

Evaluación híbrida asistida por modelo sustituto para un algoritmo genético aplicado al problema de transporte bajo demanda

Ricardo Pérez Cabrera, Rodolfo Eleazar Pérez Loaiza,
Perfecto Malaquías Quintero Flores, Edmundo Bonilla Huerta,
Paulina Galindo Garrido, Omar Atriano Venta

TecNM/Instituto Tecnológico de Apizaco, Apizaco, Tlax.,
México

m17370613@apizaco.tecnm.mx, rodolfo.pl@apizaco.tecnm.mx,
data.sci.phd@gmail.com, edmundo.bh@apizaco.tecnm.mx,
m19370503@apizaco.tecnm.mx, m25370001@apizaco.tecnm.mx

Resumen. El problema del transporte a demanda (DARP, por sus siglas en inglés) modela los servicios de transporte bajo demanda, donde una flota debe satisfacer las solicitudes de recogida y entrega sujetas a restricciones operativas. En aplicaciones reales, el coste computacional de un algoritmo genético (AG) depende en gran medida de la evaluación de cada individuo, especialmente cuando esta evaluación implica la reconstrucción temporal de rutas, tiempos de viaje basados en la red y tiempos de servicio heterogéneos. Este trabajo propone una estrategia de evaluación híbrida para un AG aplicado al DARP, en la que se utiliza un sustituto estructural basado en una matriz origen-destino (OD) para preevaluar la población y reducir la frecuencia de llamadas al evaluador real. La propuesta se implementa mediante una política basada en individuos en dos fases: una fase principal orientada al ahorro computacional y una fase de refinamiento con una mayor proporción de evaluaciones reales. La evaluación experimental se llevó a cabo en tres instancias de prueba, comparando el AG de referencia y la variante asistida por el sustituto mediante diez ejecuciones independientes por condición. Los resultados demuestran que la estrategia propuesta preserva la calidad de la solución, sin observarse diferencias estadísticamente significativas en la aptitud final, a la vez que reduce significativamente tanto el número de evaluaciones reales como el tiempo de ejecución. Estos hallazgos respaldan el uso de sustitutos estructurales simples e interpretables como un mecanismo eficaz para mejorar la eficiencia computacional de los algoritmos evolutivos en problemas de transporte con respuesta a la demanda y evaluaciones costosas.

Keywords: Dial-a-Ride problem, genetic algorithm, surrogate, origin-destination matrix, hybrid evaluation, expensive combinatorial optimization.

Hybrid Evaluation Assisted by a Surrogate Model for a Genetic Algorithm Applied to the On-demand Transport Problem

Abstract. The Dial-a-Ride Problem (DARP) models demand-responsive transport services in which a fleet must satisfy pickup and delivery requests under operational constraints. In realistic applications, the computational cost of a Genetic Algorithm (GA) depends heavily on the evaluation of each individual, especially when this evaluation involves temporal route reconstruction, network-based travel times, and heterogeneous service times. This work proposes a hybrid evaluation strategy for a GA applied to the DARP, in which a structural surrogate based on an origin-destination (OD) matrix is used to pre-evaluate the population and reduce the frequency of calls to the real evaluator. The proposal is implemented through an individual-based policy in two phases: a main phase oriented toward computational savings and a refinement phase with a higher proportion of real evaluations. The experimental assessment was conducted on three test instances, comparing the baseline GA and the surrogate-assisted variant through ten independent runs per condition. The results show that the proposed strategy preserves solution quality, with no statistically significant differences observed in final fitness, while significantly reducing both the number of real evaluations and execution time. These findings support the use of simple and interpretable structural surrogates as an effective mechanism for improving the computational efficiency of evolutionary algorithms in demand-responsive transport problems with expensive evaluation.

Keywords: Dial-a-Ride problem, genetic algorithm, surrogate, origin-destination matrix, hybrid evaluation, expensive combinatorial optimization.

1. Introducción

El problema de transporte bajo demanda (DARP, *Dial-a-Ride Problem*) constituye un marco de referencia para modelar servicios en los que una flota de vehículos debe atender solicitudes de recogida y entrega bajo restricciones de capacidad, precedencia, ventanas de tiempo y duración de ruta [1]. Este problema es especialmente relevante en contextos de transporte para personas con movilidad limitada, donde no solo importa la eficiencia operativa, sino también la calidad del servicio y el cumplimiento de condiciones temporales y logísticas asociadas a cada usuario [1,2].

Los algoritmos genéticos (GA, *Genetic Algorithm*) representan una alternativa atractiva para explorar el espacio de soluciones del DARP, debido a su capacidad para manejar estructuras combinatorias complejas y restricciones múltiples [3]. Sin embargo, en aplicaciones realistas su desempeño depende en

gran medida del costo computacional del procedimiento de evaluación. Cuando la función de aptitud (*fitness*) de cada individuo exige reconstruir rutas, verificar factibilidad temporal y calcular tiempos de servicio detallados, el proceso evolutivo puede volverse costoso incluso con tamaños de población y horizontes generacionales moderados.

Esta situación adquiere mayor relevancia en el sistema base considerado en este trabajo, donde la evaluación real integra tiempos de traslado sobre la red vial y tiempos de servicio estimados mediante un sistema de inferencia difusa (FIS, *Fuzzy Inference System*) para modelar el abordaje y descenso de pasajeros con movilidad limitada [4]. Aunque esta formulación incrementa el realismo operacional del modelo, también introduce un cuello de botella computacional que limita la eficiencia de la búsqueda evolutiva.

Para enfrentar este problema, el presente trabajo propone una estrategia de evaluación híbrida para un GA aplicado al DARP, en la que un modelo sustituto o aproximado (*surrogate*) estructural basado en una matriz origen-destino (OD, *Origin-Destination*) se utiliza para preevaluar la población y reducir la frecuencia de llamadas al evaluador real. La propuesta se implementa mediante una política basada en individuos (*individual-based*) en dos fases: una fase principal orientada al ahorro computacional y una fase de refinamiento con mayor proporción de evaluaciones reales [5,6,7].

La contribución principal del artículo consiste en mostrar que esta estrategia permite reducir de manera importante el número de evaluaciones reales y el tiempo de ejecución, sin deteriorar de forma significativa la calidad de las soluciones obtenidas. Para ello, se compara el sistema base con una variante asistida por modelo sustituto sobre varias instancias de prueba, considerando métricas de calidad, esfuerzo computacional y convergencia.

2. Antecedentes

La literatura ha abordado el DARP como un problema de transporte bajo demanda con restricciones múltiples. Jaw et al. [2] propusieron una heurística para el caso multi-vehículo con solicitudes anticipadas y ventanas de tiempo, incorporando criterios de calidad del servicio. Posteriormente, Cordeau y Laporte [1] consolidaron una formulación de referencia para el DARP estático multi-vehículo y mostraron la utilidad de enfoques heurísticos y metaheurísticos para tratar su complejidad computacional. Estos trabajos establecen la base estructural del problema sobre la que se construyen enfoques posteriores.

En el ámbito de las metaheurísticas, Jorgensen et al. [3] aplicaron un GA al DARP mediante un esquema de agrupar-primero, rutear-después (*cluster-first, route-second*), evidenciando que los métodos evolutivos pueden producir soluciones competitivas cuando se combinan con mecanismos adecuados de construcción y validación de rutas. Su aportación muestra que, en el DARP, el desempeño del GA depende no solo de los operadores evolutivos, sino también del procedimiento empleado para evaluar la calidad y factibilidad de los individuos.

Cuando la evaluación de la aptitud es costosa, la literatura ha propuesto el uso de *surrogates* o modelos aproximados como mecanismos para estimar respuestas de alto costo computacional mediante evaluaciones más baratas [8]. En el contexto del cómputo evolutivo, Jin [9] clasificó los principales niveles de aproximación y advirtió que sustituir de forma indiscriminada la función real puede inducir sesgos de búsqueda u óptimos falsos. Más adelante, Jin [5] formalizó el marco de la *surrogate-assisted evolutionary computation*, distinguiendo estrategias de gestión del modelo (*model management*) basadas en individuo, en generación y en población.

En el caso específico de la optimización combinatoria costosa, Liu et al. [7] señalaron que esta línea ha recibido menos atención que la optimización continua, a pesar de su importancia en problemas de ruteo, secuenciación y asignación. Por su parte, Hanawa et al. [6] analizaron el efecto de la precisión del *surrogate* sobre distintas estrategias de *model management* y mostraron que los enfoques basados en individuo son especialmente útiles cuando el modelo aproximado no es lo suficientemente preciso como para justificar una dependencia más agresiva.

A diferencia de enfoques basados en metamodelos de regresión o aprendizaje estadístico, el presente trabajo utiliza un modelo sustituto estructural basado en una matriz OD y una verificación barata de factibilidad. Esta elección responde a tres consideraciones del dominio: el espacio de búsqueda combinatorio del DARP no es un espacio vectorial continuo donde los metamodelos clásicos se desempeñen idealmente; la matriz OD aprovecha información estructural exacta dentro de su alcance, sin requerir entrenamiento estadístico adicional; y el enfoque resulta interpretable y libre de hiperparámetros, lo cual es relevante en sistemas operativos donde la trazabilidad del costo aproximado es importante. Esta decisión se relaciona además con el sistema base de Pérez Cabrera et al. [4], en el que un FIS incrementa el realismo operacional del evaluador pero también eleva su costo computacional, lo que justifica la necesidad de la estrategia híbrida propuesta.

3. Sistema base

El método propuesto se construye sobre un sistema base para el DARP en el que cada solución representa un conjunto de rutas asignadas a una flota de vehículos. El objetivo es atender solicitudes de recogida y entrega bajo restricciones de capacidad, precedencia y ventanas de tiempo, minimizando el costo total de operación [1,3].

El sistema base emplea un GA con operadores de selección, cruce, mutación y elitismo para explorar el espacio de soluciones. La calidad del proceso evolutivo depende no solo de los operadores aplicados, sino también del mecanismo utilizado para evaluar la factibilidad y el costo de cada individuo, aspecto crítico en aplicaciones del GA al DARP [3].

La evaluación real de una solución S se realiza mediante un procedimiento que reconstruye temporalmente cada ruta y verifica el cumplimiento de las

restricciones operativas. La aptitud real puede expresarse como

$$f(S) = C_{\text{op}}(S) + P(S), \quad (1)$$

donde $C_{\text{op}}(S)$ representa el costo operacional asociado a los tiempos de desplazamiento y servicio, y $P(S)$ agrupa las penalizaciones por violaciones de capacidad, precedencia y ventanas de tiempo.

Un componente distintivo del sistema base es la incorporación de un FIS para estimar tiempos de abordaje y descenso de pasajeros con movilidad limitada [4]. Esto implica que la evaluación real no solo considera tiempos de traslado sobre la red vial, sino también tiempos de servicio dependientes del perfil funcional del usuario. Por ello, la evaluación exacta resulta operativamente más realista pero también computacionalmente más costosa, lo que motiva la incorporación de la estrategia de evaluación híbrida descrita en la sección siguiente.

El FIS empleado por el sistema base se compone de cuatro subsistemas independientes, uno por cada tipo de herramienta de apoyo del usuario (sin herramienta, bastón, andadera y silla de ruedas). Cada subsistema toma como entrada una única variable lingüística, denominada Movilidad, definida en una escala de 1 a 10 que el observador asigna al usuario, y produce como salida el tiempo de servicio en parada (*dwelt time*) correspondiente al abordaje y descenso [4]. Los cuatro subsistemas siguen una estructura Mamdani–Singleton con cinco términos lingüísticos por universo, funciones de membresía gaussianas y una base de cinco reglas. El modelo fue calibrado a partir de 250 observaciones de campo recolectadas en hospitales de Apizaco, Tlaxcala, alcanzando coeficientes de determinación $R^2 \geq 0,86$ en validación. La incorporación del FIS al evaluador real permite representar la heterogeneidad de los tiempos de servicio entre usuarios con distinto perfil funcional, en lugar de asumir un tiempo de servicio constante. Su desventaja operativa es que la incorporación de tiempos de servicio heterogéneos durante la reconstrucción temporal de cada ruta incrementa el costo de la evaluación real, lo que motiva la incorporación del modelo sustituto descrita en la sección siguiente. Conviene aclarar que el FIS en sí solo se ejecuta una vez por cliente al inicializar el evaluador, no en cada generación del GA; la Sección 4.1 detalla esta arquitectura.

4. Metodología propuesta

4.1. Modelo sustituto basado en matriz origen-destino

El cuello de botella computacional del sistema base reside en la evaluación real de cada individuo, ya que esta requiere reconstruir temporalmente las rutas, consultar tiempos de traslado sobre la red vial y considerar tiempos de servicio del FIS. Para reducir este costo, se propone un modelo sustituto estructural basado en una matriz OD que almacena los tiempos de traslado entre los nodos relevantes de la instancia (depósito, puntos de recogida y destinos de entrega). La matriz se construye previamente y se reutiliza en evaluaciones posteriores, evitando repetir consultas costosas al modelo de red.

Sea S una solución compuesta por un conjunto de rutas $R = \{r_1, \dots, r_m\}$. La evaluación aproximada se define como

$$\hat{f}(S) = \sum_{r \in R} \sum_{(i,j) \in A(r)} t_{ij}^{\text{OD}} + \hat{P}(S), \quad (2)$$

donde $A(r)$ es el conjunto de arcos consecutivos de la ruta r , t_{ij}^{OD} es el tiempo precalculado entre los nodos i y j , y $\hat{P}(S)$ es una penalización de factibilidad de bajo costo. La evaluación aproximada es determinista para una misma solución: asigna siempre el mismo valor aproximado mientras no cambien los tiempos precalculados ni la penalización estructural. En consecuencia, el modelo sustituto no busca reemplazar el evaluador real, sino proporcionar una estimación suficientemente informativa para ordenar candidatos prometedores con un costo muy inferior.

Es importante notar que la matriz OD almacena exclusivamente tiempos de traslado entre nodos. Los tiempos de servicio del FIS no forman parte de la suma $\sum t_{ij}^{\text{OD}}$ de la aptitud aproximada, ya que dependen del usuario asignado a cada nodo y no de la secuencia de la ruta. Sin embargo, durante la inicialización del evaluador, el FIS se ejecuta una única vez por cliente para obtener su tiempo de servicio en recogida y entrega, valor que queda asociado al nodo correspondiente. La penalización barata $\hat{P}(S)$ utiliza estos tiempos precalculados únicamente para estimar de forma aproximada el tiempo de llegada a cada entrega y verificar la ventana correspondiente, sin volver a invocar al FIS. Esta arquitectura preserva el carácter económico del modelo sustituto: el FIS se ejecuta una sola vez al inicio y no durante el ciclo evolutivo, y a la vez la búsqueda no queda ciega al efecto de los tiempos de servicio sobre la factibilidad temporal.

4.2. Penalización barata de factibilidad

La evaluación aproximada incorpora un término $\hat{P}(S)$ que penaliza violaciones estructurales detectables sin ejecutar la simulación completa de la ruta. Se consideran: (i) nodos faltantes o duplicados, (ii) violaciones de precedencia, cuando un nodo de entrega aparece antes de su recogida correspondiente, (iii) exceso de capacidad, e (iv) incumplimientos aproximados de ventanas de tiempo estimados a partir del tiempo acumulado con la matriz OD. Esta penalización desincentiva individuos claramente inviables antes de asignarles una evaluación real costosa. La validez definitiva de una solución se establece exclusivamente mediante el evaluador real.

4.3. Estrategia híbrida de evaluación

Se adopta una estrategia de gestión del modelo sustituto de tipo *individual-based* [5], en la cual todos los individuos reciben primero una evaluación aproximada y solo un subconjunto es reevaluado mediante el evaluador real. La política de selección opera mediante dos componentes:

$$S_{\text{elite}} = \text{Top}_\alpha(P, \hat{f}), \quad (3)$$

formado por el $\alpha\%$ de individuos con mejor aptitud aproximada, y

$$S_{\text{aleat}} \subset P \setminus S_{\text{elite}}, \quad |S_{\text{aleat}}| = \beta|P|, \quad (4)$$

seleccionado de manera uniforme entre el resto de la población. El conjunto que recibe evaluación real en cada generación es entonces

$$S_{\text{real}} = S_{\text{elite}} \cup S_{\text{aleat}}. \quad (5)$$

Esta combinación equilibra intensificación y exploración: el componente élite dirige el esfuerzo hacia soluciones prometedoras, mientras que el componente aleatorio reduce el riesgo de sesgos sistemáticos del modelo sustituto y preserva diversidad [6].

4.4. Evaluación en dos fases

La gestión del modelo sustituto se organiza en dos fases. En la **fase principal**, correspondiente a la mayor parte del proceso evolutivo, se emplean $\alpha = 0,15$ y $\beta = 0,15$, de modo que aproximadamente el 30 % de la población recibe evaluación real en cada generación. Posteriormente, en el último 25 % de las generaciones, se activa una **fase de refinamiento** con $\alpha = 0,35$ y $\beta = 0,15$, por lo que alrededor del 50 % de la población recibe evaluación real. El proceso evolutivo favorece además a los individuos ya validados realmente durante selección y elitismo, lo que disminuye la probabilidad de que soluciones sobrevaloradas por el modelo sustituto dominen la evolución.

4.5. Pseudocódigo del método propuesto

La lógica general del método se resume en el Algoritmo 1 y se ilustra esquemáticamente en la Fig. 1. Ambos elementos son complementarios: el Algoritmo 1 formaliza el flujo de evaluación aproximada, selección del subconjunto S_{real} y transición de fases, mientras que la Fig. 1 resume gráficamente la diferencia operativa entre el GA base y la variante con modelo sustituto. La línea 1 garantiza que la población inicial sea siempre validada realmente; las líneas 5–7 realizan la preevaluación de toda la descendencia con la matriz OD; las líneas 8–12 controlan la transición a la fase de refinamiento en función del progreso generacional; y las líneas 13–19 efectúan la evaluación real selectiva sobre el subconjunto $S_{\text{real}} = S_{\text{elite}} \cup S_{\text{aleat}}$. Finalmente, la línea 20 asegura que el mejor histórico S^* siempre proviene de una evaluación real, evitando que soluciones sobrevaloradas por el modelo sustituto se conviertan en óptimo del proceso.

4.6. Análisis de complejidad

Sea $|S|$ el número total de nodos visitados por una solución. La evaluación aproximada requiere únicamente recorrer las rutas y sumar tiempos OD entre nodos consecutivos:

$$T_{\text{sust}}(S) = O(|S|). \quad (6)$$

Algoritmo 1 Esquema general del GA asistido por modelo sustituto para el DARP

Entrada: Población inicial P_0 , número de generaciones G , proporciones $\alpha_1, \beta_1, \alpha_2, \beta_2$

Salida: Mejor solución validada realmente S^*

```

1: Evaluar realmente todos los individuos de  $P_0$ 
2:  $S^* \leftarrow$  mejor individuo de  $P_0$ 
3: para  $g \leftarrow 1$  hasta  $G$  hacer
4:   Generar descendencia  $Q_g$  mediante selección, cruce y mutación
5:   para todo  $s \in Q_g$  sin evaluación real hacer
6:     Calcular  $\hat{f}(s)$  usando la matriz OD y la penalización barata
7:   fin para
8:   si  $g \leq 0,75 G$  entonces
9:      $\alpha \leftarrow \alpha_1; \beta \leftarrow \beta_1$  ▷ Fase principal
10:  si no
11:     $\alpha \leftarrow \alpha_2; \beta \leftarrow \beta_2$  ▷ Fase de refinamiento
12:  fin si
13:   $S_{\text{elite}} \leftarrow \text{Top}_\alpha(Q_g, \hat{f})$ 
14:   $S_{\text{aleat}} \leftarrow$  muestra aleatoria uniforme de  $Q_g \setminus S_{\text{elite}}$  con  $|S_{\text{aleat}}| = \beta|Q_g|$ 
15:   $S_{\text{real}} \leftarrow S_{\text{elite}} \cup S_{\text{aleat}}$ 
16:  para todo  $s \in S_{\text{real}}$  hacer
17:    Calcular  $f(s)$  con el evaluador real
18:  fin para
19:  Actualizar población mediante reemplazo y elitismo, priorizando individuos con
    evaluación real
20:  si existe  $s \in S_{\text{real}}$  con  $f(s) < f(S^*)$  entonces
21:     $S^* \leftarrow s$ 
22:  fin si
23: fin para
24: devolver  $S^*$ 

```

En contraste, la evaluación real implica reconstruir la secuencia temporal de cada ruta, verificar precedencia, capacidad y ventanas de tiempo, e incorporar tiempos de traslado y servicio mediante la red vial y el FIS:

$$T_{\text{real}}(S) = O(|S| + Q_{\text{red}}(S) + Q_{\text{serv}}(S)) \gg T_{\text{sust}}(S), \quad (7)$$

donde $Q_{\text{red}}(S)$ y $Q_{\text{serv}}(S)$ son los costos de cálculo de tiempos sobre la red y de validación temporal con tiempos de servicio heterogéneos, respectivamente. Si la población tiene tamaño N y el algoritmo evoluciona durante G generaciones, el número total de evaluaciones reales se aproxima por

$$E_{\text{real}} \approx \sum_{g=1}^G \rho_g N, \quad (8)$$

donde $\rho_g \approx 0,30$ en la fase principal y $\rho_g \approx 0,50$ en la fase de refinamiento. La propuesta reduce sustancialmente el número de evaluaciones costosas sin eliminar la retroalimentación de la aptitud real.

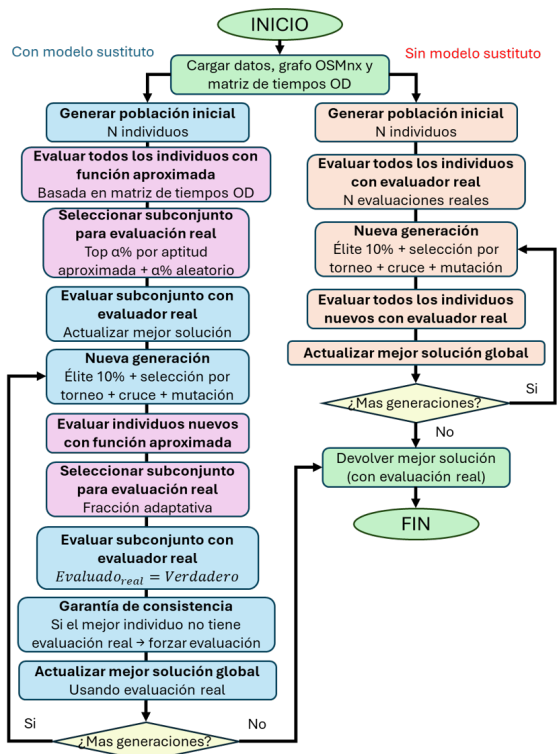


Fig. 1. Comparación esquemática entre el GA base y la variante asistida por modelo sustituto: preevaluación con matriz OD, selección parcial para evaluación real y fase de refinamiento.

5. Diseño experimental

La evaluación experimental se planteó como una comparación controlada entre dos condiciones: GA sin modelo sustituto y GA con modelo sustituto. Se utilizaron tres instancias de prueba: I1 con 30 solicitudes y 3 destinos, I2 con 40 solicitudes y 4 destinos, e I3 con 50 solicitudes y 5 destinos. Para cada instancia se realizaron 10 corridas independientes por condición, generando un total de 60 ejecuciones.

Ambas variantes mantuvieron la misma configuración general del GA: población de 200 individuos, 2000 generaciones, tasa de cruce de 0.6, tasa de mutación de 0.1, torneo de tamaño 4 y elitismo del 10 %. La única diferencia entre condiciones fue la estrategia de evaluación. En el GA sin modelo sustituto, toda la población recibió evaluación real durante todo el proceso evolutivo. En la variante con modelo sustituto, durante la fase principal se evaluó realmente aproximadamente el 30 % de la población ($\alpha = 0,15, \beta = 0,15$), mientras que en el último 25 % de las generaciones se activó la fase de refinamiento, en la que la fracción evaluada realmente aumentó a alrededor del 50 % ($\alpha = 0,35, \beta = 0,15$).

Tabla 1. Promedios de desempeño por instancia y método.

Instancia	Método	Aptitud promedio	Eval. reales promedio	Tiempo promedio (s)
I1	Sin mod. sust.	571.13	376 981.8	493.27
I1	Con mod. sust.	582.66	140 340.0	257.70
I2	Sin mod. sust.	824.60	311 180.1	529.46
I2	Con mod. sust.	809.45	140 112.9	376.93
I3	Sin mod. sust.	1 322.61	302 532.4	627.95
I3	Con mod. sust.	1 298.65	140 079.5	473.19

Tabla 2. Resultados de la prueba de Wilcoxon por instancia.

Instancia	p -valor aptitud	p -valor eval. reales	p -valor tiempo
I1	0.492	0.002	0.002
I2	0.492	0.002	0.002
I3	0.375	0.002	0.002

Las métricas analizadas fueron aptitud final, número de evaluaciones reales, número de evaluaciones aproximadas y tiempo total de ejecución. Además, se analizó la convergencia a partir de la mejor aptitud acumulada y de la aptitud promedio generacional. Para el contraste estadístico se empleó la prueba de rangos con signo de Wilcoxon, adecuada para comparar dos condiciones con tamaño de muestra pequeño sin asumir normalidad.

6. Resultados y discusión

La comparación mostró un patrón consistente en las tres instancias: la estrategia propuesta preservó la calidad de solución y redujo de manera importante tanto el número de evaluaciones reales como el tiempo de ejecución. La Tabla 1 resume los valores promedio para ambas configuraciones y la Tabla 2 presenta los resultados de la prueba de Wilcoxon.

Antes de discutir las pruebas estadísticas conviene interpretar el patrón observado en la Tabla 1. El número de evaluaciones reales del GA con modelo sustituto se estabiliza cerca de 140 000 en las tres instancias, comportamiento consistente con la política de evaluación propuesta: con población de 200, 2 000 generaciones y fracciones de evaluación real del 30 % en fase principal y 50 % en refinamiento, el total esperado de evaluaciones reales depende de los parámetros α y β y del número de generaciones, no del tamaño de la instancia. Esta independencia es una propiedad estructural de la estrategia y permite anticipar el costo de evaluación a priori.

En contraste, en el GA sin modelo sustituto el número de evaluaciones reales se mantiene también relativamente estable (entre 300 000 y 380 000), pero el costo unitario de cada evaluación crece con $|S|$, ya que la reconstrucción temporal de cada ruta debe procesar más nodos, más arcos y más tiempos de servicio. Esto

explica que el tiempo total aumente con el tamaño de la instancia y, sobre todo, que la reducción relativa de tiempo lograda por el modelo sustituto disminuya de 47.8 % en I1 a 28.8 % en I2 y 24.6 % en I3: en instancias pequeñas, el sobrecosto fijo del evaluador real domina y eliminarlo en parte de la población produce un ahorro porcentualmente mayor; en instancias grandes, el costo se distribuye entre más nodos y la ventaja relativa, aunque sigue siendo significativa, se atenúa. Este comportamiento es coherente con el hecho de que el sobrecosto fijo del evaluador real se distribuye entre más nodos en instancias grandes; la generalización a instancias de mayor escala requiere validación adicional con un banco de pruebas más amplio.

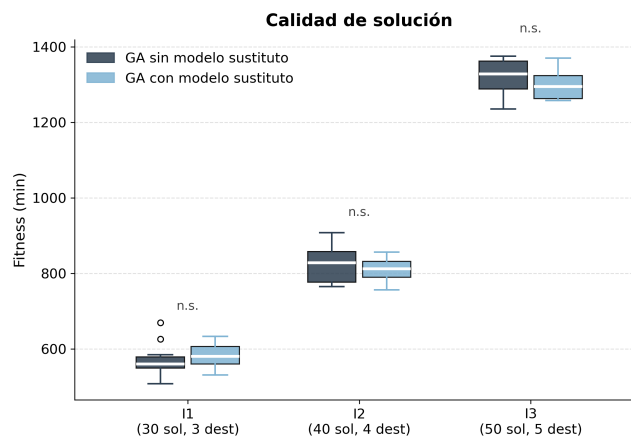


Fig. 2. Distribución de la aptitud final para el GA sin modelo sustituto y la variante con modelo sustituto en las tres instancias analizadas.

En términos de calidad de solución, ambas configuraciones produjeron resultados comparables. Los promedios de aptitud difieren ligeramente entre condiciones, pero la Tabla 2 indica que dichas diferencias no fueron estadísticamente significativas en ninguna instancia. La Figura 2 confirma que las distribuciones de aptitud permanecen próximas entre ambas variantes.

La ausencia de diferencia significativa en aptitud, junto con la reducción altamente significativa en evaluaciones reales y tiempo, sugiere que el modelo sustituto proporciona una preselección suficientemente útil para conservar soluciones competitivas, aunque su valor aproximado no replica la aptitud real exacta. La caracterización formal de la correlación de rango entre \hat{f} y f queda como trabajo futuro. Esto es coherente con la lógica de la estrategia híbrida: para seleccionar candidatos prometedores no se requiere precisión absoluta del modelo aproximado, sino que su ordenamiento aproximado mantenga utilidad práctica en la región de soluciones de alta calidad. La fase de refinamiento corrige cualquier sesgo residual al final del proceso evolutivo, lo que se refleja en que el mejor histórico S^* siempre proviene de una evaluación real.

Un hallazgo que merece atención es que en I2 e I3 el GA con modelo sustituto alcanzó en promedio una aptitud ligeramente mejor que el GA base (809,45

frente a 824,60 y 1 298,65 frente a 1 322,61, respectivamente). Aunque la prueba de Wilcoxon indica que esta diferencia no es estadísticamente significativa, el patrón sugiere que el componente aleatorio S_{aleat} del 15 % no actúa únicamente como salvaguarda contra sesgos del modelo sustituto: también puede contribuir a preservar diversidad en regiones del espacio de búsqueda que el GA base explora con menor intensidad, ya que su selección concentra esfuerzo en los mismos individuos durante muchas generaciones. La evidencia disponible no permite afirmar de manera concluyente este mecanismo, pero el patrón observado es consistente con resultados reportados en la literatura sobre el papel de la diversidad en algoritmos evolutivos asistidos por modelos sustitutos [6].

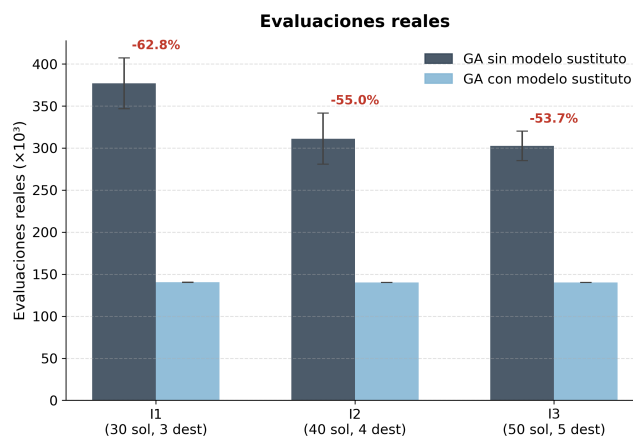


Fig. 3. Número promedio de evaluaciones reales realizadas por el GA sin modelo sustituto y la variante con modelo sustituto. Los porcentajes indican la reducción relativa obtenida con la propuesta.

El efecto más claro de la propuesta se observa en el número de evaluaciones reales. El modelo sustituto redujo este indicador de forma sustancial: de 376 981.8 a 140 340.0 en I1, de 311 180.1 a 140 112.9 en I2 y de 302 532.4 a 140 079.5 en I3, con diferencias altamente significativas ($p = 0,002$). La Figura 3 confirma que esta reducción es estable y constituye el resultado más robusto del estudio. Dado que el evaluador real integra reconstrucción temporal, verificación de restricciones, tiempos sobre red y tiempos de servicio del FIS, la disminución de evaluaciones reales implica una reducción directa del componente más costoso del proceso evolutivo.

En cuanto al tiempo de ejecución, la reducción también fue consistente: de 493.27 s a 257.70 s en I1, de 529.46 s a 376.93 s en I2 y de 627.95 s a 473.19 s en I3, con diferencias significativas en las tres instancias ($p = 0,002$). La Figura 4 muestra que la reducción del esfuerzo de evaluación sí se tradujo en una disminución observable del tiempo total de ejecución.

Las Figuras 5, 6 y 7 muestran la dinámica de convergencia de ambas configuraciones para cada instancia. En general, ambas variantes siguen trayectorias semejantes: mejora rápida en las etapas iniciales y evolución más gradual con-

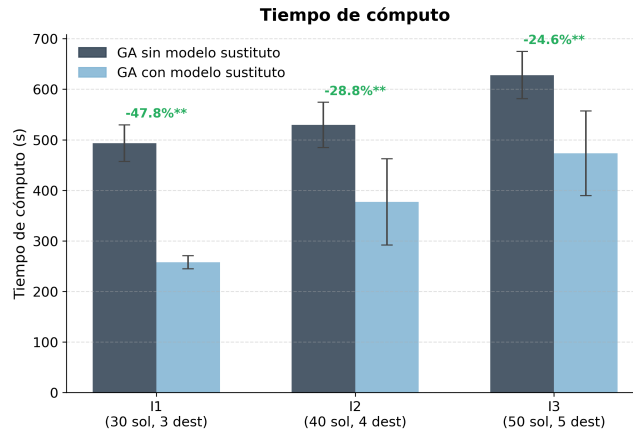


Fig. 4. Tiempo de cómputo promedio del GA sin modelo sustituto y la variante con modelo sustituto. Los porcentajes indican la reducción relativa obtenida con la propuesta.

forme la búsqueda se concentra en regiones de mayor calidad. La transición a la fase de refinamiento en la generación 1 500 no introduce inestabilidades visibles, sino una intensificación progresiva de la evaluación real, lo que indica que la estrategia híbrida no altera de forma drástica la lógica evolutiva del sistema base.

En conjunto, los resultados permiten sostener tres afirmaciones principales. Primero, la estrategia propuesta preserva la calidad de solución, ya que no se detectaron diferencias significativas en aptitud. Segundo, reduce de manera consistente y significativa el número de evaluaciones reales. Tercero, reduce también de manera significativa el tiempo de ejecución en las tres instancias consideradas.

7. Conclusiones

En este trabajo se presentó una estrategia de evaluación híbrida para un GA aplicado al DARP, en la que un modelo sustituto estructural basado en una matriz OD se integra con el evaluador real mediante una política basada en individuos en dos fases. La propuesta no sustituye el sistema base ni la aptitud real, sino que reduce el costo computacional asociado a la evaluación de individuos manteniendo la confiabilidad del proceso evolutivo.

Los resultados mostraron que la estrategia redujo de manera sustancial y consistente el número de evaluaciones reales en las tres instancias analizadas, sin deteriorar de forma estadísticamente significativa la calidad final de las soluciones. Además, esta reducción se tradujo en una disminución significativa del tiempo de ejecución en todas las instancias. El análisis de convergencia confirmó que la variante asistida por modelo sustituto conserva trayectorias evolutivas comparables a las del sistema base. La combinación entre fase principal y fase

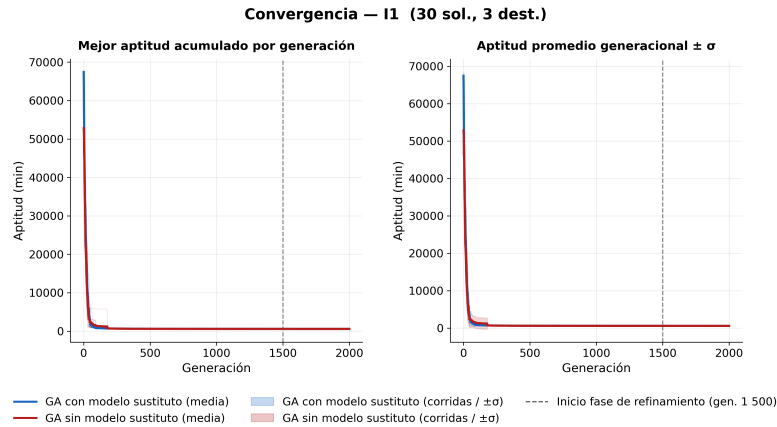


Fig. 5. Convergencia para I1: mejor aptitud acumulada (izq.) y aptitud promedio $\pm\sigma$ (der.) por generación. La línea punteada indica el inicio de la fase de refinamiento.

de refinamiento resultó adecuada para equilibrar exploración, intensificación y control del error de aproximación.

Como limitaciones del trabajo deben señalarse cuatro aspectos. Primero, los resultados deben interpretarse como una validación inicial sobre instancias representativas del sistema operativo bajo estudio (30, 40 y 50 solicitudes con 3 a 5 destinos), no como una demostración general de escalabilidad para instancias de mayor tamaño. Segundo, la ventaja observada depende de la capacidad de la matriz OD y de la penalización estructural para representar de manera razonable el costo real de la solución; en escenarios con mayor dinamismo o fuentes de variabilidad no reflejadas en dicha estructura, la precisión del modelo sustituto podría deteriorarse. Tercero, el estudio no incluye una caracterización formal de la correlación de rango entre \hat{f} y f , métrica que sería deseable en un análisis más profundo del comportamiento del modelo sustituto. Cuarto, la transición entre fase principal y fase de refinamiento se realiza de forma prefijada al 75 % de las generaciones, sin un mecanismo adaptativo que ajuste dinámicamente las proporciones α y β .

En consecuencia, como trabajo futuro se proponen tres líneas: (i) caracterizar la correlación de rango entre el modelo sustituto y el evaluador real mediante coeficientes de Spearman o Kendall sobre parejas (\hat{f}, f) obtenidas en corridas dedicadas; (ii) explorar variantes adaptativas de la política de evaluación híbrida que ajusten α y β en función del progreso evolutivo o de la calidad estimada del modelo sustituto; y (iii) extender la validación a instancias más grandes y a escenarios dinámicos que permitan evaluar la escalabilidad y robustez del enfoque.

En síntesis, los resultados respaldan la hipótesis central del artículo: es posible reducir de forma importante el esfuerzo de evaluación y el tiempo de ejecución en un GA para el DARP mediante un modelo sustituto OD gestionado de forma híbrida, sin afectar de manera significativa la calidad de las soluciones.

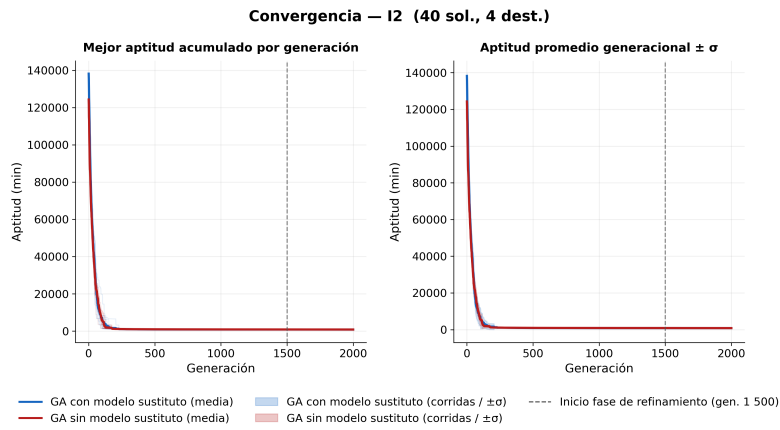


Fig. 6. Convergencia para I2: mejor aptitud acumulada (izq.) y aptitud promedio $\pm\sigma$ (der.) por generación. La línea punteada indica el inicio de la fase de refinamiento.

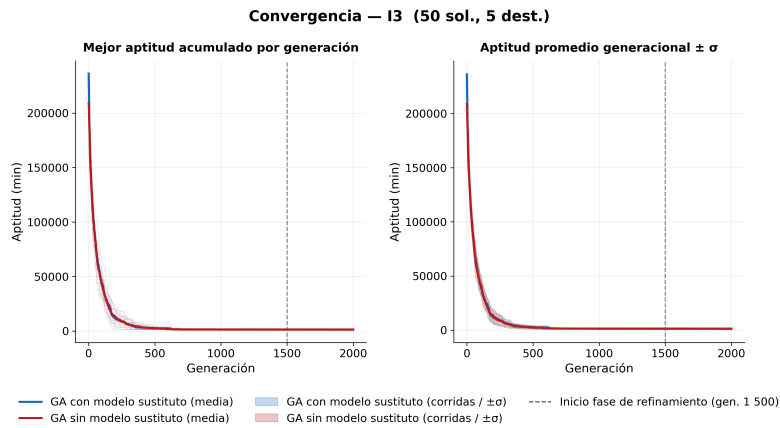


Fig. 7. Convergencia para I3: mejor aptitud acumulada (izq.) y aptitud promedio $\pm\sigma$ (der.) por generación. La línea punteada indica el inicio de la fase de refinamiento.

Referencias

1. Cordeau, J.-F., Laporte, G.: A tabu search heuristic for the static multi-vehicle dial-a-ride problem. *Transportation Research Part B: Methodological* 37(6), 579–594 (2003)
2. Jaw, J.-J., Odoni, A. R., Psaraftis, H. N., Wilson, N. H. M.: A heuristic algorithm for the multi-vehicle advance request dial-a-ride problem with time windows. *Transportation Research Part B: Methodological* 20(3), 243–257 (1986)
3. Jørgensen, R. M., Larsen, J., Bergvinsdottir, K. B.: Solving the Dial-a-Ride Problem using Genetic Algorithms. *Journal of the Operational Research Society* 58(10), 1321–1331 (2007)
4. Pérez Cabrera, R., Pérez Loaiza, R. E., Quintero Flores, P. M.: Fuzzy Inference Modeling for Estimation of Boarding and Alighting Times in Passengers with Limited Mobility. *Tecnología y Ciencia Aplicada* 8(2), 143–150 (2025)
5. Jin, Y.: Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm and Evolutionary Computation* 1(2), 61–70 (2011)
6. Hanawa, Y., Harada, T., Miura, Y.: Impact of surrogate model accuracy on performance and model management strategy in surrogate-assisted evolutionary algorithms. *Array* 27, 100461 (2025)
7. Liu, S., Wang, H., Peng, W., Yao, W.: Surrogate-assisted evolutionary algorithms for expensive combinatorial optimization: A survey. *Complex & Intelligent Systems* 10, 5933–5949 (2024)
8. Forrester, A. I. J., Sobester, A., Keane, A. J.: *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons (2008)
9. Jin, Y.: A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing* 9(1), 3–12 (2005)