

Enhanced Gameday 系統之王建民投球球路分析

許伯陽* 周靈山**

*南華大學

**吳鳳技術學院

摘要

本研究以大聯盟網站 Enhanced Gameday 系統所提供之公用數據，分析洋基隊投手王建民在 2007 年 8 月 30 日與紅襪隊對戰的球路軌跡，經由球速、水平位移分量、垂直位移分量以及變化量等四個變數，透過集群分析來分辨投球球路，分類命名後，經由變異數分析確認差異的顯著性，最後再由統計圖表來顯現各球路在相關變數上的表現。研究結果顯示，在 85 球有紀錄數據的投球分析下，分為快速球（58 球，68.2%）、變速球（7 球，8.2%）、滑球（20 球，23.6%）等三類。經由變異數分析得知三種球路在四個變數上都達顯著差異($p < .05$)。事後比較發現，除了水平位移分量上的快速球與變速球的差異不顯著外，其餘都達顯著差異($p < .05$)。最後證實，王建民投球球路可經由 EG 系統的數據，進行客觀的量化分析。

關鍵字：集群分析、快速球、變速球、滑球

通訊作者：許伯陽，臺北縣三峽鎮國際一街 40 號 4 樓之 3，
南華大學體育室。

聯絡電話：0927-557630

E-mail：poyang.hsu@msa.hinet.net

The Analysis of Chien-Ming Wang's Pitch by Enhanced Gameday

*Hsu Po-Yang**

*Ling-Shang Chou***

**Nan-Hua University*

***Wu-Feng Institute of Technology*

Abstract

This study was analyzed the pitch of Chien-Ming Wang in the game versus the Boston Red Sox on August 30th, 2007. The data was provided by Major League Baseball website from Enhanced Gameday (EG). The cluster of analysis, ANOVA, and descriptive statistical analysis were used to classify and confirm the difference through four variables, which including start speed, pfx_x, pfx_z, and break_length. Based on these analysis, there were several conclusions be defined. The first conclusion showed the pitches were classified to three categories, which including Fastball (58, 68.2%), Change up (7, 8.2%), and Slider (20, 23.6%). The second one pointed out these three kind of pitches have the significant difference in four variables ($p < .05$). Then, in post hoc, they all have the significant difference besides Fastball and Change up in pfx_x ($p < .05$). In the final, the pitch of Chien-Ming Wang can be executed in the objective quantitative analysis by using EG system data.

Keywords: cluster of analysis, fastball, change up, slider

壹、緒論

在過去的棒球學術研究中，一直無法將投手球路的飛行軌跡數據化，只能依靠專家透過攝影器材的畫面，經過研判而得知投手每一球的路線軌跡，並以各式球種稱之。後期的技術應用測速槍的功能，將投手投出每一球的速度顯示出來，供作研究者一項客觀的數據資料，但投手球路軌跡的分析，除了球速之外，還包含了變化幅度、位移幅度以及放球點、進壘點等數據的相互配合，才能做出有別於主觀質性分析之外的客觀量化分析。

以蔡文星（2005）對日本投手松坂大輔的投球分析為例，經由專家透過錄放影機慢動作分析，在 25 個方格中紀錄每一球的球種、速度與落點分佈，以此數據作為後續的 t 考驗與變異數分析等統計處理，在依此作松坂大輔投球的投球策略分析。國內其他學者，鐘宇政（1999）對 IBA 世界青棒賽投手曹錦輝之配球模式分析，楊賢銘（2000）對 1999 年亞洲盃棒球賽中、日、韓三國投手投球分析，廖文男等（2003）對 2001 年世界盃成棒賽投手球路分析等，都是透過專家與電視來判別球路，並以此來作後續的球路分析，這部分的研究較屬於主觀的質性分析輔以量化的統計分析。陳五洲（2003）認為攝錄影器材、電腦設備與嚴謹的實驗控制所獲得的訊息，提供專家與教練有效率的分析，在無法提供更為精確且客觀的判別依據下，上述的球路判別是可行的方式。而其他學者也做出類似的研究，如 Chang et al (2002)透過攝影機的鏡頭擷取影像，研究投手的投球球路，同樣利用隱藏馬可夫模式(Hidden Markov Model, HMM)來分析投球球路。以及 Gueziec (2002)利用 Sportvision 公司所設計的 K Zone 好壞球辨別系統，追蹤投球的飛行軌跡，決定每一球的好壞球的品質，以及分辨球種，最後建立投球資訊的配球模式。在上述的文獻中，對於投手的球路軌跡，主要是依據主觀的專家分析，或是小規模的使用未模式化的影像擷取技術，這些數據很難達到大量化客觀的品質，容易因專家與影像擷取的差異而產生不同的數據，而 Enhanced Gameday 系統所提供的數據正可以解決此一問題。

2007 年美國大聯盟網站正式將 Sportvision 公司所設計的 Pitch f/x 投球軌跡系統，加入網路即時轉播系統 Enhanced Gameday 中，該系統主要提升投手投球球路飛行的軌跡重建技術，使網路使用者能即時觀看每一球的飛行軌跡與進壘角度等模擬影像，輔與原本 Gameday 所擁有的文字敘述與其他強大的即時功能，帶給網路使用者更貼近比賽的臨場感。為了重建投手投球的飛行軌跡，Enhanced Gameday 建構大量且多樣的數據作為模擬影像的依據，這些數據比先前的 Gameday 系統所提供的人工紀錄更為客觀，紀錄下每位投手每一球的相關資訊，排除專家與人工紀錄的判別方式，避免人為的疏失與不客觀。

在使用多年的 Gameday 系統後，2006 年美國大聯盟季後賽在大聯盟網站(MLB. com)

正式啓用升級版的 Gameday 系統，稱爲 Enhanced Gameday。這套系統增加由 Sport vision 公司所建構的 Pitch f/x 系統，將原本只有人工紀錄的即時比賽文字訊息，增加 3D 立體的球路軌跡，使網路收視戶在得到即時文字訊息的同時，也能看到電腦模擬的球路進壘軌跡，增加網路社群收看比賽的臨場感。而此模擬技術的背後是經由多台攝影機的拍攝定位與電腦程式的分析所建立的數據資料，之後再將數據轉換成電腦即時畫面。2007 球季結束後，MLB.com 期望能在 30 個大聯盟球隊主球場建立 Enhanced Gameday 系統，目前只有七座球場建立，其他陸續建構中。洋基主場在 8/28~8/30 對紅襪三連戰時，正式啓用此系統，王建民在第三戰先發主投，Enhanced Gameday 系統紀錄每一球的數據，爲本研究數據來源。本研究即是透過 Enhanced Gameday 系統所提供在大聯盟網站的公用數據，分析洋基投手王建民的球路，透過集群分析來分辨投球球路，分類後的球路經由變異數分析確認差異的顯著性，最後再由統計圖表來顯現各球路在相關變數上的表現。經由量化的數據與統計分析建立客觀的分析模式。

貳、研究方法

一、研究對象

王建民 2007/8/30 對波士頓紅襪隊的比賽，主投七局，總用球數 103 球，70 個好球，33 個壞球，無失分，投出五次三振。Enhanced Gameday 系統在 103 球中，記錄到 85 球，沒有紀錄到的是第一局 3 球，第五局 15 球。本場比賽的比賽強度與王建民的優異表現是選爲本研究數據分析的重要因素，另外，王建民在球季後期的投球表現內容丕變，比起去年球季依賴直球的比例甚高的配球模式，本季後期的球種豐富許多，在分析上能夠得到多樣的結論，這是另一項重要因素。

二、Enhanced Gameday 系統的變數說明

Enhanced Gameday 系統爲了構成電腦即時的球路軌跡 3D 模擬影像，必須提供豐富的背景數據，這包括放球點和進壘點的速度、加速度、位置，以及軌跡變化的幅度等等，參閱圖 1，實線箭頭所畫的軌跡爲投手投球的實際拋物線，虛線箭頭所畫的軌跡爲不受球體旋轉的預測軌跡，兩者在進壘點的差距爲位移量(p_{fx})，而實際拋物線到放球點與進壘點直線的最大距離，稱爲變化量(break_length)，此爲本研究所使用到重要的變數定義，詳細的定義如下：

1. start_speed：放球點速度。本研究以放球點的速度爲球速，單位爲英哩/小時。
2. Pitch f/x (p_{fx})：位移量。棒球旋轉所造成實際進壘點與理論拋物線的差距，分爲水平和垂直分量。
 - (1) p_{fx_x} ：水平位移分量。以捕手爲面，理論拋物線進壘點爲原點，右邊爲正(+X)，左邊爲負(-X)，單位爲英吋。

(2) px_z ：垂直位移分量。以捕手為面，理論拋物線進壘點為原點，往下為負(-Z)，往上為正(+Z)，單位為英吋。

3. $break$ ：變化點。實際球路軌跡與放球點到進壘點直線之間的最大差距點。

(1) $break_length$ ：變化量。變化點到放球點與進壘點直線的距離。單位為英吋。

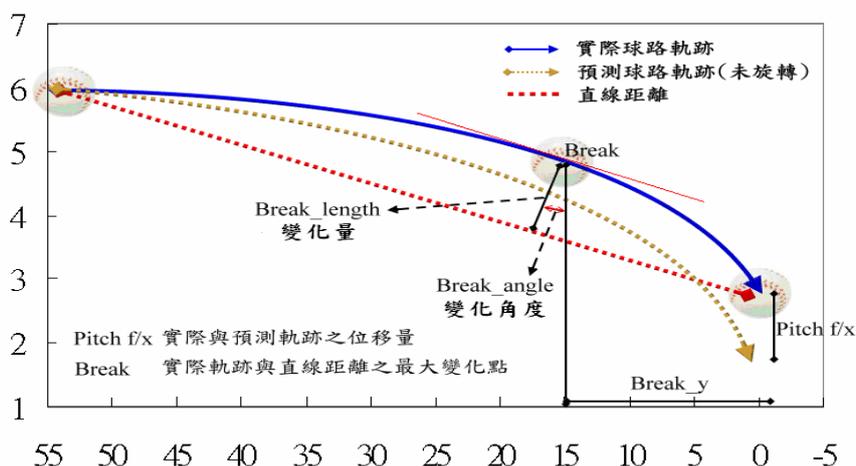


圖 1 Enhanced Gameday 系統的變數示意圖

三、資料分析

本研究依據 Enhanced Gameday 系統所提供的數據，包括球速($start_speed$)、水平位移分量(px_x)、垂直位移分量(px_z)、變化量($break_length$)等變數，使用階層式集群分析(Hierarchical cluster analysis)來分辨王建民 8/30 的投球球種。整併後的球種經由獨立樣本單因子變異數分析(One way ANOVA)，確認差異的顯著性。最後再由統計圖表來顯現各球種在相關變項上的表現。

參、結果與討論

一、階層式集群分析

使用階層式集群分析的集結法(agglomerative method)，先計算各觀察體間的距離或組內誤差矩陣，將最接近的兩個觀察體合併成一個群體，重複合併直到合併成一個集群（陳正昌等，2005）。本研究使用華滋最小變異數法(Ward's Minimum Variance)進行合併。在 85 球的分析當中，透過球速($start_speed$)、水平位移分量(px_x)、垂直位移分量(px_z)、變化量($break_length$)等變數，將相似的球路歸納成一個集群的結果，最多有 84 個集群，最少 1 個集群，圖 2 將連結距離的合併過程會製成陡坡圖，橫軸為合併集群數，縱軸為集結法的係數。由圖 2 可知，集群數由 84 個到 3 個，連結距離都是緩慢增加，3 個集群之後即大幅增加，因此 85 球分為 3 個集群較為恰當。

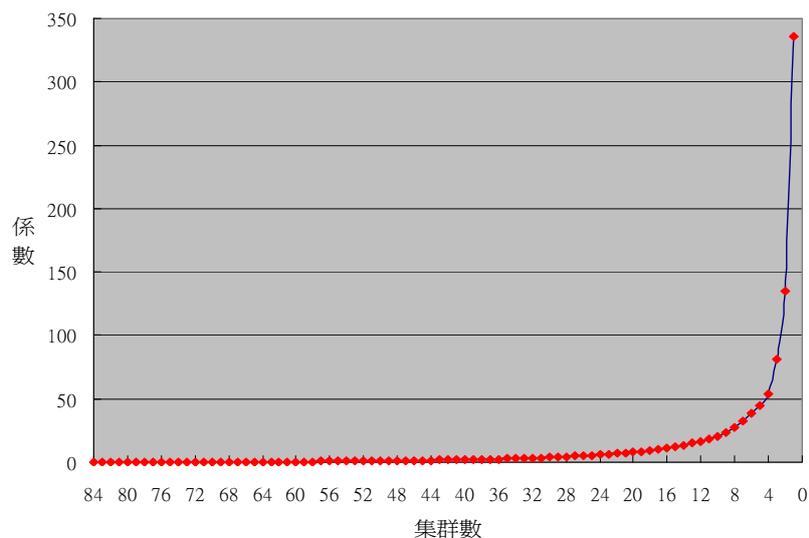


圖 2 集群分析陡坡圖

另外，表 1 為集群分析的判斷指標，BIC 與 AIC 指標以 3 個集群的指標最低(177.59、118.97)，BIC 與 AIC 改變比率(Ratio of change)除了不分群之外，以 2 組最佳，3 組次之(.043, .217)，在一般的集群分析中常會遇到當變項的數據差異性不大時，愈少的群組愈能提高各群組的組間變異。在本研究的分析中，不僅需要依靠集群分析來分類球種，也需要多樣的球種來作後續的球種表現分析，因此選擇將球種分為 3 個集群。

表 1 集群分析判斷指標摘要表 (~代表省略)

集群數	BIC	BIC 改變率	AIC	AIC改變率
1	269.205		249.664	
2	181.372	1.000	142.290	1.000
3	177.588	.043	118.965	.217
4	208.820	-.356	130.655	-.109
~	~	~	~	~

3 個集群的球速、水平、垂直位移分量以及變化量等變數的資料如表 2，發現集群 1 共有 58 球，其球速最快，達 94.37 英哩/小時，水平位移分量最大，達-6.25 英吋，負值表示往右打者內角位移，垂直位移分量居中，5.94 英吋，變化量最小，5.27 英吋，稱集群 1 為「快速球」。集群 2 共有 7 球，球速最慢，平均 79.34 英哩/小時，水平位移分量居中，為-3.89 英吋，負值表示往右打者內角位移，變化量居中，為 6.64 英吋，垂直位移分量最大，為 8.26 英吋，因 Pitch f/x 系統設定垂直位移分量的距離為預測進壘點

低於實際進壘點（參閱圖 1），所以距離愈大表示球路軌跡愈平，距離愈小表示球掉落的幅度愈大，愈接近無球體旋轉的預測進壘點，稱集群 2 為「變速球」。集群 3 共有 20 球，球速居中，平均為 84.38 英哩/小時，水平位移分量最小，為 3.24 英吋，正值表示往右打者的外角位移，垂直位移分量最小，為 .57 英吋，表示掉落的幅度最大，變化量最大，為 8.58 英吋，稱集群 3 為「滑球」。

表 2 三種球路各變項摘要表

分類	球數		球速		水平位移分量		垂直位移分量		變化量	
	數量	百分比	平均數	標準差	平均數	標準差	平均數	標準差	平均數	標準差
1	58	68.2%	94.37	1.17	-6.25	2.40	5.94	1.95	5.27	.88
2	7	8.2%	79.34	3.83	-3.89	1.35	8.26	1.10	6.64	1.03
3	20	23.5%	84.38	1.49	3.24	2.89	.57	1.61	8.58	.75
總和	85	100.0%	90.78	5.66	-3.83	4.68	4.87	3.07	6.16	1.64

二、單因子變異數分析

經由集群分析得到的球種分別是快速球、變速球與滑球等三種，後續探討三球種在四個變數上的差異顯著性，為獨立樣本單因子變異數分析(one-way ANOVA)。如表 3 所示，在球速、水平位移分量、垂直位移分量和變化量等四個變數上都達顯著差異($p < .05$)。事後比較(Post hoc test)上使用 Scheffe 法，除了水平位移分量上，快速球（平均數-6.25）與變速球（平均數-3.89）的差異不顯著外（平均數差-2.36， $p = .062$ ），其餘各組各球種均有顯著差異($p < .05$)。

表 3 三球種的變異數分析摘要表

		SS	df	MS	F 值	P 值
球速	組間	2483.37	2	1241.69	490.112	.000
	組內	207.75	82	2.53		
	總和	2691.12	84			
水平分量	組間	1339.89	2	669.94	110.566	.000
	組內	496.86	82	6.06		
	總和	1836.74	84			
垂直分量	組間	517.07	2	258.54	77.281	.000
	組內	274.32	82	3.35		
	總和	791.40	84			

續表 3 三球種的變異數分析摘要表

	組間	164.99	2	82.49	109.818	.000
變化量	組內	61.60	82	0.75		
	總和	226.58	84			

*p<.05

三、三球種的敘述統計

三球種在四個變數中的平均數、標準差、最大值、最小值等資料如表 4 所示，球速上，快速球介於 89~97 英哩/小時，平均 94.37 英哩/小時，變速球介於 71~82 英哩/小時，平均 79.34 英哩/小時，滑球介於 80~87 英哩/小時，平均為 84.38 英哩/小時。水平位移分量上，快速球介於-10.90~ -1.97 英吋，平均-6.25 英吋，變速球介於-5.24~ -1.72 英吋，平均-3.89 英吋；滑球介於-6.33~7.69 英吋，平均 3.24 英吋。垂直位移分量上，快速球介於 2.52~11.39 英吋，平均 5.94 英吋；變速球介於 6.09~9.16 英吋，平均 8.26 英吋；滑球介於-3.16~2.56 英吋，平均 0.57 英吋。變化量上，快速球介於 3.1~6.7 英吋，平均 5.27 英吋；變速球介於 5.7~8.6 英吋，平均 6.64 英吋；滑球介於 7.5~10.1 英吋，平均 8.58 英吋。

表 4 三球種敘述統計摘要表

分類	項目	球速(英哩/小時)	水平分量(英吋)	垂直分量(英吋)	變化量(英吋)
快速球	平均數	94.37	-6.25	5.94	5.27
	標準差	1.17	2.40	1.95	.88
	最小值	89	-10.90	2.52	3.1
	最大值	97	-1.97	11.39	6.7
	範圍	8	8.94	8.86	3.6
變速球	平均數	79.34	-3.89	8.26	6.64
	標準差	3.83	1.35	1.10	1.03
	最小值	71	-5.24	6.09	5.7
	最大值	82	-1.72	9.16	8.6
	範圍	11	3.53	3.07	2.9
滑球	平均數	84.38	3.24	.57	8.58
	標準差	1.49	2.89	1.61	.75
	最小值	80	-6.33	-3.16	7.5
	最大值	87	7.69	2.56	10.1
	範圍	7	14.02	5.72	2.6

四、3D 立體圖與兩變數平面圖的比較

為了解釋經由集群分析所產生三種球路的相對位置，本研究透過 3D 立體圖來說明，圖 3 為實際與預測的進壘點之間的距離(pfx)與球速之間的 3D 立體圖，X 軸為水平位移分量，Z 軸為垂直位移分量，Y 軸為球速。圖 4 則是 pfx 與變化量之間的 3D 立體圖，Y 軸改為變化量，其餘相同。由圖 3 的速度差異可以分別出三種球路的高低相關位置，滑球的水平分量最大，垂直分量最小，因此在圖 3 的 3D 圖中可明顯分辨，而快速球和變速球除了速度差異明顯外，水平和垂直分量差異不大。圖 4 中，滑球的變化量最大，配合最大的水平分量與最小的垂直分量，可明顯辨識，但是快速球和變速球的變化量、水平與垂直位移均交互出現，因此在圖 4 中不易區分，必須由兩變數間的平面圖來觀察。

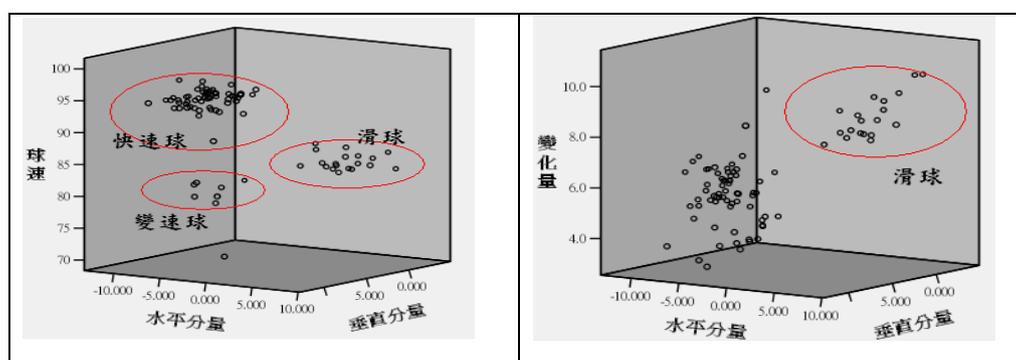


圖 3 pfx 與球速 3D 立體圖

圖 4 pfx 與變化量 3D 立體圖

兩變數平面圖顯示，在垂直分量與變化量間的關係，如圖 5，發現明顯的負相關，也就是當變化量愈大時，垂直分量愈小，甚至出現負值，最小值為-3.16，因 Pitch f/x 系統設定預設原點為每一球球路在不受任何旋轉的情況下的預設球點，而實際球點會受到球體自旋的影響而上升與旁移，大部分位移是如此，也有可能低於預設球點，愈接近預設球點，其變化幅度愈大，因此產生圖 5 的負相關，滑球的變化量與垂直分量均會大幅改變。而滑球的水平分量與其他球種也有明顯的區別，如圖 6 所示，滑球的水平分量均為正值，也就是往右打者的外角位移，其餘球種往內角位移，滑球是王建民變化幅度最大的球種。另外，變速球的變化量稍微大於快速球，因此在圖 5 與圖 6 中，變速球位於快速球的右側，觀察出兩者間的區別。

球速在三球種的變數裡是最重要的鑑別因素，球速與其他變數的相關圖均能輕易的區分三者間的差異，如圖 7 所示，速度與變化量間的相關圖，快速球球速快之外，變化量最小，滑球速度居中，但變化量最大，而變速球速度最慢，變化量居中。在圖 8 中，顯示滑球的垂直變化幅度最大，水平位移為正值，而快速球與變速球的水平分量差異不

顯著，與三球種變異數分析的事後比較結果一致。

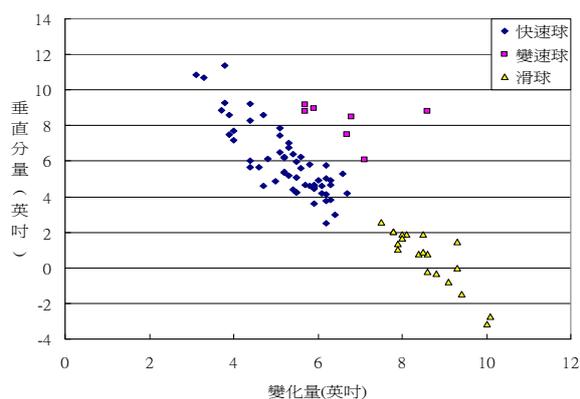


圖 5 Pfx_Z 與變化量相關圖

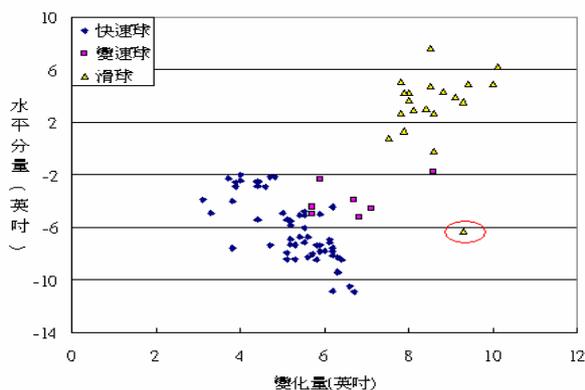


圖 6 Pfx_X 與變化量相關圖

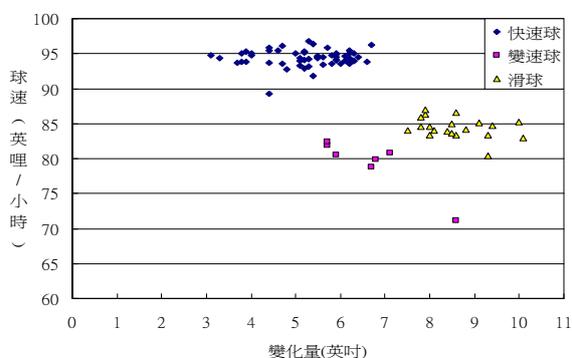


圖 7 球速與變化量相關圖

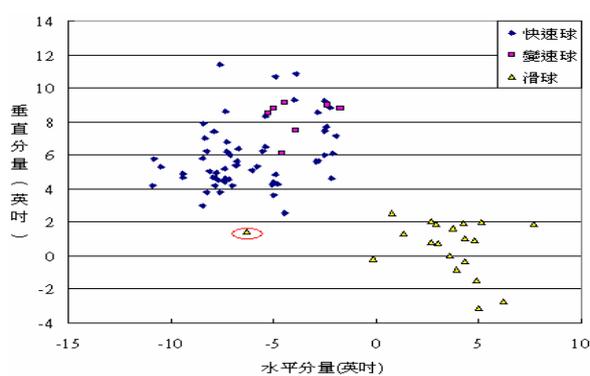


圖 8 Pfx_X 與 Pfx_Z 相關圖

肆、結論與建議

本研究以大聯盟網站 Enhanced Gameday 系統所提供之公用數據，分析洋基隊投手王建民在 2007 年 8 月 30 日與紅襪隊對戰的球路軌跡，透過集群分析將王建民的投球數據作分類，在 85 球有紀錄數據的投球分析下，分為快速球（58 球，68.2%）、變速球（7 球，8.2%）、滑球（20 球，23.6%）三類。再經由變異數分析得知在球速、水平、垂直位移分量和變化量等四個變數上都達顯著差異($p < .05$)。事後比較發現水平位移分量上，快速球（平均數-6.25）與變速球（平均數-3.89）的差異不顯著外（平均數差-2.36， $p = .062$ ）。此可說明快速球與變速球的水平分量近似，區別上要由球速或其他變數來觀察。之後依照三球種的四個變數作敘述統計的描述，球速上，快速球介於 89~97 英哩/小時；變速球介於 71~82 英哩/小時；滑球介於 80~87 英哩/小時。水平位移量上，快速球介於-10.90~-1.97 英吋；變速球介於-5.24~-1.72 英吋；滑球介於-6.33~7.69 英吋。垂直位移分量上，快速球介於 2.52~11.39 英吋；變速球介於 6.09~9.16 英吋；滑球介於-3.16~2.56 英吋。變化量上，快速球介於 3.1~6.7 英吋；變速球介於 5.7~8.6 英吋；滑球介於 7.5~10.1 英吋。除了表 4 呈述三球種在四個變項上的極值與平均值外，也透過圖 3

到圖 8 等立體圖與平面圖，使 85 球在四變數上的相關位置由圖形中呈現。

李家昶 (2006) 的研究論文指出，球速是重要的分類依據，軌跡特徵可利用隱藏馬可夫模式 (Hidden Markov Model, HMM) 建立每種球路的時間特性，再由貝氏決定論 (Bayesian Decision Theory) 以及點分配模式 (Point Distribution Models) 驗證其正確性，其研究 195 球的實例中，得到 0.8 以上的球路分辨率。本研究利用 Enhanced Gameday 系統所提供之公用數據，有效利用準確的投球資料，之後再經由統計分析工具，將投球作分類，給予命名，建立嚴謹的量化分析模式。

MLB 耗費鉅資所建構的 EG 系統除了提供即時的比賽訊息之外，由於增加 3D 立體的球路軌跡，其背後的數據對於棒球的客觀分析研究有著巨大的貢獻，以投球球路的飛行軌跡為例，棒球的學術研究很難對於各種球路下一個客觀的定義，現存的球路名詞，諸如直球或曲球等，沒有科學化的數據作依靠，而 EG 系統提供很多變數去形成 3D 立體軌跡，當洋基主場 Yankee Stadium 架設 Pitch f/x 系統後，所有在此出場的投手都會產生投球的數據，可視投手走進實驗室中投球一般，統一影像擷取定位和角度，固定投手投球與進壘的原點，避免投手丘高度的差異與飛行距離不同的問題，盡量排除外在的影響變因，取得純粹且固定的數據。這對於重視數據資料的棒球運動，會有深遠的影響。在未來的研究中，由於 EG 系統提供其他的數據，例如速度與加速度在各向量的變化量，以及在變化點 (break point) 的角度與距離等等，這些數據對於球路飛行軌跡有更深刻的描述，同樣可以做為投手球路分析的變數，期望 EG 系統的數據對於棒球運動能做更多的貢獻。

伍、參考文獻

- 李家昶 (2006)。針對電視轉播的棒球比賽之投手球路辨識，未出版之國立中正大學資訊工程研究所碩士論文，嘉義。
- 陳正昌、程炳林、陳新豐、劉子鍵 (2005)。多變量分析方法-統計軟體應用。臺北：五南圖書出版有限公司。
- 曾慶裕 (1996)。棒球投手訓練的基礎理論與方法之應用，未出版之中國文化大學運動教練研究所碩士論文，臺北。
- 楊賢銘 (2000)。1999 年第 20 屆亞洲盃棒球賽—中、日、韓三國投手投球分析研究，論文發表於 2000 國際大專運動教練科學研討會。臺北：國立臺灣師範大學。
- 蔡文星 (2005)。日本投手松坂大輔配球模式之研究—以 2003 年第二十二屆亞洲棒球錦標賽為例，未出版之國立臺灣體育學院體育研究所碩士論文，臺中。
- 廖文男、林啓川、高英傑、彭國平 (2003)。棒球投手球路與速度之分析—以 2001 年世界盃成棒賽為例，北體學報，11，87-98。
- 鐘宇政 (1999)。棒球投手曹錦輝之配球模式分析，大專體育，46，50-55。

Thomas, J. R., & Nelson, J. K. (2003). 運動研究法 (陳五洲)。臺中：華格納企業 (原著於 2000 年出版)。

Whither. (2007)。以 New Gameday 數據解析王建民的投球特性。漫步在綠海豚街上部落格。資料引自：<http://blog.roodo.com/Whither/archives/3425057.html>。

Chang, P., Han, M., & Gong, Y. (2002). *Extract Highlights from Baseball Game Video with Hidden Markov Models*. Paper presented at the IEEE Int. Conf, Rochester, NY.

Gueziec, A. (2002). Tracking Pitches for Broadcast Television. *IEEE Computer*, 35, 38-43.

MLBlogs-Gameday. (2007). *Enhanced Gameday FAQ*, MLB. com.

From http://gameday.mlblogs.com/gameday/2006/10/enhanced_gameday.html.

MLBlogs-Gameday. (2007, August 30). *Gameday FAQ - April 13, 2007*. MLB. com.

From http://gameday.mlblogs.com/gameday/2007/04/gameday_faq_apr.html.

PITCHf/x Pitch Tracking. (2007). *PITCHf/x*. Sportvision. com.

From <http://www.sportvision.com/>.