

Strike Zone: sistema para la recopilación y análisis de información *pitch by pitch* en el béisbol cubano

Strike Zone: software for the collection and analysis of *pitch by pitch* information in the Cuban baseball



<http://opn.to/a/Baxt2>

Alfredo Ríos-Fuentes ¹*, Beatriz Sánchez-Córdova ¹

¹Centro de Investigaciones del Deporte Cubano, La Habana, Cuba

RESUMEN: El béisbol cubano cuenta en este momento con un sistema de anotación de juegos de béisbol que no cumple con las necesidades y exigencias actuales de este deporte a nivel internacional ya que, entre otras debilidades, no permite recopilar información *pitch by pitch*. Con el objetivo de suplirlo por otro eficaz y apoyar con mayor carácter científico al béisbol cubano, surge la idea de crear *Strike Zone*, un sistema informático que permite guardar detalladamente cada lanzamiento del juego, mostrar reportes en tiempo real y realizar análisis de la información recopilada en cualquier momento (*offline*). Este cuenta con una interfaz de usuario sencilla para una fácil interacción y ha sido desarrollado a través de tecnologías totalmente compatibles con la infraestructura de la Federación Cubana de Béisbol Amateur. Este trabajo demuestra el proceso de desarrollo de *software* realizado para la obtención de *Strike Zone* y su utilidad para el perfeccionamiento y actualización del béisbol cubano.

Palabras clave: análisis, béisbol cubano, información, *pitch by pitch*, sistema informático.

ABSTRACT : At present, Cuban baseball has a baseball game score system. However, it does not meet the current needs and demands of this sport internationally, because among other weaknesses it does not allow to collect *pitch by pitch* information. In order to make up the aforementioned weakness and a more scientific support to the Cuban baseball, we decided to create *Strike Zone*, a computer system that allows you to save in detail each pitch of the game, showing real-time reports and analyzing collected information at any time (*offline*). The system has a simple user interface for easy interaction with it and has been developed through technologies fully compatible with the infrastructure of the Cuban Federation of Amateur Baseball. This work shows the software development procedure conducted for obtaining *Strike Zone* software and its usefulness in improving and updating the Cuban baseball.

Keywords: analysis, Cuban baseball, information, *pitch by pitch*, computer system.

INTRODUCCIÓN

“La pelota en Cuba es una síntesis de talento natural y ganas de brindar espectáculo. No puede decirse que es solo un deporte, es la prolongación cultural de un país, es lo que no perdonaría la gente que no tuviéramos”. De esta manera el reconocido cronista Elio Menéndez expresaba lo que significa el béisbol para los cubanos, por lo tanto, es una gran responsabilidad para el Instituto Nacional de Educación Física y Recreación (INDER) y la Federación Cubana de Béisbol Amateur (FCBA) continuar con los éxitos alcanzados en el pasado que han ubicado a Cuba como una de las principales potencias del béisbol *amateur* en el mundo. La FCBA es la encargada de realizar varios procesos, entre ellos, la anotación y recopilación de información de nuestro pasatiempo nacional.

Desde el año 1977 la FCBA puso en práctica un sistema de anotación de juegos de béisbol denominado “Béisbol”, el cual resulta de gran importancia, pues en él se recopila una serie de datos y estadísticas necesarias para el desarrollo de este deporte. Este sistema de anotación desarrollado en MS-DOS, está provisto de una interfaz gráfica por consola que dificulta la interacción del anotador con el sistema, pero en el momento de su creación constituyó un proyecto innovador de gran ayuda para las tareas de gestión y anotación de los juegos de béisbol, que se realiza en papel y dificulta el trabajo de los anotadores y la gestión de las estadísticas que se generan a partir de la información recogida en cada partido ([Guerrero Ramírez & Stevenson Borrell, 2011](#)).

*Autor para correspondencia: Alfredo Ríos-Fuentes. E-mail: alfredo.rios.fuentes@gmail.com

Recibido: 21/01/2019

Aceptado: 15/02/2019

La anotación de los juegos como se ha explicado es en papel; el anotador debe tener un amplio conocimiento del Sistema Codificado de Anotación Cubano para desempeñar su tarea y paralelamente, insertar en el sistema “Béisbol” cada jugada durante el encuentro, por lo que se realiza el mismo trabajo en dos ocasiones. Este sistema introdujo la posibilidad de guardar estadísticas en formato digital, sin embargo se pierde información del juego pues los lanzamientos no se registran ya que de manera general permite guardar solamente la información de un partido *play by play*.

Con el desarrollo de las tecnologías de la información y las comunicaciones (TIC) y la evolución estadística y científica del béisbol surge la necesidad de crear nuevos sistemas capaces de recoger una mayor cantidad de datos e información ajustados a los nuevos paradigmas. En la actualidad, dominan el mercado sistemas como: *Statcast*, *Rapsodo* y *FlighScope*, los cuales permiten recopilar casi todo lo que ocurre en un terreno de béisbol. Estas herramientas son costosas, necesitan condiciones especiales para su eficiente aplicación, y el análisis de la información está limitado a las opciones propuestas por los desarrolladores del sistema. Esas características obstaculizan la aplicación, en Cuba, de alguno de estos sistemas.

De esta manera surge la idea de crear un software para apoyar el sistema de anotación existente a partir de la recopilación y análisis de información *pitch by pitch*, compatible con la infraestructura tecnológica.

La Federación Cubana de Béisbol Amateur lleva a cabo una serie de procesos vinculados al béisbol. La anotación de estos juegos es uno de los procesos más importantes, base de la recopilación de estadísticas que constituye un aspecto fundamental para este deporte. La aplicación vigente que se utiliza para realizar la anotación no cumple con las tendencias actuales, ya que no recopila información necesaria para el estudio de contrarios.

Por lo tanto, es importante permitir que los juegos se anoten *pitch by pitch*, con lo que se lograría recopilar mayor cantidad de información, lo que no se hace en Cuba, al insertar datos en la aplicación de anotación solo cuando se completa una jugada (*play by play*). Condicionados por

estas circunstancias, el estudio de las herramientas utilizadas en las diferentes ligas de béisbol del mundo y las investigaciones sabermétricas sobre la importancia del *pitch by pitch* en el béisbol moderno, surge la idea de crear *Strike Zone*, un sistema que permita la recopilación y análisis de información *pitch by pitch* en el béisbol cubano.

MATERIALES Y MÉTODOS

Strike Zone fue desarrollado a través de varias herramientas como:

- Microsoft Visual Studio 2013.
- Lenguaje de programación: C Sharp (C#).
- eXtensible Markup Language (XML), para los archivos que exporta e importa el *software*.
- Framework .Net 3.5

Las herramientas utilizadas fueron seleccionadas por las ventajas que ofrecen y en relación con la infraestructura actual de los centros de la FCBA, lo que determinó que para su funcionamiento máximo no se necesitara de gran cantidad de recursos; se puede ejecutar sobre cualquier sistema operativo de la familia *Windows* con el Framework 3.5 instalado como únicos requisitos. Presenta una interfaz de usuario de fácil manejo y adaptada a los conocimientos beisboleros de los anotadores, e incluso puede ser comprensible para cualquier persona con conocimientos básicos de béisbol ya que sus funciones se encuentran reflejadas de manera muy intuitiva.

Este *software* comenzó a desarrollarse en el año 2014 por el Grupo Independiente para la Investigación de Béisbol (GIIB), y en su primera versión probada en la Serie Nacional de Béisbol 54, se comenzaron a realizar adaptaciones para llegar a registrar poco a poco todos los eventos que fueron surgiendo *pitch by pitch* durante un juego de béisbol. Este deporte cuenta con una gran variedad de eventos que pueden ocurrir en cualquier momento, inicialmente este trabajo se enfocó en aquellos que ocurren desde que el lanzador está en el montículo hasta que realiza el lanzamiento.

RESULTADOS

Strike Zone cuenta con una serie de funcionalidades dirigidas a dos objetivos fundamentales: la anotación y el análisis de información *pitch by pitch*.

La novedad de la anotación de este *software* es que permite recoger todos los detalles de lo que ocurre en cada lanzamiento, por ejemplo:

- Identificador del torneo.
- Lanzador (número de camiseta, equipo, mano de lanzar).
- Bateador (número de camiseta, equipo, mano de batear).
- Conteo en que se realiza cada lanzamiento.
- Tipo de lanzamiento.
- Velocidad.
- Resultado (strike cantado, strike tirándole, *foul*, bola, contacto, pelotazo, *foul fly out*, interferencia).
- Descripción (intento de robo, intento de toque de bola, sorprendido en base, *wild pitch*, *passed ball*, *balk*).
- Localización del lanzamiento vista desde atrás de home.
- Otras acciones que no constituyen un lanzamiento (Bases por bolas intencionales, out sin lanzar, strike/bola por regla, *balk* sin lanzar, viraje).
- En la parte derecha de la pantalla utilizada para la anotación del juego se van mostrando los datos de los lanzamientos en tiempo real. Cada vez que se realiza un lanzamiento el analizador se actualiza automáticamente.

Estas estadísticas se pueden enlazar a través de los diferentes filtros que ofrece el programa sin afectar la anotación del juego. Se pueden filtrar los lanzamientos por el tipo de lanzamiento,

conteos (primer lanzamiento, lanzador encima, lanzador debajo, dos strikes, tres y dos, parejo), *swings* (fallados y contactos), bateador en contra, o por comparencias al bate (respecto al lanzador).

Para el estudio de contrarios genera reportes de bateadores y lanzadores con sus tendencias, debilidades y fortalezas.

Strike Zone ha sido aplicado de manera sistemática por el Grupo para la Investigación del Béisbol (GIIB) desde la Serie Nacional de Béisbol 54, y en los eventos internacionales en los que participó la selección nacional de Cuba (Series del Caribe 2015-2018, Juegos Panamericanos 2015, PREMIER 12, Can-Am 2016-2017) (ver [Tabla 1](#)).

ANÁLISIS Y DISCUSIÓN

A partir de 69216 lanzamientos recopilados entre 2015 y 2018 por el sistema *Strike Zone*, se ha podido determinar que la velocidad promedio de la recta en Series Nacionales de Béisbol es de 85 mph. Esta velocidad se encuentra por debajo del promedio de los eventos internacionales en los que participa la selección nacional de Cuba, donde se lanzan rectas que sobrepasan las 88 mph según los datos recopilados por este mismo sistema, mientras que en las Grandes Ligas de Béisbol (MLB) fueron de 93 mph en 2017, con tendencia ascendente durante los últimos 5 años.

A partir de los datos recopilados se determinaron los índices de correlación de Pearson entre las variables lanzamiento a lanzamiento. En el caso del bateo se encontró la correlación más fuerte ($R=0.76$ con $p<0,05$) entre el porcentaje de ponches del total de comparencias al bate (%K) y el porcentaje de swings abanicados del total de lanzamientos (%WHIFF). Por lo que se puede afirmar que en Cuba los equipos que más se ponchan son los que más swings fallan del total de lanzamientos.

Tabla 1. Lanzamientos recopilados en *Strike Zone* (2015-2018).

Eventos	TL	%B	%S	Strikes		Swings		
				Cantados	Swings	Tirándole	Foul	En juego
Nacionales	21068	39%	61%	28%	72%	15%	35%	50%
Internacionales	48148	36%	64%	28%	72%	25%	33%	42%

Fuente: Elaboración propia.

TL: Total de Lanzamientos, %B: Porcentaje de Bolas del TL, %S: Porcentaje de Strikes del TL.

Sin embargo, en otra ligas de mayor nivel como la MLB, según (Murphy, 2012) los equipos que más se ponchan son los que más swing fallan del total de swings realizados. Evidentemente, las características de cada liga influyen en los resultados de este tipo de investigaciones.

Otro fenómeno demostrado fue que a mayor cantidad de swings en una misma comparecencia al bate (CB) menor efectividad ofensiva reflejada a través de métricas como el promedio de bateo (AVE), promedio de embasado (OBP), el *slugging* (SLG), el *on base plus slugging* (OPS), y la frecuencia de jonrones por comparecencia (CB/HR) (ver [Tabla 2](#)).

Otras investigaciones como (Ring, 2018) han demostrado este mismo comportamiento pero en la MLB, exponiendo las experiencias de bateadores que han comenzado a intentar disminuir la cantidad de swings en búsqueda de mayor rendimiento ofensivo.

El desconocimiento de este fenómeno por parte de entrenadores y atletas no permite la eficiente corrección de errores. Por ejemplo, la Selección Nacional de Béisbol de Cuba durante su participación en las Series del Caribe de 2016 hasta 2019, se ha mantenido entre los 2 equipos que más swings hacen del total lanzamientos (Ver [Tabla 3](#)). Esta tendencia, como se demostró anteriormente, atenta contra la obtención de

Tabla 2. Relación entre la cantidad de swings que realiza un bateador en una comparecencia al bate con indicadores ofensivos (17264 comparecencias al bate).

Situación	AVE	OBP	SLU	OPS	CB/HR
Swings == 1	0.306	0.344	0.459	0.803	39.4
Liga	0.261	0.334	0.382	0.716	55.3
Swings == 2	0.232	0.267	0.330	0.597	67.3
Swings >= 3	0.210	0.250	0.296	0.546	92.0

Fuente: Elaboración propia

(AVE: promedio de bateo, OBP: porcentaje de embasado, SLU: *slugging*, OPS: OBP+SLU, CB/HR: frecuencia de jonrones).

Tabla 3. Comportamiento de la cantidad de swings y estadísticas de rendimiento ofensivo en Series del Caribe de Béisbol (2016-2019).

TORNEO	EQ	CB	AVG	OBP	SLU	OPS	SWING%_(TL)
SCB 2016	PUR	210	0.244	0.305	0.342	0.647	0.51
SCB 2016	VEN	131	0.282	0.359	0.436	0.795	0.48
SCB 2016	DOM	150	0.216	0.273	0.273	0.546	0.49
SCB 2016	CUB	198	0.200	0.273	0.239	0.512	0.49
SCB 2016	MEX	156	0.286	0.359	0.400	0.759	0.46
SCB 2017	VEN	193	0.261	0.311	0.378	0.689	0.50
SCB 2017	MEX	175	0.252	0.291	0.350	0.641	0.49
SCB 2017	PUR	193	0.240	0.319	0.351	0.670	0.48
SCB 2017	CUB	175	0.242	0.276	0.376	0.652	0.51
SCB 2017	DOM	146	0.203	0.292	0.227	0.519	0.43
SCB 2018	PUR	128	0.356	0.398	0.568	0.966	0.47
SCB 2018	MEX	114	0.279	0.330	0.346	0.676	0.51
SCB 2018	DOM	81	0.297	0.346	0.365	0.711	0.47
SCB 2018	VEN	174	0.370	0.439	0.623	1.062	0.46
SCB 2018	CUB	75	0.306	0.356	0.403	0.759	0.49
SCB 2019	PAN	30	0.207	0.233	0.276	0.509	0.56
SCB 2019	CUB	180	0.189	0.249	0.238	0.487	0.50
SCB 2019	VEN	99	0.186	0.265	0.233	0.498	0.49
SCB 2019	MEX	109	0.208	0.243	0.228	0.471	0.47
Todos	LIGA	2717	0.253	0.313	0.353	0.666	0.49

Fuente: Elaboración propia

EQ: Equipo, CB: Comparecencia al bate, AVG: Promedio de bateo. OBP: porcentaje de embasado, SLU: *slugging*, OPS: OBP+SLU, SWING%_(TL): porcentaje de swings del total de lanzamientos.

mejores resultados ofensivos, aunque estos están influidos por otros factores ajenos al objetivo de este artículo.

El análisis de los datos recopilados por el software *Strike Zone* demuestra que los resultados registrados jugada a jugada pueden estar influidos por lo que ocurre lanzamiento a lanzamiento, por lo que se manifiesta la importancia del análisis de información lanzamiento a lanzamiento para la eficiente evaluación del rendimiento deportivo en el béisbol cubano.

CONCLUSIONES

Con el desarrollo de *Strike Zone*, se logra recopilar y analizar información *pitch by pitch* en el béisbol cubano, de modo que se convierte en la única herramienta en Cuba capaz de llevar a cabo estas funcionalidades.

Ahora, el sistema de anotación utilizado en el país cuenta con el sustento de esta herramienta que, entre otras funciones, ayuda a la toma de decisiones, apoya presentes y futuros estudios saberométricos y permite realizar *scouting* exhaustivo ajustado a las tendencias actuales, así como detectar debilidades, fortalezas y tendencias de lanzadores y bateadores.

REFERENCIAS BIBLIOGRÁFICAS

Guerrero Ramírez, D.E., & Stevenson Borrell, R.E. (2011). Sistema para la anotación de juegos de béisbol. Facultad Regional de Artemisa de la Universidad de las Ciencias Informáticas.

Murphy, B. (2012). WHIFF! Strikeout Rates Explained. Beyond the Box Score.

Ring, S. (2018). Javier Baez is doubling down. Fangraphs.

Albert, J. (2010). Baseball Data at Season, Play-by-Play, and Pitch-by-Pitch Levels. Journal of Statistics Education. doi: 10.1080/10691898.2010.11889592

Albert, J. (2010). Sabermetrics: The Past, the Present, and the Future. doi: 10.5948/UPO9781614442004.002

Armijo, T., Gao, X., Lovette, B., & Siemers, K. (2015). Analytic Versus Non-Analytic Decision Makers. Universidad de San Francisco.

Beneventano, P., Berger, P. D., & Weinberg, B. D. (2012). Predicting Run Production and Run Prevention in Baseball: The Impact of Sabermetrics. International Journal of Business, Humanities and Technology, 2(4).

Davenport, T. H. (2011). Competir mediante el análisis. *Harvard Business Review*. Disponible en hbrl.com.

Dewan, J. (2017). 2017 Shift Slowdown. Bill James Online.

Fry, M. J., & Ohlmann, J. W. (2012). Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications. *Interfaces*, 42(2), 105-108.

González Castellanos, M., & Soto Valero, C. (2015). Sabermetría y nuevas tendencias en el análisis estadístico del juego de béisbol. Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación.

Harrison, W. K., & Salmon, J. L. (2017). Bullpen Strategies for Major League Baseball. MIT SLOAN SPORTS ANALYTICS CONFERENCE.

Lahman, S. (2010). Sean Lahman's Baseball Archive. from <http://www.baseball1.com>

Lewis, M. (2003). Moneyball: The art of Winning an Unfair Game.

Murphy, B. (2012). WHIFF! Strikeout Rates Explained. Beyond the Box Score.

Pérez Martínez, I. (2008). Metodología para la evaluación del rendimiento competitivo en el béisbol (ERC-Béisbol). Instituto Superior de Cultura Física Manuel Fajardo. Facultad de Cultura Física de Matanzas.

Ring, S. (2018). Javier Baez is doubling down. Fangraphs.

Soto-Valero, C., González-Castellanos, M., & Pérez-Morales, I. (2017). A predictive model for analysing the starting pitchers' performance using time series classification methods. *International Journal of Performance Analysis in Sport*. doi: 10.1080/24748668.2017.1354544

Tango, T. M., Litchman, M. G., & Dolphin, A. E. (2006). *The Book. Playing the Percentages in Baseball*.

Thorn, J., Palmer, P., & Reuther, D. (1985). *The hidden game of baseball* (p. 4). New York.

Ing. Alfredo Ríos-Fuentes, Ingeniero en Ciencias Informáticas graduado en la Universidad de las Ciencias Informáticas (UCI, 2015). Titulado del Curso de “Métodos Estadísticos en la Investigación Científica” recibido en la Universidad de las Ciencias Informáticas (UCI, 2016). Ha impartido conferencias en el Curso de Directores de Equipos de la Serie Nacional de Béisbol (55, 57, 58). Tema impartido: “La sabermetría como herramienta para la toma de decisiones en el béisbol”. Participó como ponente en los siguientes eventos: VI Convención Internacional de Actividad Física y Deportes (AFIDE), VI Forum Científico de la Escuela Nacional de Béisbol, VII Forum Científico de la Escuela Nacional de Béisbol, VI Convención Internacional de Actividad Física y Deportes (AFIDE), CEADES 2018, Cubamotricidad 2018. Participó como desarrollador de software y analista del equipo nacional de béisbol en los siguientes eventos: Serie del Caribe Puerto Rico 2015, Premier 12, Juegos Panamericanos de Toronto, Serie del Caribe República Dominicana 2016, Tope Cuba contra el Tampa Bay Rays, Serie del Caribe México 2017, Clásico Mundial de Béisbol 2017, Torneo Panamericano Sub 23 de Béisbol 2017, Serie del Caribe México 2018, Juegos Centroamericanos de Barranquilla 2018. E-mail: alfredo.rios.fuentes@gmail.com

Dra. C. Beatriz Sánchez-Córdova. E-mail: bsc23863.bs@gmail.com

Los autores de este trabajo declaran no presentar conflicto de intereses.

Este artículo se encuentra bajo licencia [Creative Commons Reconocimiento-NoComercial 4.0 Internacional \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)