ISSN: 2306-2495 | RNPS: 2343_____________________________________*Serie Científica de la Universidad de las Ciencias Informáticas*
http://publicaciones.uci.cu Vol. 14, No. 9, Mes: Septiembre, 2021, Pág. 1-15

Tipo de artículo: Artículo original

Sistema para la generación de calendarios para la Serie Nacional de Béisbol

Schedules generation system for the National Baseball Series

Mario Herrera Almira [1*], https://orcid.org/0000-0001-8698-2014
Rafael Alejandro Gayoso Rodríguez [2], https://orcid.org/0000-0001-9155-1198
Alejandro Rosete Suárez [3], https://orcid.org/0000-0002-4579-3556
Luis Suárez González [4], https://orcid.org/0000-0001-7194-6117

[1] Instituto Superior Politécnico José Antonio Echeverría. Facultad de Ingeniería Informática. La Habana. Cuba.
[2] Instituto Superior Politécnico José Antonio Echeverría. Facultad de Ingeniería Informática. La Habana. Cuba.
[3] Instituto Superior Politécnico José Antonio Echeverría. Facultad de Ingeniería Informática. La Habana. Cuba.
[4] Instituto Superior Politécnico José Antonio Echeverría. Facultad de Ingeniería Informática. La Habana. Cuba.

*** Autor para correspondencia: lsuarezgo@ceis.cujae.edu.cu**

Resumen

La Serie Nacional de Béisbol de Cuba es un evento deportivo que consume una gran cantidad de combustible. Disminuir el costo de transportación es uno de los principales objetivos del país debido a las largas distancias a recorrer en el transporte de los equipos. Este trabajo propone un método computacional para la generación de calendarios para la Serie Nacional de Béisbol, el cual argumenta la necesidad de crear un sistema informático para realizar esta tarea, y las ventajas que ofrece con respecto a la actual forma de crear los calendarios. Se plantea el problema de optimización que corresponde para luego solucionarlo utilizando metaheurísticas mediante la biblioteca de clases BiCIAM. En este trabajo se implementan dos nuevas heurísticas para generar soluciones iniciales del problema, una basada en enfrentamientos y otra en fechas, siendo la primera la que ofrece mejores resultados. Se realizan experimentos empleando las metaheurísticas Escalador de Colinas, Estrategias Evolutivas y Búsqueda Aleatoria, el Escalador de Colinas obtiene los mejores resultados. El sistema ahorra hasta el 16.27% de la cantidad de kilómetros a recorrer, con respecto a los calendarios oficiales diseñados por el personal del INDER.

Palabras clave: calendario deportivo; metaheurísticas; mutaciones; problemas de optimización; problema de transportación en torneos.

Abstract

Cuba's National Baseball Series is a sporting event that consumes a large amount of fuel. Reducing transportation costs is one of the country's main objectives due to the long distances involved in transporting the teams. This work proposes a computational model calendar generation system for the National Baseball Series, which argues the need to create a computer system to perform this task, and the advantages it offers over the current way of creating calendars. The corresponding optimization problem is posed and then solved using metaheuristics through the BiCIAM class library. To complement the metaheuristics used in this work, two new heuristics are implemented in order tocreate the initial solutions for theoptimization process, one is matches based and the other is dates based, out of the two of them the first one offers better results. Throughout this work the metaheuristics Hill Climbing, Evolutionary Strategies, and Random Search are used to perform the experiments, and the results *show that the Hill Climbing metaheuristic obtains the best results. The system saves up to 16.27% of the number of kilometers to be covered, with respect to the official calendars designed by INDER staff.*

Keywords: *metaheuristics; mutations; optimization problem; sport schedule; traveling tournament problem.*

Recibido: 01/07/2021
Aceptado: 01/08/2021

Introducción

En los últimos años se han llevado a cabo distintas investigaciones que tienen como objetivo la reducción de la cantidad de combustible que se emplea en diversos en sectores del panorama mundial (Lambers et al., 2021; Pérez et al., 2021). El deporte constituye uno de los sectores donde se evidencia la necesidad de reducir el costo de transportación. La distancia total que deben recorrer todos los equipos participantes durante la celebración del torneo influye considerablemente en el gasto de combustible debido a la distancia que se debe cubrir entre sede y sede. Esta situación constituye un problema de optimización el cual se encuentra definido como el Problema de Transportación en Torneo (Travelling Tournament Problem, TTP por sus siglas en inglés) (Rameez Naqvi et al., 2021), que se encuentra incluido dentro de los problemas de calendarios deportivos.

El TTP tiene una complejidad elevada debido a la cantidad de variantes posibles de calendarios. Constituye un problema combinatorio perteneciente a la clase NP-Hard, por este motivo se ha enfrentado con anterioridad utilizando métodos aproximados como heurísticas y metaheurísticas (Alarcón et al., 2017; Rosete et al., 2013). Las heurísticas y metaheurísticas existentes que dan solución a problemas similares de ruteo (Khelifa et al., 2017; Rameez Naqvi et al., 2021; Rosete et al., 2013), se pueden utilizar como ideas para crear nuevas heurísticas que construyan mejores soluciones iniciales o para hacer nuevos operadores de mutación que se puedan aplicar en este tipo de problemas. Las metaheurísticas constituyen algoritmos de optimización de propósito general. Tienen como objetivo brindar una solución aproximada con un costo computacional razonable(Talbi, 2009).

En el ámbito nacional, particularmente en la Serie Nacional de Béisbol (SNB), los calendarios deportivos para la competición se desarrollan de manera manual. Este proceso consume una gran cantidad de tiempo y solo permite explorar un número pequeño de soluciones ya que el espacio de búsqueda es muy abarcador. La situación descrita anteriormente demuestra que hay un problema, y su solución genera un gran interés y posibilitaría la exploración de una mayor cantidad de soluciones en busca de la solución óptima, o al menos, una solución cercana a la óptima.

En (Khelifa et al., 2017), (Pace, 2021), (Kynga¨s et al., 2017) y (Fazayeli et al., 2018) se plantean distintas soluciones precedentes para el TTP, pero ninguno se adecúa a la problemática cubana ya que las características de los problemas planteados, no coinciden con el ámbito nacional. En(Rosete et al., 2013), se plantea una solución que sirve para resolver el problema existente en Cuba pero es difícil de usar y configurar, ya que no cuenta con una interfaz gráfica para la interacción con el usuario y no permite obtener soluciones con distintas metaheurísticas.

Para resolver la problemática planteada se logró definir como objetivo general, desarrollar un método computacional basado en metaheurísticas paragenerar los calendarios de la Serie Nacional de Béisbol en Cuba.

El presente trabajo está estructurado de la siguiente forma:

En la sección Materiales y métodos se explica el diseño de la investigación, como se realizó la misma y como se lleva a la práctica. En la sección Resultados y discusión se muestran los resultados obtenidos después de aplicar las técnicas seleccionadas y demuestra que la propuesta planteada en esta investigación, logra cumplir con el objetivo general de este trabajo, resuelve la problemática planteada anteriormente y se adecúa a los requerimientos existentes en la Serie Nacional de Béisbol de Cuba.

Materiales y métodos

Dentro de la estructura de la Serie Nacional de Béisbol (SNB), se definen un conjunto de enfrentamientos entre dos equipos en el estadio de uno de ellos. Estos enfrentamientos se organizan dentro de una serie de fechas determinadas. Una fechase encuentra formada por un conjunto de enfrentamientos que se llevarán a cabo el mismo día. Por tanto, se puede definir un calendario como el conjunto de fechas donde se llevarán a cabo todos los enfrentamientos del torneo.

En esta propuesta se presentan tres tipos de restricciones fundamentales:

- Restricciones que se deben cumplir en todo momento:
    - Cada equipo solo puede aparecer una vez en cada fecha. El incumplimiento de esta restricción significaría que un equipo se encuentra jugando en dos lugares diferentes al mismo tiempo lo cual no tiene sentido.
    - Si el calendario es a simple vuelta cada enfrentamiento solo puede aparecer una sola vez, pero si el calendario es a doble vuelta cada enfrentamiento debe aparecer dos veces intercambiando al equipo visitador por el equipo sede en la segunda ocasión.
- Restricciones que se penalizan añadiendo kilómetros extras al calendario cada vez que se incumplen:
    - Un equipo no puede jugar en su sede una cantidad de veces seguidas determinada.
    - Un equipo no puede jugar fuera de su sede una cantidad de veces seguidas determinada.
    - Los equipos seleccionados como campeón y subcampeón de la serie pasada deben enfrentarse en la primera fecha de haberse seleccionado esta opción.

- Un equipo no debe recorrer más de cierta distancia en un solo viaje. Esta restricción existe para evitar viajes muy largos que agoten a los jugadores.

- Restricciones adicionales:
  - Se incluye descanso o no. Un descanso en la terminología de los calendarios deportivos significa que todos los equipos deben regresar a sus sedes en una fecha determinada, y en esta fecha no se realizan enfrentamientos. Esta interrupción en el calendario provoca un aumento considerable en la cantidad de kilómetros totales.

Para garantizar el cumplimiento del objetivo general planteado anteriormente, es necesario definir la forma de representar un calendario. Cada calendario C se representa como un conjunto de n dimensiones correspondientes a las fechas del calendario. La $i-ésima$ fecha F_i es un vector con una cantidad m de enfrentamientos. El $j-ésimo$ enfrentamiento de una fecha F_i, representado por E_{ij}, constituye un par ordenado donde se enfrentan los equipos L y V, siendo L el equipo local y V el equipo visitante. Para representar las distancias entre las sedes de los equipos que participan en el torneo generalmente se utiliza una matriz de distancias(Rosete et al., 2013). La matriz de distancias, representada por M, es una matriz cuadrada y simétrica donde hay una fila y una columna por cada equipo. En la intercepción,($M_{ij}: i! = j$) se encuentra la distancia en kilómetros entre las sedes de los dos equipos (Rosete et al., 2013). Formalmente se puede plantear:

$$C = \{F_1, F_2, \dots, F_i, \dots, F_n\} \qquad F_i = \{E_{i1}, E_{i2}, \dots, E_{ij}, \dots, E_{im}\} \qquad E_{ij} = (L_{ij}, V_{ij}) \qquad L_{ij}, V_{ij} \in Equipos\ participantes$$

Función objetivo

La función objetivo está dada por la suma de la cantidad de kilómetros que recorre cada uno de los equipos durante el torneo. En este problema se intenta minimizar esta cantidad de kilómetros lo cual se traduce en una disminución en el combustible empleado para transportar los equipos entre las sedes a lo largo del torneo. Las fórmulas siguientes fueron tomadas del artículo (Rosete et al., 2013), además se agregaron dos fórmulas nuevas (ecuaciones (6) y (7)) para incluir restricciones que se incorporan en esta propuesta, en estas fórmulas se pueden observar los componentes que conforman la función objetivo:

$$KMT = \sum KM_e\ , \ \ e \in \text{Equipos} \qquad (1)$$

En la ecuación (1) se define que la cantidad de kilómetros totales (KMT) está dada por la suma de los kilómetros recorridos por cada uno de los equipos (KM_e). En las ecuaciones (2), (3) y (4) se puede observar que la cantidad de kilómetros recorridos por cada equipo está dada por la suma de las distancias recorridas entre las sedes consecutivas del itinerario del equipo analizado:

$$KM_e = \sum_{i=0}^{m-1} distancia\big(sede(e,i), sede(e,i+1)\big)$$
$$+distancia\big(ciudad(e), sede(e,1)\big) \quad (2)$$
$$+distancia\big(sede(e,m), ciudad(e)\big).$$

$$ciudad(e) = ciudadqueessedeoficialdelequipoe \quad (3)$$

$$sede(e,i) = \begin{cases} ciudad(q) si (q,e) \in F_i \\ ciudad(e) si (e,q) \in F_i \\ ciudad(e) si (e,q) \notin F_i, (q,e) \notin F_i \end{cases} \quad (4)$$

En el caso de la función $sede(e,i)$ se devuelve la sede del enfrentamiento que corresponde al equipo e en la fecha i. Es necesario notar que en las ecuaciones se tiene en cuenta que todos los equipos parten inicialmente desde sus correspondientes sedes, y se trasladan hacia la sede determinada en la primera fecha del calendario. Lo mismo sucede al final del calendario cuando todos los equipos deben regresar a sus sedes cuando hayan terminado de realizar todos los enfrentamientos que les correspondan. En muchas ocasiones los calendarios poseen descansos, esto significa un período de tiempo en que todos los equipos regresan a sus sedes y no realizan ningún enfrentamiento y luego de terminado este período retoman el torneo por donde lo dejaron. Estos descansos se tienen en cuenta en la última condición de la función sede. La función $distancia(A, B)$devuelve la distancia entre las dos sedes que recibe por parámetro, esta distancia es una de las entradas del algoritmo.

Lo último a tener en cuenta para la evaluación de la función objetivo son las violaciones de las restricciones. Como la restricción de si el campeón y el subcampeón de la serie pasada se enfrentan en la primera fecha ($IECS$) no necesita un cálculo, sino solo una comprobación, el valor de la variable $IECS$ es 1 si se incumple la restricción y 0 si no se incumple. Las ecuaciones (5), (6) y (7) corresponden al cálculo de la cantidad de veces que se violan el resto de las restricciones a lo largo del calendario.

La ecuación (5) corresponde al incumplimiento de la restricción de encuentros seguidos como visitador (IEV):

$$IEV = \sum IEV_e \text{ , } e \in Equipos \quad (5)$$

IEV_ecuenta las veces que en el itinerario un equipo e juega más de V juegos de visitador de manera consecutiva.

La ecuación (6) corresponde al incumplimiento de la restricción de encuentros seguidos como sede (IES):

$$IES = \sum IES_e \text{ , } e \in Equipos \quad (6)$$

IES_ecuenta las veces que en el itinerario un equipo e juega más de S juegos como sede de manera consecutiva.

La ecuación (7) corresponde al incumplimiento de la restricción no se debe recorrer más de cierta distancia en un solo viaje:

$$IED = \sum IED_e , e \in Equipos \tag{7}$$

IED_e cuenta las veces que en el itinerario un equipo $e$recorre más de una cierta distancia Den su itinerario.

Finalmente, para calcular el valor de la Función Objetivo (FO) se suma la cantidad de veces que se incumple cada una de las restricciones planteadas y ese valor se multiplica por el término P que representa un valor para penalizar los calendarios por incumplir restricciones. En este trabajo se le asigna un valor a $P = 100\,000$ para que, al incumplir mayor cantidad de restricciones, se obtenga un mayor valor de la Función Objetivo y así garantizar que los calendarios sean mejores mientras menos restricciones incumplan. Cuando dos calendarios estén empatados en la cantidad de restricciones violadas, entonces tienen preferencia los de menor cantidad de kilómetros recorridos por los equipos (KMT). Por último, se suma la cantidad de kilómetros del calendario KMT para obtener el valor de la Función Objetivo (Ver Ecuación (8)).

$$FO = KMT + P * (IEV + IES + IED + IECS) \tag{8}$$

Modelado e Implementación

Para el modelado e implementación de este problema, se hace uso de BiCIAM (Chaveco et al., 2021): un marco de trabajo que contiene un modelo unificado de algoritmos metaheurísticos.

Construcción de soluciones

Para la construcción de un calendario inicial correcto y factible, primero deben ubicarse todos los enfrentamientos posibles en cada una de las fechas disponibles siempre y cuando se cumpla con la restricción de que un equipo no puede aparecer en dos enfrentamientos dentro de la misma fecha. Para esto se implementaron dos algoritmos para la construcción de una solución inicial: uno basado en seleccionar en cada momento el enfrentamiento que menos distancia añada al calendario y otro que se encarga de ordenar las fechas ya conformadas en cada una de las posiciones posibles en el calendario seleccionando siempre la que menos distancia en kilómetros añada a la solución. A continuación, se presentan los seudocódigos pertenecientes a estos algoritmos.

```
Entrada:lista_enfrentamientos, n cantidad de equipos
Salida: Calendario calendario_inicial
Elegir enfrentamiento (e) aleatorio ∈ lista_enfrentamientos
Crear fecha(f )
Añadir e  a f
i = 0
Mientras f  no esté completa hacer
                              e_x ← lista_enfrentamiento[i]
  sie_x cumple restricción entonces
    Añadire_xa f
  fin si
  i=i+1
fin mientras
Añadir f   a calendario_inicial
Mientras calendario_inicialno esté completo hacer
  f  =fecha_vacía
  Mientras f  no esté completa hacer
    Para j= 0 hasta n/2
      lista_posibles_enfrentamientos =  posiblesEnfrentamientos(f)
      Ordenarlista_posibles_enfrentamientos en cuanto a menor distancia
                              e_x ← lista_posibles_enfrentamientos[0]
      sie_xcumple restricción entonces
        Añadire_xaf
      fin si
    fin para
  fin mientras
  Añadir f   a calendario_inicial
  fin mientras
  Devolver inicial
```

Algoritmo 1: Seudocódigo de la Heurística de enfrentamientos.

```
Entrada: lista_fechas
Salida: Calendario calendario_inicial
“La lista_fechas que se recibe como parámetro, se obtiene como resultado del
algoritmo de construir la solución inicial a partir de los enfrentamientos”
Extraer fecha(f ) aleatoria ∈lista_fechas
Añadir f  a calendario_inicial
n ← tamaño de lista_fechas
Para i= 0hastan
  Mientras calendario_inicialno esté completo hacer
    pos_{f_x}← pos FechaMenosDistancia(calendario_inicial, lista_fechas)
    f_x ← lista_fechas[pos_{f_x}]
    Añadirf_xa calendario_inicial(pos_{f_x})
  fin mientras
fin para
Devolver inicial
```

Algoritmo 2: Seudocódigo de la Heurística de Fechas

Mutación de soluciones

En este trabajo se utilizaron ocho mutaciones diferentes para transformar la estructura de los calendarios y así generar variedad en las soluciones y poder realizar el proceso de búsqueda de los mejores resultados. Estas mutaciones se encargan de realizar algún cambio, ya sea en el orden de las fechas o en los enfrentamientos dentro de las fechas. Al realizar alguno de estos cambios se tiene en cuenta que se respeten las dos restricciones fundamentales mencionadas anteriormente. Si al realizar algún cambio se llega a violar una de estas dos restricciones, se corrige el calendario para mantenerlo factible y que se mantenga cumpliendo con las dos restricciones fundamentales. A continuación, se describen las mutaciones utilizadas. Para referirse a cada una de ellas se utiliza la notación $\boldsymbol{M_e}$, utilizando $\boldsymbol{M}$ para referirse a mutación, y el subíndice e , para indicar el número de mutación. $\boldsymbol{M_1}$ se encarga de seleccionar una fecha de entre todas las del calendario y cambiarla de posición.$\boldsymbol{M_2}$selecciona dos fechas del calendario e intercambia sus posiciones.$\boldsymbol{M_3}$escoge dos fechas del calendario e intercambia sus posiciones.$\boldsymbol{M_4}$ toma un enfrentamiento de una fecha y lo coloca en otra fecha diferente, reajustando ambas fechas de ser necesario, para que no existan soluciones no factibles. En ocasiones, se comporta de manera similar a $\boldsymbol{M_2}$. $\boldsymbol{M_5}$selecciona dos equipos e intercambia el uno por el otro en todas las ocasiones que aparezcan en el calendario. $\boldsymbol{M_6}$selecciona una fecha y cambia en cada enfrentamiento de la fecha al equipo que se encuentra de visitador por el equipo que se encuentra de sede y viceversa.$\boldsymbol{M_7}$selecciona un enfrentamiento determinado, y solo en ese enfrentamiento intercambia al equipo que juega de visitador por el equipo que juega de sede. Por último$\boldsymbol{M_8}$selecciona un equipo determinado y en todos los enfrentamientos en los que aparezca este equipo lo intercambia de visitador a sede o de sede a visitador.

A partir de lo descrito anteriormente, se puede observar que existen mutaciones que modifican las sedes de enfrentamientos como son $\boldsymbol{M_5}$, $\boldsymbol{M_6}$, $\boldsymbol{M_7}$y $\boldsymbol{M_8}$. Existen SNB que poseen calendarios simples por lo que debe respetarse las sedes prefijadas con anterioridad y solo están disponibles las mutaciones $\boldsymbol{M_1}$, $\boldsymbol{M_2}$, $\boldsymbol{M_3}$y $\boldsymbol{M_4}$. En las otras variantes de calendarios de la SNB, todas las mutaciones están disponibles. Cabe destacar que, en cada paso para aplicar una mutación, esta se escoge de maneara aleatoria.

Caso de Estudio

Para estudiar el comportamiento de las metaheurísticas en el problema, se estudiaron veinte casos de calendarios de la SNB, los cuales poseen distintas características (simple o doble vuelta, juego inaugural, campeón contra subcampeón, etc.), siendo 11 calendarios oficiales utilizados desde la serie 52 hasta la serie 60 y el resto resultando una variación de algunos de estos calendarios mencionados anteriormente. Ejemplos de estas variaciones pueden ser: añadir un juego inaugural, añadir un juego entre el campeón y subcampeón de la serie anterior, entre otros. En los casos de la serie 53 hasta la 60, los calendarios fueron creados manualmente, por expertos de la Comisión Nacional de Béisbol. En el caso de la serie 52, este calendario fue generado de manera semiautomática: primero se generó una versión preliminar utilizando el software del artículo(Rosete et al., 2013) y luego se le aplicaron modificaciones por parte de los expertos manualmente. En el caso de la serie 60-v17, corresponde a una versión totalmente generada por el software presentado en(Rosete et al., 2013).Para consultar los datos de los calendarios utilizados, se puede acceder mediantehttps://drive.google.com/drive/folders/1raNHuRW9Pd3gxoUMHOG2eWGvRHRt9QHY?usp=sharing.

Metaheurísticas a comparar

Existen muchas metaheurísticas que pueden ser utilizadas para solucionar el TTP, pero es muy difícil extrapolar las soluciones existentes a este problema para ser utilizadas en otros casos diferentes. Esto se debe a que cada uno tiene en cuenta diseños de calendarios y restricciones particulares a satisfacer (Rosete et al., 2013). En este estudio las metaheurísticas utilizadas son Escalador de Colinas(Talbi, 2009), Estrategia Evolutiva (Talbi, 2009; Wei et al., 2019) y Búsqueda Aleatoria (Talbi, 2009). Estas metaheurísticas pueden ser utilizadas en diferentes tipos de problemas, pero necesariamente necesitan la definición de heurísticas subordinadas para obtener las soluciones iniciales o para obtener las siguientes soluciones candidatas. Por ejemplo, el Escalador de Colinas constituye una metaheurística, que continuamente se mueve en la dirección de los valores que sean mejores que el anterior. En este caso se definen heurísticas de obtención de soluciones iniciales y heurísticas de mutación que ejercen cambios diferentes sobre los calendarios y permiten obtener las siguientes soluciones candidatas. Estas heurísticas son específicas del problema en cuestión, a diferencia del algoritmo Escalador de Colinas que puede ser utilizado en problemas diferentes solo cambiando el método de obtener la solución inicial y de obtener las siguientes soluciones candidatas. En este trabajo se utiliza como método de construcción de la solución inicial, el algoritmo que ofrece mejor desempeño de los descritos en la sección anterior, y se utilizan las ocho heurísticas de mutación descritas en la sección Mutación de soluciones.

Para la comparación de los resultados obtenidos, se definen 10 mil iteraciones para cada una de las metaheurísticas. Dado el carácter no determinista de las mismas, se realizan 20 ejecuciones del algoritmo para obtener la evaluación media de la función objetivo.

Criterios para analizar y comparar la calidad de resultados

Para determinar cuál metaheurística ofrece un mejor desempeño se aplica una prueba de Friedman1xN para determinar si existen diferencias significativas en el rendimiento de las heurísticas. Como hipótesis nula definida se plantea que no existen diferencias significativas entre las distintas heurísticas utilizadas entre las evaluaciones de la función objetivo, mientras que la hipótesis alternativa plantea que existen diferencias significativas. Para corroborar los resultados obtenidos por la prueba de Friedman, se lleva a cabo un análisis post-hoc, para determinar diferencias significativas entre las restantes heurísticas. Para ajustar los distintos p-valores se llevan a cabo los procedimientos de ajustes Holm , Finner y Li .Para el procesamiento de los datos y la realización de las pruebas estadísticas se utiliza la herramienta KEEL (Singh et al., 2021).

Resultados y discusión

En Figura 1 se muestran las evaluaciones de la función objetivo obtenida por cada una de las metaheurísticas para las instancias analizadas, utilizando como método de obtención de la solución inicial la Heurística de Enfrentamientos, ya que este constituye el algoritmo de construcción de la solución inicial que mejores resultados arrojó derivado de la prueba de Friedman y el análisis post-hoc correspondiente.

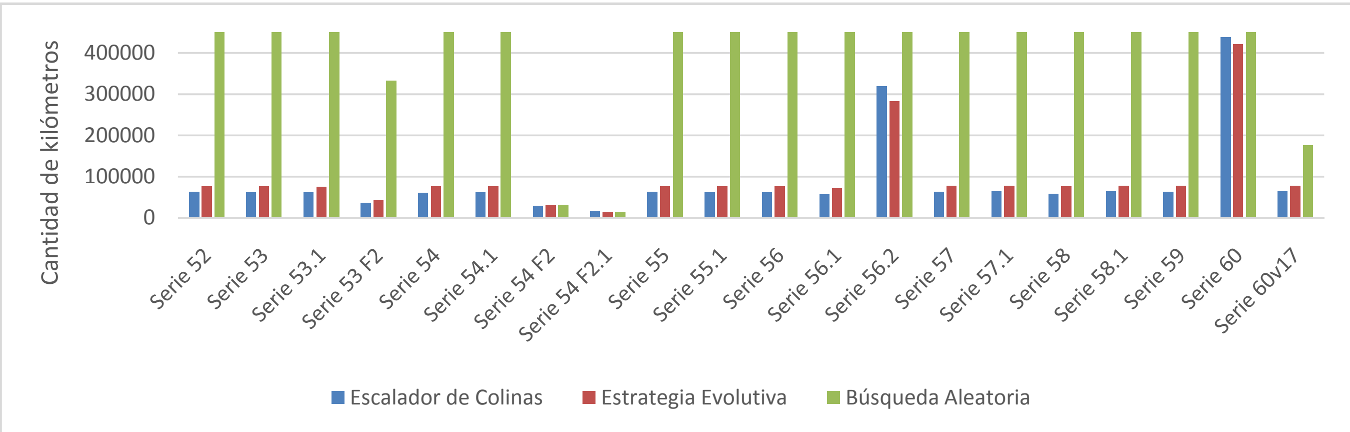

Figura 1: Evaluación de la Función Objetivo para las Metaheurísticas utilizadas

En la Tabla 1 se evidencia los resultados obtenidos tras la aplicación de la prueba de Friedman 1xN y su posterior análisis post-hoc para determinar la metaheurística que ofrece mejores resultados. Para tomar la decisión de aceptar la hipótesis nula que plantea que no existen diferencias entre los algoritmos, se considera un nivel de significancia de 0.05.

Tabla 1: Resultados del ranking obtenido por la prueba de Friedman 1xN para las metaheurísticas

Algoritmo	**Friedman (Ranking)**	**Holm**	**Finner**	**Li**
Escalador de Colinas	**1.3**	*	*	*
Estrategia Evolutiva	**1.9**	**0.05778**	**0.05778**	**0.05778**
Búsqueda Aleatoria	**2.8**	**0.000004**	**0.000002**	**0.000002**
p-valor	**0.000011**			

El * indica el algoritmo de control

Como se observa en la Tabla 1, el p-valor obtenido por la prueba de Friedman es menor que el nivel de significancia establecido para aceptar la hipótesis nula y esta es rechazada. Por tanto, se puede concluir que existen diferencias significativas entre las metaheurísticas analizadas y el algoritmo que ofrece mejores resultados es el Escalador de Colinas

Tomando un nivel de significancia de 0.05 para el análisis post-hoc, la Tabla 1 muestra que no existen diferencias significativas entre la Estrategia Evolutiva y el algoritmo utilizado como método de control, mientras que si existen diferencias entre la Búsqueda Aleatoria y dicho algoritmo. Siendo más permisivos y utilizando un nivel de significancia de 0.10, se evidencia que existen diferencias significativas entre estos algoritmos y el algoritmo de control.

Estos análisis demuestran que, en algunos casos los algoritmos no logran encontrar calendarios que cumplan completamente con las restricciones de cantidad de veces seguidas que un equipo puede jugar de local o visitante. A pesar de esto, el Escalador de Colinas es la metaheurística que mejor resultados ofrece y también la que logra en la mayoría de los casos menos violaciones de las restricciones.

Comparación entre los calendarios obtenidos por los métodos metaheurísticos y los calendarios oficiales

En la Tabla 2, se puede apreciar una comparación entre los mejores resultados obtenidos con el Escalador de Colinas y los resultados de los calendarios oficiales. La columna *KM(O)*, muestra el total de kilómetros en el Calendario Oficial obtenido manualmente, la columna *KM(MHP)* muestra el total de kilómetros promedio de la solución que fue encontrada con el Escalador de Colinas , la columna *KM(MHM)* muestra el total de kilómetros de la mejor solución que fue encontrada con el Escalador de Colinas, la columna *AHORRO*, muestra la diferencia de kilómetros existentes entre las dos primeras columnas, y la columna *%AHORRO* muestra el porciento de ahorro respecto a *KM(O).*

Tabla 2: Comparación entre los resultados oficiales y los mejores resultados obtenidos por el Escalador de Colinas.

Calendarios	KM(O)	KM(MHP)	KM(MHM)	AHORRO	%AHORRO
Serie 52	65281	**62375**	**58531**	2906	4.5
Serie 53	60620	61757	**59402**	-	-
Serie 53 Fase 2	38488	**36172**	**35035**	2316	6.02
Serie 54	62623	**60798**	**56219**	1825	2.9
Serie 54 Fase 2	34380	**28788**	**26871**	5592	16.27
Serie 55	66024	**62453**	**59462**	3571	5.4
Serie 56	66167	**62030**	**59122**	4137	6.3
Serie 57	67918	**62901**	**59519**	5017	7.4
Serie 58	61826	**58356**	**58824**	3470	5.6
Serie 59	**58231**	62488	59827	-	-
Serie 60	**105114**	438062	113096	-	-

Para procesar los resultados obtenidos por el Escalador de Colinas se debe tomar la evaluación promedio de los resultados para obtener una medición más precisa y reducir el error experimental. Esto se debe a la naturaleza estocástica del algoritmo ya que se utiliza la aleatoriedad dentro de las heurísticas de obtención de la solución inicial, y también en la forma en que se seleccionan los operadores de mutación y los componentes que estos requieren (las fechas, enfrentamientos o equipos que se transforman dentro de cada mutación son seleccionados de forma aleatoria). Las heurísticas definidas en este trabajo no son deterministas, y por ende las metaheurísticas que utilizan estas heurísticas subordinadas tampoco lo son.

En la Tabla 2 se puede apreciar que, en 9 de las 11 instancias utilizadas, las soluciones encontradas por el Escalador de Colinas mejoran los resultados oficiales obtenidos. Los resultados de esta metaheurística ahorran entre 2.00% y el 21.84% de kilómetros en comparación con los resultados de los calendarios oficiales utilizados en la SNB.

También en la Tabla 2 se muestra que, en 8 de las 11 instancias utilizadas, las soluciones promedio encontradas por el Escalador de Colinas mejoran los resultados oficiales obtenidos. Los resultados de esta metaheurística ahorran entre 2.9% y el 16.27% de kilómetros en comparación con los resultados de los calendarios oficiales utilizados en la SNB.

Según los resultados obtenidos en las pruebas, el software ofrece una mejoría de un 6.79% de ahorro en cantidad de kilómetros como promedio con respecto a los calendarios oficiales de la SNB, elaborados de forma manual por expertos en el tema. En general, estos resultados muestran que las soluciones ofrecidas por Escalador de Colinas obtienen mejores resultados que los calendarios oficiales en las instancias analizadas. Cabe destacar que solo en la Serie 60 se encontró un calendario no factible ya que no cumplen con las restricciones establecidas.

Conclusiones

Con la realización de este trabajo se cumplieron los objetivos trazados y se pueden extraer las siguientes conclusiones:

- El método actual que posee Cuba para generar los calendarios de la Serie Nacional de Béisbol no es eficiente. Esto se debe a que no poseen una herramienta informática sencilla de utilizar por lo que deben realizar los calendarios de forma manual.
- Se proponen tres variantes de heurísticas de construcción de soluciones iniciales, la primera basada en enfrentamientos, la segunda basada en fechas y la tercera combina las anteriores seleccionando la estrategia de manera aleatoria. La primera variante ofrece los mejores resultados, presentando una mejora del 29.6% respecto a la segunda y una mejora del 24.16% respecto a la tercera.
- Se utilizan tres metaheurísticas para optimizar las soluciones, la primera es el Escalador de Colinas, la segunda es la Estrategia Evolutiva y la tercera es la Búsqueda Aleatoria. La primera metaheurística ofrece los mejores resultados, presentando una mejora del 13.9% con respecto a la segunda y una de mejora del 76.28% respecto a la tercera.
- La nueva propuesta mostró mejores resultados en comparación con el método manual utilizado por los expertos del INDER. Superando en la mayoría de los experimentos al método manual con un 6.79% de ahorro promedio y un 16.27% de ahorro máximo.

Conflictos de intereses

Los autores del presente artículo declaran que no poseen conflicto de intereses.

Contribución de los autores

1. Conceptualización: Alejandro Rosete Suárez, Luis Suárez González, Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
2. Curación de datos: Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
3. Análisis formal: Alejandro Rosete Suárez, Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.

4. Investigación: Alejandro Rosete Suárez, Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
5. Metodología: Alejandro Rosete Suárez, Luis Suárez González, Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
6. Administración del proyecto: Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
7. Recursos: Luis Suárez González.
8. Software: Luis Suárez González, Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
9. Supervisión: Alejandro Rosete Suárez, Luis Suárez González.
10. Validación: Alejandro Rosete Suárez, Luis Suárez González.
11. Visualización:Alejandro Rosete Suárez, Luis Suárez González,Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
12. Redacción – borrador original: Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.
13. Redacción – revisión y edición: Alejandro Rosete Suárez, Luis Suárez González, Mario Herrera Almira, Rafael Alejandro Gayoso Rodríguez.

Financiamiento

La investigación no requirió fuente de financiamiento externa.

Referencias

Alarcón, F., Durán, G., Guajardo, M., Miranda, J., Muñoz, H., Ramírez, L., Ramírez, M., Sauré, D., Siebert, M., Souyris, S., Weintraub, A., Wolf, R., & Zamorano, G. (2017). Operations Research Transforms the Scheduling of Chilean Soccer Leagues and South American World Cup Qualifiers. *INFORMS Journal on Applied Analytics*, *47*(1), 52-69. https://doi.org/10.1287/inte.2016.0861

Chaveco, T. G., Nodarse, C. P., & Suárez, A. R. (2021). Nodos Knime para ajustar modelos usando la biblioteca de clases BiCIAM. *Revista Cubana de Transformación Digital*, *2*(1), 115-129.

Fazayeli, S., Eydi, A., & Nakhai, I. (2018). A model for distribution centers location-routing problem on a multimodal transportation network with a meta-heuristic solving approach. *J Ind Eng Int*, *14*, 327–342. https://doi.org/s40092-017-0218-6 (The Author(s) 2017.)

Khelifa, M., Boughaci, D., & Aimeur, E. (2017). An enhanced genetic algorithm with a new crossover operator for the traveling tournament problem [Proceedings]. *2017 4th International Conference on Control, Decision and Information Technologies (CoDIT)*, 1072-1077. https://doi.org/10.1109/CoDIT.2017.8102741

Kynga¨s, J., Nurmi, K., Kynga¨s, N., Lilley, G., Salter, T., & Goossens, D. (2017). Scheduling the Australian Football League [Creando el Cronograma de la Liga de Futbol Australiana] [Cientific]. *Journal of the Operational Research Society (2017)*, *68(8)*, 973–982. https://doi.org/10.1057/s41274-016-0145-8

Lambers, R., Rothuizen, L., & Spieksma, F. C. (2021). The Traveling Social Golfer Problem: The Case of the Volleyball Nations League. International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research,

Pace, G. (2021). *Randomized Construction Approaches for the Traveling Tournament Problem using Lower Bound Based Heuristics* Wien].

Pérez, A. C. P., Ansola, E. S., & Rosete, A. (2021). Una solución metaheurística al problema de planificación de rutas de auto-buses escolares con flota homogénea y selección de paradas.

Rameez Naqvi, S., Ahmad, A., Riazul Islam, S., Akram, T., Abdullah-Al-Wadud, M., & Alamri, A. (2021). Towards Prevention of Sportsmen Burnout: Formal Analysis of Sub-Optimal Tournament Scheduling. *arXiv e-prints*, arXiv: 2106.09627.

Rosete, A., Paredes, D., & Sánchez, E. (2013). Reducción del costo de transportación en las Series Nacionales de Béisbol empleando metaherísticas [Metaheuristics for the reduction of transportation cost in Baseball National

Series]. *Ingeniería Mecánica*, *16*, 257-271.

Singh, A. P., Gupta, C., Singh, R., & Singh, N. (2021). A Comparative Analysis of Evolutionary Algorithms for Data Classification Using KEEL Tool. *International Journal of Swarm Intelligence Research (IJSIR)*, *12*(1), 17-28.

Talbi, E.-G. (2009). *Metaheuristics: From Desing to Implementation. John Wiley & Sons, Inc.* Published by John Wiley & Sons, Inc., Hoboken, New Jersey and Canada. https://www.wiley.com/en-us/Metaheuristics%3A+From+Design+to+Implementation+-p-9780470278581

Wei, F.-F., Chen, W.-N., Hu, X.-M., & Zhang, J. (2019). An empirical study on evolutionary algorithms for traveling salesman problem. 2019 9th International Conference on Information Science and Technology (ICIST),