Ingeniería Mecánica. Vol. 16. No. 3, septiembre-diciembre, 2013, p. 257-271 ISSN 1815-5944

Artículo Original

Reducción del costo de transportación en las Series Nacionales de Béisbol empleando metaheurísticas

Metaheuristics for the reduction of transportation cost in Baseball National Series

Alejandro Rosete-Suárez, David Paredes-Miranda, Eduardo Sánchez-Anzola

I. Instituto Superior Politécnico José Antonio Echeverría. Facultad de Ingeniería Informática. La Habana. Cuba.
Correo electrónico: rosete@ceis.cujae.edu.cu

Recibido: 19 de abril de 2013 Aceptado: 12 de junio de 2013

Resumen

Este trabajo demuestra cómo se pueden disminuir el costo de transportación en las Series Nacionales de Béisbol de Cuba reduciendo la distancia total recorrida por los equipos. Se planteó el problema de optimización correspondiente para luego enfrentar su solución usando varias metaheurísticas, en el problema de la construcción de los calendarios de las tres Series Nacionales de Béisbol más recientes. Los resultados mejores los obtuvieron las metaheurísticas RRT y Escalador de Colinas. Los calendarios obtenidos ahorran entre un 13 y un 29% de la distancia total recorrida y se obtienen en menos del 1% del tiempo empleado por los expertos humanos que diseñaron los calendarios oficiales de esas competencias. Estos resultados muestran que el uso más eficiente de los portadores energéticos puede enfocarse desde la perspectiva de la optimización de los kilómetros recorridos usando metaheurísticas.

Palabras claves: metaheurísticas, problema de optimización combinatoria, problema de la transportación en un torneo, calendario deportivo.

Abstract

This paper shows that the transportation cost of Baseball National Series may be reduced using metaheuristics. This may be achieved by the reduction of the total distance of the itineraries of the teams. First, the corresponding optimization problem is defined. Then, the solution is faced using metaheuristics. Many metaheuristics were compared in the optimization of the schedules of the last three Baseball National Series. The best results were obtained by RRT and Hill Climbing. The schedules obtained by these metaheuristics allow to save between 13 and 29% of the overall distance travelled by teams, respect to the official schedules designed by human experts. Also, the time to obtain the schedules with metaheuristics is about 1% of the time used by human experts. These results show that a better used of combustible can be also achieved by reducing the overall distance travelled, facing the corresponding optimization problem using metaheuristics.

Key words: metaheuristics, combinatorial optimization problem, traveling tournament problem, sport timetabling.

Introducción

En los años recientes se han intensificado las investigaciones enfocadas al mejor uso de los portadores energéticos, con la introducción de nuevas tecnologías para disminuir su consumo. Esto es muy importante para reducir la contaminación ambiental y mejorar la eficiencia de las organizaciones. En el caso del uso del combustible en la transportación hay otra dimensión a considerar: la reducción de la distancia total recorrida.

En el caso concreto la Serie Nacional de Béisbol (SNB) de Cuba, que es el evento deportivo más importante y complejo del deporte nacional, la ubicación de los enfrentamientos entre los equipos en las distintas fechas del evento influye significativamente en la distancia total recorrida. La ubicación de los enfrentamientos en las fechas implica distintos itinerarios para los equipos, con diferentes distancias a recorrer. En general este problema de optimización está definido como el Problema de la Transportación en un Torneo (en inglés: *Travelling Tournament Problem*, TTP), que se incluye dentro de los problemas de calendarios deportivos (en inglés: *sports timetabling*), y ha recibido mucha atención en el mundo. Este problema ha sido enfrentado usando metaheurísticas debido a su complejidad, provocada por la gran cantidad de variantes posibles de calendarios y la dificultad para crear algoritmos exactos de solución,.

Las metaheurísticas son algoritmos de optimización de propósito general. Su objetivo es obtener el mejor (mayor o menor) valor posible de la función a optimizar (llamada función objetivo) dentro del dominio del problema. Las metaheurísticas son de propósito general, ya que no exigen que la función cumpla con ciertas propiedades, ni que las variables que definen cada solución tengan un dominio particular. Estas características diferencian a las metaheurísticas de los algoritmos heurísticos particulares para cierto tipo de función a optimizar o dominio de las variables. Existen muchas metaheurísticas, y entre las más famosas se pueden incluir los Algoritmos Evolutivos, el Recocido Simulado, el Escalador de Colinas, etc. Las metaheurísticas no garantizan la obtención del valor óptimo de la función objetivo. Sin embargo, logran obtener soluciones buenas en diversos dominios de aplicación. Para una explicación más detallada y obtener una visión más general sobre este tema puede consultarse [21, 24].

Algunas de las metaheurísticas que se usarán en este trabajo se comentan de forma general a continuación y para cada una de ellas se coloca el término en inglés correspondiente. Todas tienen en común que toman varias decisiones de manera aleatoria, y que dan como respuesta la mejor de las soluciones que son evaluadas durante la ejecución del algoritmo. Todas parten de una solución inicial cualquiera, generada de algún modo, generalmente aleatorio. Luego se van analizando nuevas soluciones, sucesivamente, hasta que se llega a una determinada cantidad de soluciones evaluadas. La diferencia entre estas metaheurísticas está en la forma en que exploran el espacio de soluciones posibles. Muchas utilizan el principio de búsqueda local, donde las nuevas soluciones exploradas no son aleatorias, sino que son soluciones parecidas a la anterior, obtenidas a partir de modificaciones de esta.

El Camino Aleatorio (*Random Walk*) parte de una solución dada y a partir de ahí le realiza una modificación que produce una solución nueva, que luego sirve de base para una nuevas modificaciones. A diferencia del Camino Aleatorio, el Escalador de Colinas (*Hill Climbing*), luego de modificar la solución actual y obtener una solución nueva, no siempre utiliza la nueva para la próxima modificación. Se acepta la nueva solución para ser la base de la próxima modificación solo si la nueva es mejor que la anterior. Hay una variante de este algoritmo que acepta también las nuevas soluciones si estas tienen una calidad igual.

A pesar de que la idea del Escalador de Colinas parece muy lógica, tiene una limitación esencial que es su tendencia a quedarse estancado en óptimos locales, cuando todas las soluciones alrededor de la solución actual son peores que esta. Para evitarlo, hay otras metaheurísticas que aceptan soluciones vecinas (resultantes de modificaciones) que sean peores que la actual. Las metaheurísticas que se describen a continuación siguen esa idea. La Aceptación por Umbral (*Threshold Accepting*) acepta soluciones peores que la actual, siempre que el valor del empeoramiento sea menor que un valor de umbral, que es un parámetro del algoritmo. Por su parte, el Recocido Simulado (*Simulated Annealing*) acepta soluciones peores con una probabilidad que es inversamente proporcional al empeoramiento de la función objetivo, y directamente proporcional a un parámetro llamado temperatura, que va decreciendo gradualmente. Esta metaheurística se inspira en el recocido o temple de los metales para garantizar que el material no quede en un estado frágil asociado a un mínimo local de energía. La forma usual de descender la temperatura es multiplicándola por un factor de reducción (valor entre 0 y 1), que es un parámetro de esta metaheurística.

A diferencia de los algoritmos Aceptación por Umbral y Recocido Simulado, que aceptan soluciones peores en relación con la diferencia entre la solución actual y la nueva, existen otras metaheurísticas que la aceptación de las soluciones depende de otros aspectos. El algoritmo RRT (*Record-to-Record Travel*) acepta las soluciones nuevas si la diferencia de calidad entre la solución nueva y la mejor encontrada hasta ese momento es menor que un valor definido como parámetro. El criterio de aceptación de RRT es similar al que emplea la Aceptación por Umbral excepto que la comparación en un caso es con solución actual y en otro con la mejor solución encontrada. Por su parte, la metaheurística conocida como Algoritmo del Gran Diluvio (*Great Deluge Algorithm*) acepta las nuevas soluciones si ellas están por encima de un valor llamado nivel del agua. El nivel del agua parte de un valor que es un parámetro del algoritmo, y se va incrementando sumándole cada vez un valor, definido por otro parámetro llamado lluvia.

Todas la metaheurísticas anteriores mantienen una sola solución de referencia para generar nuevas. Hay otro grupo de metaheurísticas, muchas veces llamadas poblacionales, que en lugar de una sola solución mantienen varias soluciones y generan las nuevas soluciones a partir de esas. Entre las metaheurísticas poblacionales se destacan las Estrategias Evolutivas (*Evolution Strategies*). Esta metaheurística comienza generando una población inicial de soluciones, generalmente de manera aleatoria. Luego, se repite un ciclo en que se encogen las mejores de ellas, para luego ser usadas para generar una nueva población realizándole modificaciones a estas mejores, hasta llenar de nuevo una población completa. Luego, con esta nueva población se repite el proceso. El tamaño de la población y la cantidad de soluciones mejores que se escogen para generar la nueva población son parámetros de esta metaheurística.

Por la complejidad y amplitud del tema, la explicación anterior no entra en detalles, ni en variantes que pueden ser consultadas con mayor detalle en la abundante bibliografía del tema, por ejemplo en [21,24]. En estas mismas referencias puede encontrarse una amplia variedad de aplicaciones. También pueden encontrarse recomendaciones sobre cómo deben ajustarse los parámetros, así como la influencia de estos en el comportamiento de las metaheurísticas. Es importante acotar que no hay ninguna demostración de que una metaheurística sea superior siempre a otra. Particularmente, el Teorema NFL (*No Free Lunch*) [22] establece que si una metaheurística supera a otra en un grupo de problemas, entonces debe esperarse que haya una cantidad similar de problemas donde ocurra lo contrario. Por tanto, en ausencia de otra información, en cada problema concreto deben compararse varias metaheurísticas para determinar cuál es la mejor en el caso.

Para aplicar una metaheurística en un problema concreto deben definirse algunos aspectos importantes:

- La representación de una solución del problema, con las variables de decisión que al combinarse definen el espacio de soluciones posibles.
- La forma de construir soluciones del problema, generalmente de manera aleatoria o pseudo-aleatoria.
- La forma de modificar soluciones del problema, para obtener otras nuevas. Cuando las modificaciones dan una nueva solución a partir de la variación de una sola solución de entrada, generalmente reciben el nombre de operadores de mutación.
- La forma de evaluar las soluciones del problema, es decir, la función objetivo.

El Problema de la Transportación en un Torneo se define por primera vez en [9] motivado por un estudio realizado de los calendarios de las principales ligas deportivas de Estados Unidos, en particular la MLB (*Major League Baseball*), para tratar de satisfacer las exigencias de la mayoría de los participante. Luego se ha aplicado en competencias deportivas de diferentes deportes y países como son: el futbol de Brasil [1,12], el voleibol de Argentina [6], el futbol de Bélgica [11], el tenis [16], entre otras.

El problema consiste en la definición de un calendario para n equipos, siendo n par, donde todos los equipos se enfrentan al menos una vez contra todos los demás equipos de la competencia. Un enfrentamiento entre dos equipos ocurre normalmente en la sede de uno de los dos equipos. Al equipo que efectúa el enfrentamiento en su propia sede se le llama "*home club*" o sede, mientras al otro se le llama "visitador". Para el cálculo de los recorridos de los equipos se necesita de una matriz cuadrada de orden n, con las distancias entre las sedes de los distintos equipos. El principal objetivo de este problema es reducir los recorridos de los equipos durante las competencias.

Otro elemento distintivo en los torneos es la cantidad de vueltas completas a realizarse durante la competencia, lo que ha motivado las siguientes clasificaciones [8, 14, 15]:

- Un "todos contra todos" simple (en inglés: *Simple Round Robin*, SRR): Es una competencia donde cada equipo se enfrenta una sola vez contra los restantes equipos. Se realiza una sola vuelta completa.
- Un "todos contra todos" doble (en inglés: *Double Round Robin*, DRR): Es una competencia donde cada equipo se enfrenta dos veces contra los restantes equipos, jugando una vez como *home club* y otra como visitador, o viceversa. En este caso tienen lugar dos vueltas completas.
- Se han propuestos variantes de cómo enfrentar el problema. Algunas de las más conocidas son:
- Minimizar la distancia total, obtenida de sumar las distancias recorridas por cada equipo [6, 9, 10].
- Minimizar la distancia máxima recorrida por el equipo que tiene el recorrido más largo [6, 8].
- Minimizar la diferencia entre la distancia recorrida por los equipos con el recorrido mayor y el menor [6]

- Maximizar la asistencia de público a los enfrentamientos [2].
- Se han definido otras variantes del problema como son, por ejemplo las siguientes:
- Variante de calendario reflejado (en inglés: *Mirrored Traveling Tournament Problem*, MTTP): en los torneos dobles, o a dos vueltas, se tiene una distribución del calendario para la primera mitad (n-1 primeros enfrentamientos) que se repite en la segunda mitad (n-1 restantes), sólo intercambiando las sedes [8, 9].
- Variante de calendario sin repeticiones (en inglés: *No Repeaters Traveling Tournament Problem*, NRTTP): un enfrentamiento entre dos equipos en el calendario no puede ser precedido o antecedido por otro enfrentamiento de ese mismo par de equipos [9].
- Variante de calendario con sedes predefinidas (en inglés: *Traveling Tournament Problem with Predefined Venues*, TTPPV): en esta variante las sedes de la competencia están predefinidas de antemano y son neutrales para todos los equipos [3,7].
- Variante de calendario sin restricciones (en inglés: *Unconstrained Traveling Tournament Problem*, UTTP): agrupa calendarios sin restricciones de una cantidad limitada de juegos como visitador o *home club*, y que dos equipos puedan enfrentarse consecutivamente, más de una vez [13].

Otro elemento variable en la literatura es la cantidad de equipos. Se han estudiado problemas artificiales con 4, 6 y 8 equipos [9], otros casos reales y artificiales con 12 y 24 equipos [9,12], etc.

Para resolver estos problemas, las formas más empleadas para representar las soluciones expresan los enfrentamientos entre dos equipos, diferenciando quien es *home club* o visitador. Las más utilizadas son:

- Una secuencia de fechas de enfrentamientos, donde cada elemento de la secuencia es un par de equipos que jugará en esa fecha. Generalmente el primero de los equipos que aparece en el par, es *home club*, y el segundo, es visitador [17, 18].
- Una matriz donde las filas son los equipos y las columnas las vueltas simples, siendo el elemento de la fila i y la columna j, el equipo contra el que se enfrenta el equipo i en la vuelta j. Para indicar que el equipo es *home club*, o visitador, se utiliza un signo positivo o negativo [7, 20] o una matriz auxiliar [19].
- El espacio de soluciones del problema es explorado a través de varios operadores o mutaciones. Algunos de los operadores más comunes son [7, 8, 20]:
- Intercambio de *home club* por visitador: en este caso se escoge un enfrentamiento y se intercambia el equipo *home club* y el visitador.
- Intercambio de equipos: se intercambian un equipo por otro en toda la planificación.
- Inserción de juego: se selecciona un juego y se inserta dentro una vuelta, teniendo que ajustar después los restantes enfrentamientos que hagan incorrecta la nueva solución generada.

Se han aplicado diversos métodos para resolver el problema. Entre las metaheurísticas que han sido empleadas están: búsqueda tabú [5, 6], recocido simulado [1, 4], algoritmo genético [16, 18, 20], búsqueda local iterada [7], métodos híbridos entre metaheurísticas como GRASP y búsqueda local iterada [8], métodos híbridos entre colonias de hormigas y búsqueda tabú [17], entre otras.

Como se ha visto, el Problema de la Transportación en un Torneo ha recibido mucha atención en los últimos años. Al existir muchas variantes de él, es muy difícil extrapolar las soluciones reportadas a otro caso, ya que tienen cada una en cuenta diferentes diseños de calendarios o intereses a satisfacer. Estas variaciones del problema hacen que la selección de la metaheurística más conveniente para cada caso sea un problema abierto, lo cual queda reforzado por el Teorema NFL [22].

En la sección siguiente se explican las particularidades de la Serie Nacional de Beisbol en Cuba, y en base a eso, se explica cómo se definen cada uno de los elementos necesarios para optimizar el calendario de estas usando metaheurísticas. Luego, se presentan los resultados experimentales de la comparación entre diferentes configuraciones de metaheurísticas, y la comparación con las soluciones obtenidas por expertos.

Materiales y Métodos

Un calendario

Según la forma de competencia que se use, en cada Serie Nacional de Béisbol (SNB) se definen un conjunto de enfrentamientos que deben efectuarse entre los equipos que participen. Se entiende por enfrentamiento a la realización de tres juegos entre un par de equipos en el estadio de uno de ellos. Por ejemplo, en la SNB 50 cada uno de los 16 equipos participantes se enfrentó dos veces con cada uno de los 15 restantes, una vez en su sede (*home club*), y otra vez en la sede del otro (visitador). Eso implica que cada equipo tenía que realizar 30 enfrentamientos. Para poder ejecutar todos los enfrentamientos, el calendario define 30 fechas para realizar esos 30 enfrentamientos de cada equipo. En cada una de las fechas se realizan varios enfrentamientos. Por ejemplo, en la SNB 50 en cada fecha se realizaron 8 enfrentamientos, que incluyen en total a los 16 equipos. En general, un calendario consiste en la asignación de cada uno de los enfrentamientos a cada una las fechas. Para cada equipo existen tantos calendarios como posibles asignaciones hay de sus enfrentamientos a las fechas. Por ejemplo, para uno de los equipos de SNB 50 hay $30! = 2.65 * 10^{32}$ calendarios posibles, dado por las posibles permutaciones (ordenamientos) de sus enfrentamientos en las fechas. Sin embargo, a la hora de conformar el calendario completo de la competencia debe tenerse en cuenta que en cada fecha un equipo solo puede efectuar un enfrentamiento.

Como se vio en la sección anterior, hay distintos tipos de calendarios. En la variante "todos contra todos" doble, la cantidad de fechas n está dada por la ecuación (1), donde t es la cantidad total de equipos.

$$n = 2(t-1) \quad (1)$$

Esa fue la variante usada en las SNB 50 y 51. No obstante, hubo una diferencia entre ambas SNB, pues la SNB 51 uso la variante de calendario reflejado y no fue así en la SNB 50. Por su parte, en la SNB 52, se empleó la variante "todos contra todos" simple, donde la cantidad de fechas n está dada por la ecuación (2).

$$n = t-1 \quad (2)$$

Adicionalmente, hay otras singularidades en las SNB cubanas como las siguientes:

- Juego inaugural entre finalistas o no: En los últimos años se incorporó una novedad a las SNB que consiste en que se agrega una fecha extra al inicio del calendario normal, para realizar un juego que inaugura la SNB en que participan el equipo campeón y el subcampeón de la SNB anterior. En esa fecha solo se realiza ese juego. Luego, cuando según el calendario le corresponda el enfrentamiento entre esos equipos, el enfrentamiento solo incluye otros dos juegos, para completar los tres juegos de este. Por ejemplo, en las SNB 50 el juego inaugural fue entre los equipos Industriales y Villa Clara, en la SNB 51 fue entre los equipos Pinar del Río y Ciego de Ávila, y en la SNB 52 fue entre Ciego de Ávila e Industriales.
- Se incluye o no descanso. Si la SNB incluye una cantidad impar de equipos, en cada fecha tiene que haber un equipo que descansa, es decir no juega en la fecha y el equipo regresa a su sede a descansar. Por ejemplo, la SNB 51 que incluyó $t = 17$ equipos tuvo descanso. En las SNB 50 y 52 no lo hubo.
- Restricciones en los enfrentamientos seguidos como visitador: Comúnmente esta cantidad es de 4 enfrentamientos.
- Las dos primeras de estas singularidades no se han encontrado en la literatura consultada.

Representación

Para poder resolver el problema de optimizar el calendario de una SNB es necesario definir la forma en que se representa cada solución (calendario). En este trabajo, cada calendario C se representa como una vector de n dimensiones que son las fechas. La $i-ésima$ fecha F_i es un conjunto con una cantidad de m enfrentamientos, donde el $j-ésimo$ enfrentamiento de la fecha F_i, es decir el enfrentamiento $\mathrm{E_{ij}}$, es un par ordenado que incluye a los equipos que se enfrentan S y V, siendo S la sede y V el visitante. Formalmente se puede decir:

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

La expresión 3 asegura que un equipo no aparezca dos veces en la misma fecha.

$$\forall_i \forall_j \forall_k\, [S_{ij} \neq V_{ij}, S_{ij} \neq S_{ik}, S_{ij} \neq V_{ik}, V_{ij} \neq V_{ik}] \qquad si \;\; j \neq k;\ i,j = 1,2, \dots m \quad (3)$$

En cuanto a la cantidad de enfrentamientos m por cada fecha F_i, si t es par y no hay descansos se calcula con la ecuación (4), y si t es impar y, por tanto, hay descanso, se calcula con la ecuación (5).

$$m = \frac{t}{2} \quad (4)$$

$$m = \frac{(t-1)}{2} \quad (5)$$

Por ejemplo, un calendario para una hipotética SNB con $Equipos = \{A, B, C, D\}$ y calendario doble, usando las expresiones anteriores, quedaría $t = 4$, $n = 6$ y $m = 2$. Un calendario podría representarse como sigue:

$$Calendario\ 1 = [\{(A,B),(C,D)\},\{(A,C),(B,D)\},\{(D,A),(C,B)\},\{(B,A)(D,C)\},\{(C,A)(D,B)\},\{(A,D)(B,C)\}]$$

Para facilitar la comprensión, en la tabla 1 se muestran ese y otros dos calendarios de esa hipotética SNB. Todos estos calendarios son dobles, restringidos, sin juego inaugural, ni descanso, ni otras restricciones. En cada enfrentamiento $X-Y$ se entiende que el primero de ellos X es el equipo que juega como sede (o "*home club*") mientras que Y es el visitador. La tabla 2 muestra los itinerarios a recorrer por cada uno de los equipos para el *Calendario 1* de la tabla 1.

Tabla 1. Tres calendarios posibles para una hipotética SNB de 4 equipos.

Calendario	**Fecha 1**	**Fecha 2**	**Fecha 3**	**Fecha 4**	**Fecha 5**	**Fecha 6**
Calendario 1	*A-B, C-D*	*A-C, B-D*	*D-A, C-B*	*B-A, D-C*	*C-A, D-B*	*A-D, B-C*
Calendario 2	*C-B, A-D*	*C-A, B-D*	*D-C, A-B*	*B-C, D-A*	*A-C, D-B*	*C-D, B-A*
Calendario 3	*A-C, B-D*	*A-B, C-D*	*D-A, C-B*	*C-A, D-B*	*B-A, D-C*	*A-D, B-C*

Tabla 2. Itinerarios a recorrer por los equipos en Calendario 1 de tabla 1.

Calendario	**Fecha 1**	**Fecha 2**	**Fecha 3**	**Fecha 4**	**Fecha 5**	**Fecha 6**
Calendario 1	*A-B, C-D*	*A-C, B-D*	*D-A, C-B*	*B-A, D-C*	*C-A, D-B*	*A-D, B-C*
Itinerario de A	*A*	*A*	*D*	*B*	*C*	*A*
Itinerario de B	*A*	*B*	*C*	*B*	*D*	*B*
Itinerario de C	*C*	*A*	*C*	*D*	*C*	*B*
Itinerario de D	*C*	*B*	*D*	*D*	*D*	*A*

Cuando la cantidad de equipos aumenta, aumentan también de manera notable las variantes de calendarios. En este ejemplo con 4 equipos, 6 fechas y con calendario doble con restricciones cada equipo tiene solo $3! = 6$ variantes de itinerarios. En la SNB 50 con 16 equipos, 30 fechas y sin restricciones cada equipo tiene $30! = 2.65 * 10^{32}$ posibles itinerarios. Esto permite percibir la complejidad del proceso de optimizar un calendario para minimizar el costo de transportación entre sedes.

Costos asociados a un calendario

La ejecución de un calendario incluye distintos costos como los de alojamiento, iluminación de los estadios, etc. Este trabajo se concentra en uno de estos aspectos: el costo de transportación de los equipos entre las sedes. Puede observarse que muchos de los otros costos son fijos y no dependen del calendario. Sin embargo, el costo de transportación entre las sedes varía con el calendario, porque el itinerario de un equipo dependerá del enfrentamiento que corresponde a cada fecha.

En las SNB la transportación de los equipos se realiza en ómnibus, moviéndose cada equipo de la sede en que ocurra un enfrentamiento en una fecha dada a la sede del enfrentamiento de la fecha siguiente. Esto implica que el costo total de la transportación entre sedes está determinado por el itinerario que debe recorrer cada equipo para cumplir el calendario. El costo de transportación dependerá del costo del combustible empleado, del desgaste de los neumáticos, etc. Sin embargo, todo esto depende en esencia de la cantidad de kilómetros recorridos por todos los equipos. En la SNB 50 la distancia total recorrida entre todos los equipos fue de más de 100 mil kilómetros, y en la SNB 51 fue de más de 175 mil. Esta diferencia entre ambas estuvo marcada por dos aspectos importantes. En la SNB 51 hubo un equipo más, y además esto hizo que la cantidad de equipos fuera impar. Esto implicó que en cada fecha el equipo que descansaba tenía que regresar a su sede, y luego partir de ahí a la sede de su enfrentamiento en la fecha siguiente.

La figura 1 muestra, gráficamente, las implicaciones de cada calendario en costo. Cada uno de ellos es un grafo donde los nodos representan las sedes de cada equipo, y al lado de cada arco que une las sedes se ha colocado una línea de un color dado por cada vez que el equipo identificado con ese color tiene que recorrerlo. La tabla 3 muestra los costos de transportación para ciertas distancias entre los puntos. Este ejemplo simple muestra las notables diferencias en costos asociados a los calendarios.

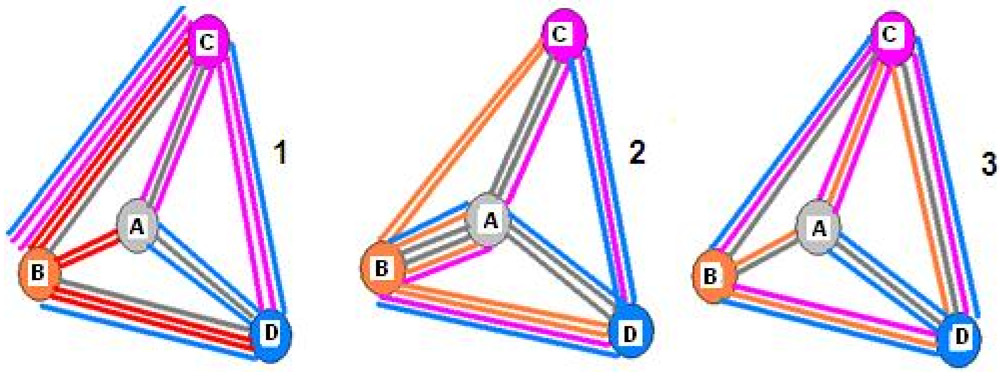

Fig. 1. Representación gráfica de los tres calendarios que se muestran en la tabla 1.

Tabla 3. Distancia entre arcos en el ejemplo hipotético y las veces que se recorren en los tres calendarios

			Costo de los arcos				
Calendario	**AB**	**AC**	**AD**	**BC**	**BD**	**CD**	**Costo**
	2.5	**3.5**	**3.5**	**6.0**	**5.5**	**5.5**	
1	2	3	3	6	4	3	100.5
2	6	3	3	2	4	3	86.5
3	2	3	3	3	3	4	82.5

Para emplear las metaheurísticas para resolver este problema, hay que definir la forma de evaluar cada uno de los calendarios. En este caso, la evaluación incluirá los dos aspectos.

El primer aspecto a incluir en la evaluación es el total de kilómetros recorridos (KMT). Este consiste en la suma de los kilómetros recorridos por cada equipo como muestra la ecuación (6), donde la cantidad de kilómetros recorridos por cada equipo está dada por la suma de las distancias recorridas entre las sedes consecutivas en el itinerario, según se muestra en la ecuación (7), la ecuación (8) y la ecuación (9).

$$KMT = \sum KM_e,\ e \in Equipos \tag{6}$$

$$KM_e = \sum_{\mathrm{i=1}}^{\mathrm{m-1}} \mathrm{distancia}\big(\mathrm{sede(e,i)}, \mathrm{sede(e,i+1)}\big) + \mathrm{distancia}(\mathrm{ciudad(e)}, \mathrm{sede(e,1)}) + \mathrm{distancia}(\mathrm{sede(e,m)}, \mathrm{ciudad(e)}\ ). \tag{7}$$

$$ciudad(e) = ciudad\ que\ es\ sede\ oficial\ del\ equipo\ e. \tag{8}$$

$$sede(e,i) = \begin{cases} ciudad(q)\ si\ \ (q,e) \in F_i \\ ciudad(e), si\ (e,q) \in F_i \\ ciudad(e), si\ (e,q) \notin F_i, (q,e) \notin F_i \end{cases} \tag{9}$$

La función $sede(e,i)$ devuelve la ciudad que es sede del enfrentamiento que corresponde al equipo e en la fecha i, que dependerá de si le corresponde ser sede o visitador en el enfrentamiento de esa fecha. Para calendarios con descansos, en cada fecha que un equipo descansa se considera equivalente a si jugara como sede, porque implica un regreso a su ciudad, y es la razón de la última condición de la función sede.

Debe notarse que KM_e, que es el costo total en kilómetros recorridos por el equipo e, suma también el costo inicial del movimiento de la sede del equipo e a la sede del enfrentamiento en la primera fecha, así como el movimiento final de regreso luego de la última fecha. Cuando hay semanas de descanso general para todos los equipos (por ejemplo en fin de año), en esas fechas los equipos regresan a su sede, lo cual se incluye también en el itinerario, y consecuentemente en los costos de transportación.

La función $distancia(X,Y)$ calcula la distancia entre dos ciudades, y se asume como entrada.

El otro aspecto a incluir en la evaluación es el incumplimiento de la restricción de encuentros seguidos como visitador (IEV), que se muestra en la ecuación (10). Aunque un calendario sea correcto, puede ocurrir que aparezca en él un equipo que tenga una cantidad de enfrentamientos seguidos como visitador que exceda una cantidad fijada por criterios de diseño. Esto se hace para evitar largos períodos fuera de su sede de cada equipo.

$$IEV = \sum IEV_e \, , e \in Equipos. \tag{10}$$

IEV_e cuenta la veces que en el itinerario un equipo e juega más de V juegos como visitador de manera consecutiva. Por ejemplo, Itinerario de A que aparece en la tabla 2 que es A, A, D, B, C, A si $V = 2$ habría un incumplimiento porque se incluye la secuencia de fechas D, B, C en que A es visitador. Esto hace que IEV_A=1. Puede notarse en la misma tabla que en los demás equipos se cumple la restricción para $V = 2$, por lo que el único incumplimiento es con el equipo A, quedando finalmente $IEV = 1$. En las SNB se trabaja normalmente con $V = 4$, lográndose calendarios que generalmente respeten esto, como en la SNB 50. Sin embargo, en la SNB 51 se incumplió esta restricción una vez ($IEV = 1$), porque una vez el equipo Villa Clara tuvo 5 enfrentamientos seguidos como visitador.

Para combinar ambos aspectos en la función objetivo se usará una suma ponderada de ambos aspectos, quedando como aparece en la ecuación (11).

$$FO = -(KMT + P * IEV) \tag{11}$$

En esta función objetivo FO, P es un valor para penalizar los calendarios que incumplan con la restricción de diseño. El valor de penalización es importante para las metaheurísticas. Si P es pequeño, el algoritmo podría devolver soluciones que violen las restricciones pero que tiene buenos valores en otros aspectos. Para garantizar que no se violen las restricciones se usa un valor grande para P. En este caso se usa $P = 1\,000\,000$, que garantiza que los calendarios son mejores mientras menos incumplan las restricciones. Dentro de los que tengan un incumplimiento IEV igual, son preferibles las de menos kilómetros recorridos KMT. Puede notarse que FO es siempre negativa, porque lo ideal es que KMT y IEV tengan el menor valor posible. En este trabajo se enfocará la metaheurística a maximizar esta función negativa, lo cual es equivalente a minimizar la función positiva opuesta.

Construcción

Para construir un calendario inicial correcto, deben ubicarse todos los enfrentamientos posibles en las fechas, sin tener en cuenta el valor de la función objetivo.

El conjunto de enfrentamientos es $TE = \{(x, y) | x \neq y, \; x, y \in Equipos\}$, para el caso general de un calendario doble. En ese caso, w es el total de enfrentamientos (cardinalidad de TE) y queda como muestra la ecuación (12). Para el calendario hipotético con $t = 4$ de la tabla 2 se cumple $w = 12$.

$$w = \frac{t!}{(t-2)!} \tag{12}$$

En el caso del calendario simple, como se juega un solo enfrentamiento entre cada par de equipos entonces se calcula como aparece en la ecuación (13).

$$w = \frac{t!}{(t-2)!\,2!} \tag{13}$$

En este caso, w solo cuenta las combinaciones de posibles parejas, sin diferenciar entre quien está en cada rol pues entre cada par de equipos hay un solo juego. En estos calendarios con un solo enfrentamiento entre cada par de equipos, es la dirección de la SNB la que decide donde se jugará cada enfrentamiento.

En cualquier caso, el algoritmo de construcción de un calendario debe recibir una lista de enfrentamientos TE y debe producir un calendario. Este consiste en la asignación de los enfrentamientos a las fechas, cumpliendo las condiciones que lo hacen ser un calendario correcto, es decir que este cumpla con la condición de que no se repita un equipo en cada fecha, y que se ubiquen todos los juegos posibles. En este punto, no se tiene en cuenta el costo del calendario, ni ninguna otra restricción.

Para lograrlo se implementó un algoritmo de construcción que parte de un calendario inicial vacío (sin enfrentamientos por fecha) y de una lista de enfrentamientos por asignar EPA (que contiene a cada uno de los enfrentamiento de TE). El algoritmo va formando el calendario incrementalmente, agregando cada vez un nuevo enfrentamiento a una fecha, garantizando que este enfrentamiento no contenga (ni como sede, ni como visitador) a un equipo que ya esté contenido (como sede o como visitador) en los juegos anteriormente asignados a esa fecha. Cada vez que se agrega un enfrentamiento a una fecha, este se elimina de EPA.

En cada momento puede existir más de un juego potencialmente asignable a una fecha, comenzando por la situación trivial inicial en que no haya aún enfrentamientos en una fecha, y que por tanto cualquier enfrentamiento puede ser asignado. Para decidir cuál de los enfrentamientos posibles se asigna a la fecha, el algoritmo los toma en el orden en que aparecen en EPA. Esto implica que este algoritmo de construcción produce diferentes calendarios según el orden de los enfrentamientos de TE dentro de EPA. En este trabajo se experimentará con un orden fijo de occidente a oriente, donde los enfrentamientos en EPA aparecen ordenados según el orden de occidente a oriente de los equipos, ordenando primero por la sede y dentro de los de igual sede por el visitador.

Modificación

Una vez definida la representación y la forma en que se construye la solución inicial, para utilizar una metaheurística deben definirse operadores que modifican una solución dada para generar una nueva solución que sea una variación de ella. A estos operadores se les llama comúnmente operadores de mutación o simplemente mutaciones. Aquí, como se parte de un calendario construido correctamente, los operadores de mutación deben garantizar que las nuevas soluciones generadas sigan siendo correctas. Las mutaciones que se emplearán en este caso son las siguientes:

- M_0: cambiar una fecha de lugar (por ejemplo, la fecha 4 pasa a ser la 2 y se desplazan las restantes fechas en orden)
- M_1: cambiar un equipo por otro.
- M_2: cambiar en una fecha los que están de visitador por sede.
- M_3: en una fecha, intercambiar en un enfrentamiento el que está de visitador por el que está de sede.
- M_4: invertir las sedes (visitador por sede, viceversa) de un equipo.
- M_5: invertir el orden de una sección de fechas consecutivas, por ejemplo (F_1, $\boldsymbol{F_2}$, $\boldsymbol{F_3}$, $\boldsymbol{F_4}$, $\boldsymbol{F_5}$, F_6) pasaría a ser (F_1, $\boldsymbol{F_5}$, $\boldsymbol{F_4}$, $\boldsymbol{F_3}$, $\boldsymbol{F_2}$, F_6).
- M_6: intercambiar dos fechas, pasando cada una al lugar de la otra.
- M_7: intercambiar un enfrentamiento de fecha con uno de otra fecha, ajustando los demás enfrentamientos para que sigan siendo correctas ambas fechas.

Cada una de estas mutaciones es realmente una familia de mutaciones, porque los parámetros concretos de cada una de ellas se generan aleatoriamente. Por ejemplo, M_6 incluye el intercambio de cualquier par de fechas, y por tanto incluye realmente $t(t-1)$ variantes de pares de fechas a intercambiar. Para el ejemplo hipotético de 4 equipos y 6 fechas habría 30 variantes de M_6, por ejemplo intercambiar 2 y 5, o intercambiar 4 y 6. Lo mismo ocurre en las demás, con las posibilidades de variar los equipos en M_1 y M_4, o el enfrentamiento en M_3. En la tabla 4 aparecen ejemplos de las mutaciones anteriores. Como los calendarios deben seguir siendo correctos después de aplicarse las mutaciones, cuando se realiza una mutación como M_2 y M_3 que modifica el equipo que es sede en una fecha dada también deben cambiarse los otros enfrentamientos entre ese par de equipos. Es curioso notar que la M_7 produce en este caso lo mismo que si se hubiera intercambiado las fechas 1 y 2. Esto pasa porque para lograr que las fechas sigan siendo correctas y se mantengan las sedes hubo que cambiar todos los juegos. Solo para ejemplificar mejor esta mutación M_7 se agrega el ejemplo siguiente con un calendario con más encuentros por sede que permite mostrar lo que ocurre. Suponga un calendario $\{\ldots, F_x, \ldots, F_y, \ldots\}$ con 8 equipos $(A, B, C, D, E, F, G, y\ H)$, donde los puntos suspensivos representan fechas del calendario que no se modificarán y siendo:

- $F_x: A-B, C-D, E-F, G-H$
- $F_y: A-C, B-D, E-G, H-F$
- Si se aplica la mutación M_7 seleccionando las fechas F_x y F_y, e intercambiando los primeros enfrentamientos de estas, esto hará que el calendario nuevo pase a ser el siguiente (…,F_x', …, F_y', …) siendo
- F_x': $A-C, B-D, E-F, G-H$
- F_y': $A-B, C-D, E-G, H-F$

Puede notarse que los cambios solo afectaron a los equipos involucrados en los enfrentamientos que serán intercambiados, y no a la fecha completa.

Tabla 4. Ejemplos de las mutaciones sobre Calendario 1 de la tabla 2, subrayando los cambios

Mutación	Fecha 1	Fecha 2	Fecha 3	Fecha 4	Fecha 5	Fecha 6
Calendario 1 (sin cambios)	*A-B, C-D*	*A-C, B-D*	*D-A, C-B*	*B-A, D-C*	*C-A, D-B*	*A-D, B-C*
M_0: la fecha 4 pasa a posición 2	*A-B, C-D*	*B-A, D-C*	*A-C, B-D*	*D-A, C-B*	*C-A, D-B*	*A-D, B-C*
M_1: cambiando A y C	*C-B, A-D*	*C-A, B-D*	*D-C, A-B*	*B-C, D-A*	*A-C, D-B*	*C-D, B-A*
M_2: fecha 3	*A-B, C-D*	*A-C, B-D*	*A-D, B-C*	*B-A, D-C*	*C-A, D-B*	*D-A, C-B*
M_3: fecha 3 (enfrentamiento 1)	*A-B, C-D*	*A-C, B-D*	*A-D, C-B*	*B-A, D-C*	*C-A, D-B*	*D-A, B-C*
M_4: equipo A	*B-A, C-D*	*C-A, B-D*	*A-D, C-B*	*A-B, D-C*	*A-C, D-B*	*D-A, B-C*
M_5: fechas 2, 3, 4 y 5	*A-B, C-D*	*C-A, D-B*	*B-A, D-C*	*D-A, C-B*	*A-C, B-D*	*A-D, B-C*
M_6: fechas 2 y 4	*A-B, C-D*	*B-A, D-C*	*D-A, C-B*	*A-C, B-D*	*C-A, D-B*	*A-D, B-C*
M_7: fecha 1 (enfrentamiento 1) y fecha 2 (enfrentamiento 1)	*A-C, B-D*	*A-B, C-D*	*D-A, C-B*	*B-A, D-C*	*C-A, D-B*	*A-D, B-C*

A partir de lo anterior puede verse que hay ciertas mutaciones que modifican las sedes de los enfrentamientos, que son M_1, M_2, M_3 y M_4. Como en las SNB que tienen calendarios simples debe respetarse las sedes prefijadas, entonces para obtener los calendarios de estas SNB solo se emplean las mutaciones M_0, M_5, M_6 y M_7, por ejemplo en la SNB 52. En las demás, todas las mutaciones están disponibles, y se selecciona en cada paso la mutación a usar de manera aleatoria.

Casos de estudio

Para estudiar el comportamiento de las metaheurísticas en el problema, se estudiaron tres casos de calendarios de SNB, que son los correspondientes a las últimas y que además tienen suficiente diferencia entre ellos. La tabla 5 resume estas diferencias, manteniendo la nomenclatura donde t es la cantidad de equipos, n es la cantidad de fechas, m es la cantidad de encuentros por fecha. Las demás columnas indican si el calendario fue doble (D) o simple (S), si incluyó descansos antes de determinadas fechas (DET), si hubo equipos que descansaban en cada fecha (EDF).

Tabla 5. Datos de los tres calendarios de SNB estudiados

SNB	T	n	M	D/S	DET	EDF	KMT	IEV
50	16	30	8	D	9, 18	No	100391	0
51	17	34	8	D	9, 20	Sí	176008	1
52	16	15	8	S	9	No	65281	0

En todas hubo juego inaugural entre el campeón y subcampeón de la SNB anterior. Adicionalmente debe decirse que aunque las SNB 50 y 52 tienen el mismo total de equipos ($t = 16$), hay 14 de ellos que son iguales, pero hay dos diferentes. Las columnas KMT y la columna IEV se corresponden con los dos aspectos considerados en la evaluación de los calendarios: total de kilómetros recorridos, e incumplimiento de la restricción de juegos seguidos como visitador. La evaluación de estos aspectos se realizó sobre los calendarios oficiales de ambas SNB. Se incluyen como referencia respecto a los resultados que se mostrarán más adelante obtenidos por las metaheurísticas. En los casos de las SNB 50 y 51 los calendarios fueron creados manualmente por expertos de la Comisión Nacional de Béisbol que llevan años realizando esta tarea, para la cual emplean varias semanas. En el caso de la SNB 52, el calendario que se refiere es el correspondiente a la primera etapa. Este calendario fue realizado de una manera semiautomática, partiendo de una versión preliminar de esta propuesta con modificaciones hechas por los expertos manualmente.

Metaheurísticas a comparar

Según el Teorema NFL [22], no existe una metaheurística que sea superior a las demás en la totalidad de problemas. Esto lleva a la necesidad de experimentar con varias metaheurísticas para poder saber cuál se comporta mejor en el problema de interés. En este trabajo se compararán varias metaheurísticas, usando los parámetros mostrados en la tabla 6, dando lugar a diferentes configuraciones de metaheurísticas a ser comparadas. Se ha preferido mantener los nombres identificadores por sus siglas en inglés por ser más conocidos, y para cada caso se indica el nombre en inglés y su traducción al español.

Tabla 6. Configuraciones de metaheurísticas a comparar en tres calendarios de la SNB

Nombre	**Metaheurística base**	**Parámetros**
ES100-20	Estrategias Evolutivas (*Evolution Strategies*)	Reemplazo generacional, tamaño de la población 100, mejores a seleccionar 20
ES50-10	Estrategias Evolutivas (*Evolution Strategies*)	Reemplazo generacional, tamaño de la población 50, mejores a seleccionar 10
ES5-1	Estrategias Evolutivas (*Evolution Strategies*)	Reemplazo generacional, tamaño de la población 5, mejores a seleccionar 1
SA99	Recocido Simulado (*Simulated Annealing*)	Temperatura inicial 60811 Factor de reducción 0.99
SA999529	Recocido Simulado (*Simulated Annealing*)	Temperatura inicial 60811 Factor de reducción 0.999529
SHCFA	Escalador de Colinas (*Hill Climbing*)	Primer ascenso (*First Ascent*), y acepta soluciones con igual evaluación
SHCFAJI	Escalador de Colinas (*Hill Climbing*)	Primer ascenso (*First Ascent*), y acepta solo las soluciones que son mejores (*Just improvement*)
RW	Camino Aleatorio (*Random Walk*)	
TA	Aceptación por umbral (*Threshold accepting*)	Umbral para aceptación respecto a la solución actual de 1000
GDA	Algoritmo del Gran Diluvio (*Great Deluge Algorithm*)	Nivel del agua inicial -10000000, Lluvia 993
RRT	Algoritmo RRT (*Record-to-Record Travel*)	Umbral para aceptación respecto a la mejor solución encontrada de 1000

Cada una de las variantes de configuración de parámetros de metaheurísticas de la tabla 6 partieron de criterios recomendados para cada una de ellas [21,24], y se probaron ajustes de los mismos hasta llegar a los valores mostrados.

Criterios para analizar y comparar la calidad de los resultados

Para hacer justa la comparación de estas configuraciones, debe fijarse una cantidad igual de evaluaciones de la función objetivo. En este caso, este valor se fijó en 10 mil, debido a que luego de esta cantidad ninguna de las configuraciones mostró mejorías significativas.

En todos los casos, se realizaron 30 repeticiones debido a que las 11 configuraciones tienen un comportamiento estocástico, no devolviendo siempre el mismo resultado. De cada repetición de una configuración se obtuvo el máximo valor de la función objetivo, entre las 10 mil soluciones evaluadas.

Para integrar y comparar los resultados de cada una de las configuraciones, se usaron cuatro medidas estadísticas para integrar los resultados de las 30 repeticiones de cada una de ellas: media aritmética, mediana, mínimo y máximo. Las columnas MA, M, MIN y MAX en las tablas 7, 8 y 9 se corresponden a estas medidas, respectivamente.

También se unieron las $30 * 11 = 330$ soluciones mejores encontradas en cada una de las 30 repeticiones de las 11 configuraciones y se ordenaron, dándole un valor de rango a cada una, obteniendo el valor de rango 1 la mejor (máximo) y de 330 la peor (mínimo). Analizando el rango de las soluciones generadas por cada metaheurística permitirá saber la tendencia a obtener soluciones mejores, en adición a las medidas de tendencia central y extrema de cada configuración. Las columnas P10, P25, P50 y P100 en las tablas 7, 8 y 9 se corresponden con el porciento de las 10, 25, 50 y 100 mejores soluciones que son obtenidas usando cada configuración. Adicionalmente, se compararon las configuraciones usando la prueba de la suma de rangos de *Wilcoxon* con una significación de 0.005, tomando como poblaciones a comparar las 30 repeticiones de cada una. Esto sigue la recomendación dada en [23] de usar pruebas no paramétricas para comparar metaheurísticas.

Resultados

Tabla 7. Comparación de las metaheurísticas en la SNB Nº 50

Configuración	MIN	MAX	MA	M	P10	P25	P50	P100
ES100-20	-110870	-97711	-104240	-104565		4	6	16
ES50-10	-123296	-100469	-108113	-107365			2	6
ES5-1	-140388	-116813	-133273	-134799				
GDA	-22097431	-20097723	-20164380	-20097723				
RRT	-111115	-90088	**-100054**	-100603	40	40	30	24
RW	-157328	-145091	-153413	-153604				
SA99	-117503	-98270	-107466	-107453			4	7
SA999529	-136342	-117884	-126801	-126149				
SHCFA	-110988	-89678	-102123	-102633	40	20	24	20
SHCFAJI	**-110380**	**-87043**	-100453	**-100502**	20	36	32	21
TA	-119436	-99561	-110183	-111251			2	6

Tabla 8. Comparación de las metaheurísticas en la SNB Nº 51

Configuración	MIN	MAX	MA	M	P10	P25	P50	P100
ES100-20	-159240	-136823	-147370	-148332				3
ES50-10	-158151	-138350	-148832	-149147				3
ES5-1	-169075	-158785	-164637	-165217				
GDA	-24105666	-24105666	-24105666	-24105666				
RRT	-145052	**-124893**	**-135165**	-136904	70	41	32	29
RW	-193261	-173546	-181528	-181555				
SA99	**-143120**	-131185	-139177	-140088		9	9	17
SA999529	-163863	-143817	-155177	-156117				
SHCFA	-147575	-125489	-136478	**-135743**	20	37	36	22
SHCFAJI	-145646	-125128	-138342	-138644	10	13	21	21
TA	-150540	-133238	-142845	-142204			2	5

Tabla 9. Comparación de las metaheurísticas en la SNB 52

Configuración	MIN	MAX	MA	M	P10	P25	P50	P100
ES100-20	**-61075**	-55415	-58960	-58774	11	13	12	20
ES50-10	-61858	-55768	-59039	-59134	22	13	14	18
ES5-1	-68278	-61998	-65971	-65966				
GDA	-75319	-70747	-73479	-73655				
RRT	-61597	-55093	**-58153**	**-58470**	45	35	30	29
RW	-75647	-65563	-73236	-73901				
SA99	-64808	-57797	-61311	-61021			3	1
SA999529	-71155	-63105	-67451	-68054				
SHCFA	-62635	**-54051**	-59138	-59529	22	22	23	16
SHCFAJI	-63040	-56739	-59635	-59924		17	16	15
TA	-65790	-57634	-63080	-63467			2	1

Discusión

Puede verse que el comportamiento de las distintas configuraciones es muy similar en las tres SNB analizadas, pudiendo agruparse las configuraciones según la calidad de los resultados de cada una.

- Ganadores: RRT, SHCFA, SHCFAJI. Estas tres configuraciones superan a las demás en la mayoría de los criterios y casos. RRT es la que generalmente obtiene los mejores resultados en las distintas medidas, siendo siempre superior a todas las metaheurísticas de los demás grupos usando la prueba de Wilcoxon, lo cual ocurre también generalmente para SHCFA y SHCFAJI. Usando la prueba de Wilcoxon no se obtuvo diferencia significativa entre las tres metaheheurísticas de este grupo. Debe notarse que estas metaheurísticas fueron las que obtuvieron la mejor solución para cada SNB aunque las dos del grupo quedaron cerca. RRT parece tener ligera ventaja en general, a pesar de no poderse establecer una superioridad estadísticamente significativa, por lo que se consideran ganadoras a las tres.
- Perdedores: RW, GDA. Estas metaheurísticas son superadas de forma general por las otras. En el caso de RW este resultado es común, debido que usan muy poco la información anterior para guiar la búsqueda. Para GDA este resultado es inesperado, y se deben a la dificultad para parametrizar esta metaheurística en este caso, llegando a obtener soluciones no factibles para dos SNB (valores por encima de 1000000).
- Medios: Las otras seis configuraciones (basadas en Recocido Simulado, Aceptación por Umbral y Estrategias Evolutivas) tienen un comportamiento intermedio, superándose indistintamente entre ellas, y quedando generalmente superadas por las metaheurísticas del grupo de los Ganadores y superando a las del grupo de los Perdedores. Puede destacarse en este grupo, que las configuraciones ES100-20 y SA99 quedaron cerca del grupo de los ganadores en algunos indicadores.

Finalmente la tabla 10 muestra una comparación de los resultados obtenidos en comparación con los calendarios obtenidos de forma manual por los expertos de la Comisión Nacional de Béisbol y que fueron las que se usaron como calendarios oficiales en las tres SNB. Para construir los calendarios de las SNB 50 y 51, los expertos emplearon varias semanas. En el caso de la SNB 52 se empleó como base una variante previa de esta propuesta que luego fue modificada manualmente por los expertos.

La columna KMT(O) muestra el total de kilómetros en el Calendario Oficial obtenido manualmente, mostrado en la tabla 5. La columna KMT(M) muestra el total de kilómetros de la solución mejor que fue encontrada con metaheurísticas, según lo mostrado en las tablas 7, 8 y 9. La columna Ahorro muestra la diferencia entre KMT(O) y KMT(M), y a la derecha el porciento del ahorro respecto a KMT(O). La columna CS muestra cuántas configuraciones de las 11 empleadas superan KMT(O) en cada SNB, y a su derecha el porciento que representa respecto a 11.

En la tabla 10 puede verse los notables ahorros en kilómetros que pueden obtenerse con la propuesta, especialmente en las SNB 51 y 52. Estas SNB, al tener variaciones en la forma de su calendario respecto a las SNB anteriores, los expertos pudieron emplear menos su experiencia pues se enfrentaban a la tarea de construir calendarios a los que no estaban acostumbrados. Igualmente es significativo que para la SNB 50 se haya obtenido un ahorro de más del 10% a pesar de que para construirlo los expertos pudieron emplear en él todo el conocimiento de años anteriores, pues la forma general del calendario hasta esa SNB se había usado por más de 10 años. Es notable que 6 de las 11 configuraciones sean mejores que este calendario oficial de la SNB 50. Estas configuraciones fueron RRT, SHCFA, SHCFAJI, ES100-20, SA99 y TA, para un 55% de las configuraciones. En el caso de las SNB 51 y 52 la inmensa mayoría de las configuraciones logran superar la calidad del calendario oficial. Este resultado demuestra la conveniencia de enfocar la solución de este problema con metaheurísticas, más allá incluso de la metaheurística particular que se emplee. En el caso de la SNB 51 incluso ocurrió un incumplimiento de la restricción de que ningún equipo tuviera más de 4 enfrentamientos seguidos como visitador. Por tanto, para este caso, los calendarios encontrados en este trabajo superaran también al calendario oficial en este aspecto.

En general estos resultados muestran que las metaheurísticas constituyen una alternativa conveniente para reducir el costo de transportación en este tipo de problemas, incluso en los casos en que los expertos humanos hayan acumulado una notable experiencia resolviendo este tipo de problemas.

Tabla 10. Comparación de los resultados obtenidos frente a los calendarios oficiales

SN	KMT (O)	KMT (M)	Ahorro	%	CS	%
50	100391	87043	13348	13%	6	55%
51	176008	124893	51115	29%	10	91%
52	65281	54051	11230	17%	9	82%

Otro aspecto destacable es que la solución propuesta mejora el tiempo para obtener los calendarios, que no excede los 5 minutos en una computadora de prestaciones básicas (procesador Pentium, 1.73 GHz de velocidad, 1 Gb de memoria RAM). Este tiempo es ínfimo comparado con las dos o tres semanas que le lleva normalmente a los expertos la construcción de un calendario. Como referencia puede notarse que 5 minutos constituyen el 0.2% del tiempo de una semana de trabajo de 8 horas.

En el caso de la SNB 52 el calendario oficial de la primera etapa se construyó a partir de una versión preliminar de este trabajo, usando metaheurísticas. Ese hecho implica una aceptación por los expertos del valor de la propuesta. Adicionalmente, el tiempo corto para obtener los calendarios usando metaheurísticas tuvo un gran valor en este caso, debido al poco tiempo que medió entre que se decidió la estructura que iba a tener el calendario y la oficialización de este, que hubiera hecho muy compleja la elaboración totalmente manual en tan poco tiempo, o que hubiera implicado la obtención de un calendario peor. Posteriormente, se han construido calendarios usando los resultados presentados en este trabajo para la segunda etapa de la SNB 52.

Conclusiones

En este trabajo se ha demostrado cómo se puede hacer un uso más eficiente de los recursos necesarios para la organización de un evento deportivo complejo, como es la Serie Nacional de Béisbol. En muchas ocasiones se enfoca el uso mejor de los combustibles y en general de los recursos, desde la perspectiva de las tecnologías que hacen mejor empleo de los portadores energéticos para recorrer la misma cantidad de kilómetros. Sin embargo, se obvia otro componente en que puede trabajarse, que es la reducción de los kilómetros totales que se recorren.

Este trabajo enfoca este problema desde la perspectiva de la optimización usando metaheurísticas, lográndose soluciones que superan a las construidas por los expertos humanos permitiendo un ahorro entre el 13% y el 29% del total de kilómetros recorridos, obteniéndose además los calendarios en tiempos inferiores a los 5 minutos, respecto a las dos o tres semanas que debe emplear un experto humano para esto. Los resultados presentados en este trabajo han sido aplicados en la elaboración los calendarios que se han empleado de la SNB 52.

Agradecimientos

Los autores desean agradecer el entusiasta apoyo y la colaboración recibida de la Comisión Nacional de Béisbol para la realización de este trabajo, especialmente a los expertos Ybrahim Averoff y Carlos del Pino.

Referencias

1. Fabricio L. Biajoli, M. J. F. S., Antonio A. Chaves, *et al*. "Scheduling the Brazilian Soccer Championship. A Simulated Annealing Approach". En: 2004.
2. Nor Hayati Abdul Hamid, G. K. "Maximising Stadium Attendance: A Case Study of Malaysian Football ". En: 2008.
3. E.K. Burke, D. D. W., J.D. Landa Silva and C. Raess. "Applying Heuristic Methods to Schedule Sports Competitions on Multiple Venues". En: 2004.
4. Graham Kendall, L. W., Barry Mccollum y Frederico Cruz. "A Multiobjective Approach for UK Football Scheduling". En: 2008.
5. Frederick Ryckbosch, G. V. B., Graham Kendall. "A Heuristic Approach for the Traveling Tournament Problem using Optimal Traveling Salesman Tours". En: 2008.
6. Flavia Bonomo, A. B., Andres Cardemil, Guillermo Duran, Javier Marenco. "An application of the traveling tournament problem: the Argentine volleyball league". En: 2008.
7. Fabricio N. Costa, S. U., Celso C. Ribeiro. "An ILS heuristic for the traveling tournament problem with predefined venues". En: 2008.
8. Celso C. Ribeiro, S. U. "Heuristics for the Mirrored Traveling Tournament Problem". En: 2004.
9. Kelly Easton George Nemhauser, M. T. "The Traveling Tournament Problem Description and Benchmarks". 2001, nº DOI
10. Amotz Bar-Noy, D. M. "A Tiling Approach for Fast Implementation of Traveling Tournament Problem ". En: 2006.
11. Dries Goossens, F. C. R. S. "Scheduling the Belgian Soccer League". En: 2006.
12. Celso C. Ribeiro, S. U. "Scheduling the Brazilian Soccer Championship". En: 2006.
13. Shinji Imahori, T. M., Ryuhei Miyashiro "An Approximation Algorithm for the Unconstrained Traveling Tournament Problem ". En: 2010.
14. Graham Kendall, B. M., Frederico Cruz, Paul Mcmullan. "Scheduling English Football Fixtures: Consideration of Two Conflicting Objectives". En: *PATAT 2012*. 2010. ISBN Disponible en:
15. Shinji Imahori, T. M., Ryuhei Miyashiro. "An Approximation Algorithm for the Unconstrained Traveling Tournament Problem". En: *PATAT 2010*. 2010. ISBN Disponible en:
16. Jorn Schonberger, D. C. M., Herbert Kopfer. "Automated Timetable Generation for Rounds of a Table-Tennis League". 2002, nº DOI
17. Mieke Adriaen, N. C., Greet Vanden Berghe. "An Agent Based Metaheuristic for the Travelling Tournament Problem". 2004, nº DOI
18. Choubey, N. S. "A Novel Encoding Scheme for Traveling Tournament Problem using Genetic Algorithm". En: *Evolutionary Computation for Optimization Techniques 2010*. 2010.
19. Ayami Suzuka, R. M., Akiko Yoshise, Tomomi Matsui. "Semidefinite Programming Based Approaches to Home-awayAssignment Problems in Sports Scheduling". 2005, nº DOI
20. Fabricio Lacerda Biajoli, L. A. N. L. "Mirrored Traveling Tournament Problem: An Evolutionary Approach". 2004, nº DOI
21. Doerner, K. F.,Gendreau, M.,Greistorfer, P.*, et al. Metaheuristics: Progress in Complex Systems Optimization.* OPERATIONS RESEARCH/COMPUTER SCIENCE INTERFACES SERIES. Editado por: Sharda, R. y Voß, S. Springer Science+Business Media, LLC., 2007. p.
22. Wolpert, D. H. y Macready, W. G. "No Free Lunch Theorems for Optimization". *IEEE Transactions on Evolutionary Computation*. April 1997 1996, vol. 1, nº 1, p. 67-82. DOI
23. García, S.,Molina, D.,Lozano, M.*, et al.* "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study". *Journal of Heuristics*. 2009, vol. 15, nº 6, p. 617–644. DOI DOI 10.1007/s10732-008-9080-4
24. Talbi, E. G. *Metaheuristics: From Design to Implementation.* John Wiley & Sons, Inc., 2009. p.