without any additional data?

Abarbanel [1] in *Analysis of Observed Chaotic Data* provides a framework for interpreting a time series and extracting underlying dynamics in order to understand the governing structure. One of the main tenants is to construct a shadow manifold made from the time delay of a signal in order to help create a phase space reconstruction. This reconstruction will in turn yield further insight into the overall dynamics of the system including parameter causality and interactions. Recent work with ecosystem dynamics [17] bolsters the validity of this approach. Figure 5 walks through the process of shadow manifold construction for the batting signal of Player A.

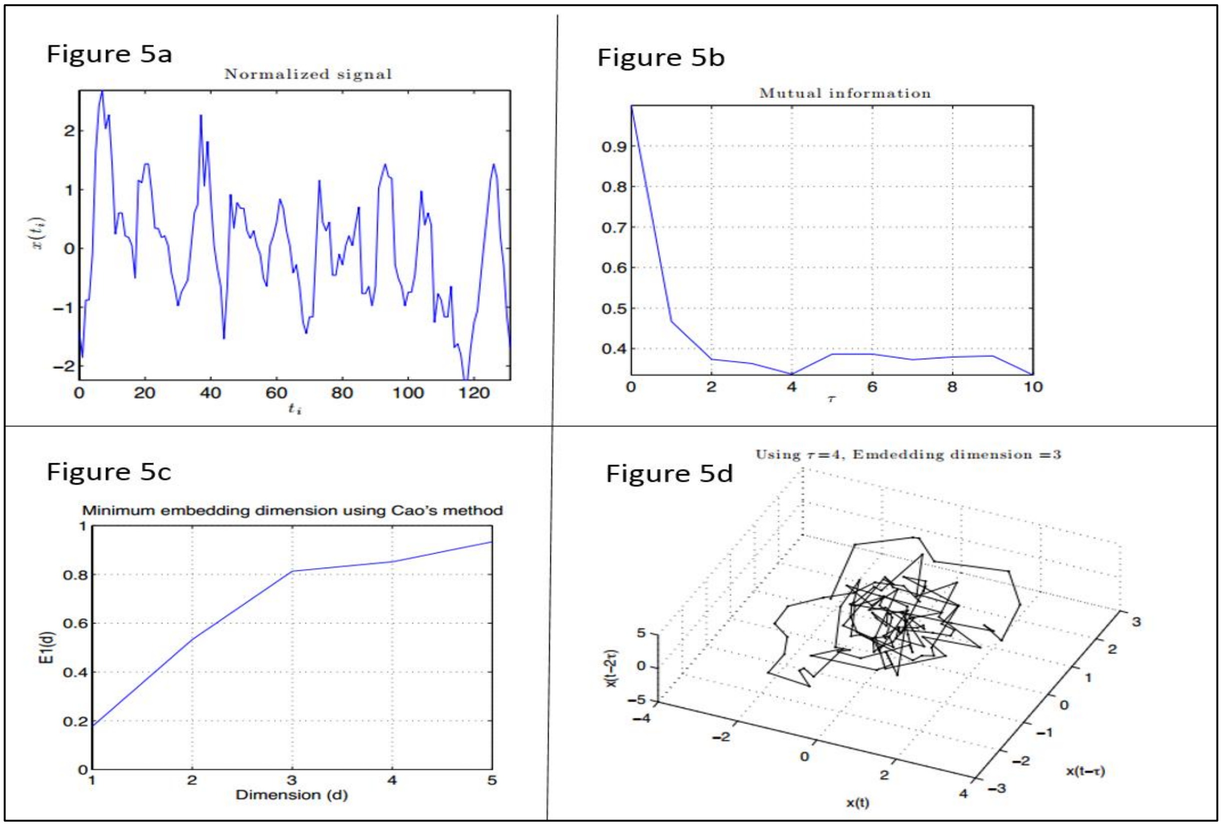


Figure 5a presents the normalized curve of Player A's 5-game averaged hitting performance. The mutual information curve is given in **Figure 5b**, which gives a prediction of what time shift, τ, will work well for constructing a shadow manifold. **Figure 5c** uses Cao's method to predict how many dimensions are necessary to form the nonlinear system. **Figure 5d** is the constructed shadow manifold under the two time shift τ's. All of these figures were created from TSTOOL, an open source MATLAB package.

From Figure 5c shows that the underlying dynamics are caused from a 3 dimensional structure. This theory is reinforced in Figure 5d where a somewhat well-defined structure can be identified. The fact that the batting performance is the consequence of three dynamic parameter input factors give us the following insights:

1) **It gives a complex understanding of how many driving factors there are in influencing performance.** In this case, the number three could be indicative of any three internal metrics, for instance sleep, psychological comfort, mental awareness, etc. It should be noted that further isolation of these metrics and uncovering underlying dynamics, say for instance of periodic psychological behavior, can parse out dominant variables to help complete the full dynamical picture.
2) **Dominant dynamical features could be delicately tuned to influence better performance once the overall system dynamics is better understood.** Take, for instance, the system in which small perturbations in psychological feelings of comfort at the plate highly influence the positive or negative feedback loops of the overall batting performance. Once this highly sensitive aspect of

the system is isolated via shadow manifold analysis, it can be targeted and manipulated in order to best influence performance.

V. Conclusions and Future Research

This work reviewed classic statistical approaches to understanding streaks in baseball performance. Then it expanded upon previous work by treating a baseball rolling batting average as a signal for an underlying dynamical system. This signal was then processed and analyzed in order to note for trends in predicting critical transitions. Finally, a reconstructed phase space and shadow manifold were created in order to ascertain information about underlying dynamical structure for a player's batting performance.

Future work will be directed in two main areas. One will be to use higher order dynamical system tools to identify attractors within the reconstructed phase space. This will yield insight into the nature and duration of streaks. Secondly, the psychological link between mental well-being and athletic performance will be explored as it pertains to phase space. The fact that the two are linked is bolstered by numerous professional teams having psychologists on staff and the success of mentally disciplined athletes such as Phil Jackson and his Zen approach. Choosing a mental metric, even if qualitative, and attempting to construct some phase space correlation between that and batting performance as shown in Figure 6 would give great insight into the underlying dynamics of the full system.

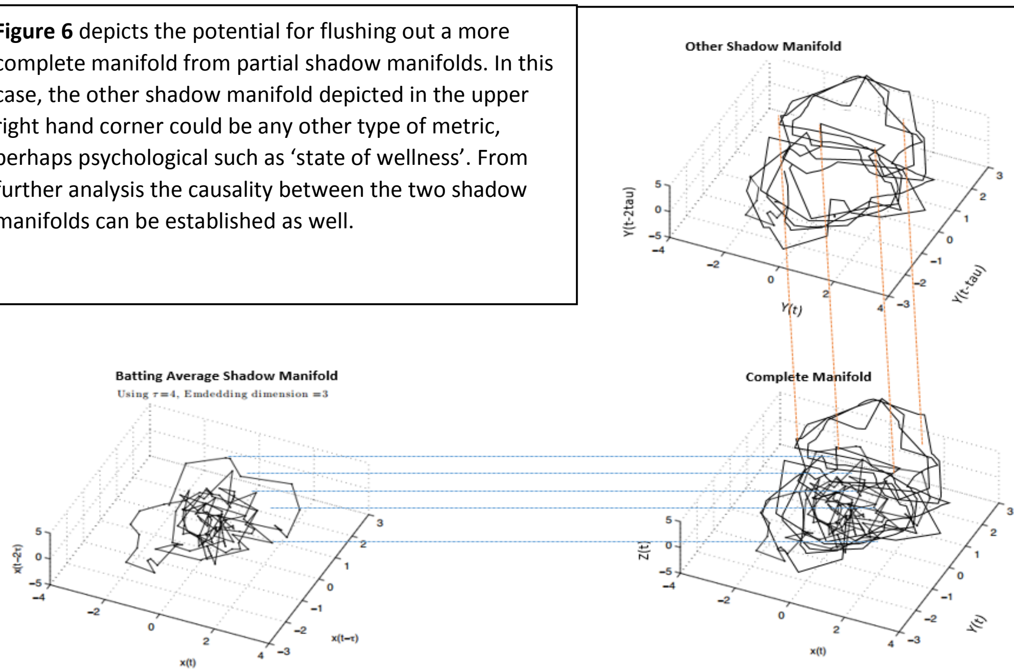


Figure 6 depicts the potential for flushing out a more complete manifold from partial shadow manifolds. In this case, the other shadow manifold depicted in the upper right hand corner could be any other type of metric, perhaps psychological such as 'state of wellness'. From further analysis the causality between the two shadow manifolds can be established as well.

Treating batting performance as a resulting chaotic signal provides insight into hidden patterns and structure otherwise lost in a linear or statistical processing approach. Considering a player's performance as the amalgamation of multiple input factors can help better predict, and potentially influence slumps in order to increase chances for success. This, coupled with an understanding of team network dynamics (briefly explored in Appendix A) can potentially yield both great insight into system behavior, and practical success. Future research will focus on categorizing the various dynamical structured derived from the time lag method with player's performance in order to extrapolate rules for streaky behavior.

Appendix A

Exploring Team Network Dynamics

A team's performance, whether measured by offensive production, win/loss, or some combination of metrics can be considered the resulting signal of a complex network of interactions. This network is formed from the numerous interactions and relationships between players on a team. When coupled with a given player's personal performance dynamics presented in the body of the paper, this can create a vastly complicated system of interactions.

Take, for instance, the sample network presented below:

Sample Model Parameters

For simplicity sake, let us look at a magical model wiffleball team with only four players. One can categorize these players as P1, P2, P3, and P4, which will correspond to their batting order. Furthermore, let us define the following:

G+ = the ability of a player to project positive influence

G- = the ability of a player to project negative influence

I+ = the ability of a player to be positively influenced

I- = the ability of a player to be negatively influenced

And let us consider each of these metrics normalized to some 1-10 scale for G values and from a -5 to 5 scale for I values.

So, on our model wiffeball team, let us consider the following players:

P1 is the 'voice of the team' and is the typical loud, in your face, athletic leader. Let us say P1 has a G+ = 8, G- = 8, I+ = 2, I- = 2.

P2 is the 'grizzled veteran' who is highly respected, but is by no means bombastic. P2 is profiled as G+ = 5, G- = 8, I+=-4, I-=-5

P3 is the 'scared rookie' who is susceptible to influence easily. P3 has a player profile of G+=2, G-=1, I+=3, I-=5

P4 is the 'steady baseline' that acts as a neutral buffer on the team. P4 can be described as G+=5, G-=5, I+=-1, I-=-1.

These relationships are expressed in the governing equations below, which may be coded and simulated with various initial conditions to determine network behavior.

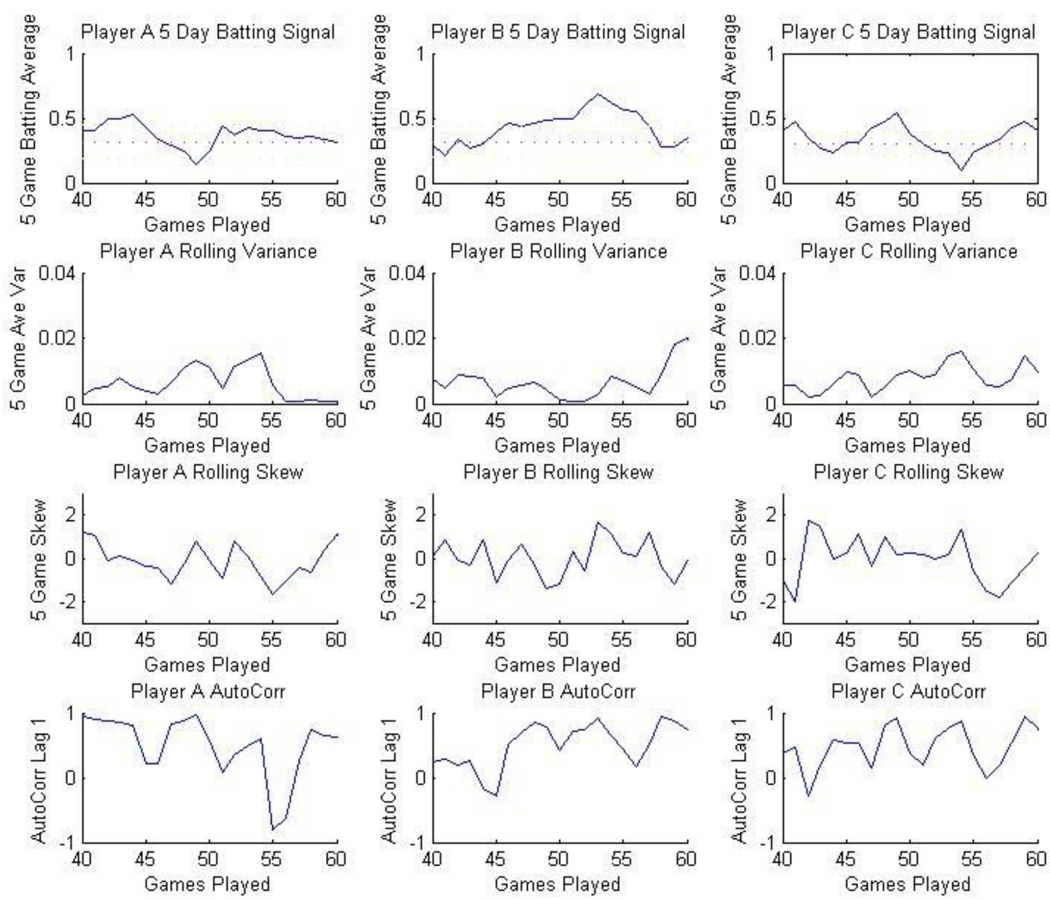
Governing Equations

Player 1:		Player 3:	
Positive	P1(t+1) = 7*P2(t) + 7*P4(t)	Positive	P3(t+1)=11*P1(t)+8*P2(t)+8*P3(t)
Negative	P1(t+1) = 6*P2(t) + 3*P4(t)	Negative	P3(t+1)=13*P1(t)+6*P2(t)+3*P3(t)
Player 2:		**Player 4:**	
Positive	P2(t+1) = 4*P1(t) + P4(t)	Positive	P4(t+1)=7*P1(t)+4*P2(t)+P3(t)
Negative	P2(t+1) = 3*P1(t) + 3*P4(t)	Negative	P4(t+1)=7*P1(t)+6*P2(t)

One can see from these equations that pivotal players, such as Player 2, have a high impact on other players on the team. Therefore, understanding Player 2's batting dynamics (such as those presented in Figure 5) are especially important for managing team performance. Further study into the nuanced relationship between team interactions is necessary to fully understand and control these processes.

Appendix B

Zoom in on batting performance with critical transition metrics associated. Note that there is no clear alignment between a rise in AutoCorrelation Lag 1 and variance, or one sidedness of the skew, with a drop in the batting signal.



References

1. Abarbanel, Henry. *Analysis of Observed Chaotic Data*. New York. Institute For Nonlinear Science. 1997. Print.
2. Albert, J. (2008) "Streaky Hitting in Baseball". *Journal of Quantitative Analysis in Sports*. 4(1): 3.
3. Biggs R, Carpenter SR, Brock WA (2009) "Turning back from the brink: Detecting an impending regime shift in time to avert it." *PNAS* 106: 826–831.
4. Brock WA, Carpenter SR (2010) "Interacting regime shifts in ecosystems: Implication for early warnings." *Ecol Monogr* 80: 353–367.
5. Carpenter, S. R. & Brock, W. A. (2006) "Rising variance: a leading indicator of ecological transition". *Ecol. Lett.* 9: 308–315.
6. Carpenter SR, Cole JJ, Pace ML, Batt R, Brock WA, et al. (2011) Early warnings of regime shifts: A whole-ecosystem experiment. Science 332: 1079–1082.
7. Chisholm, R. A. & Filotas, E. "Critical slowing down as an indicator of transitions in two-species models". *J. Theor. Biol.* 257:142–149
8. Foley JA, Coe MT, Scheffer M, Wang G (2003) "Regime shifts in the Sahara and Sahel: Interactions between ecological and climatic systems in northern Africa." *Ecosystems* 6: 524–532.
9. Guttal V, Jayaprakash C (2008) "Changing skewness: an early warning signal of regime shifts in ecosystems." *Ecol Lett* 11: 450–460.
10. Hand, Kevin. Personal Interview. 6 April 2014.
11. May, R. M. (1977) "Thresholds and breakpoints in ecosystems with a multiplicity of stable states". *Nature* 269, 471–477.
12. May, R. M., Levin, S. A. & Sugihara, G. (2008) "Ecology for bankers". *Nature* 451: 893–895.
13. Rinaldi, S. & Scheffer, M. Geometric analysis of ecological models with slow and fast processes. *Ecosystems* 3: 507–521.
14. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2001) "Catastrophic shifts in ecosystems". *Nature* 413: 591–596.
15. Scheffer, M, et al. (2009) "Early-warning signals for critical transitions" *Nature.* 461: 53-59.
16. Sugihara et al. (2012) "Detecting Causality in Complex Ecosystems". *Science*. 338(6106): 496-500.
17. van de Leemput IA, et al. (2014) "Critical slowing down as early warning for the onset and termination of depression". *PNAS* 111(1):87–92.