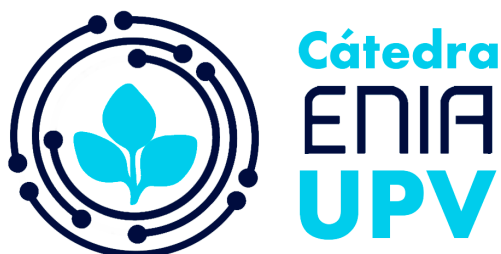


Informe de Prácticas Extracurriculares

Grado en Ciencia de Datos



Juan Carlos Trujillo Rodríguez

2026-2027

Tutora UPV: Paula Bastida Molina
Tutor Empresa: Víctor Sánchez Anguix



Contents

1 Descripción concreta y detallada de las tareas, trabajos desarrollados y departamentos de la entidad a los que ha estado asignado	2
2 Valoración de las tareas desarrolladas con los conocimientos y competencias adquiridas en relación con los estudios universitarios	2
3 Relación de los problemas planteados y de los procedimientos seguidos para su resolución	3
4 Identificación de las aportaciones que, en materia de aprendizaje han supuesto las prácticas	6
5 Evaluación de las prácticas y sugerencias de mejora	7

1 Descripción concreta y detallada de las tareas, trabajos desarrollados y departamentos de la entidad a los que ha estado asignado

Las prácticas se han desarrollado en el marco de la Cátedra ENIA-UPV, dentro de su línea de investigación en Movilidad Eléctrica Sostenible, y dan continuidad al proyecto iniciado en las prácticas anteriores, cuyo objetivo es facilitar la transición hacia el vehículo eléctrico (EV) en España mediante herramientas de soporte a la decisión para investigadores y administraciones públicas. En aquella primera etapa diseñé un simulador basado en agentes (ABM) sobre grafos urbanos extraídos de OpenStreetMap, capaz de evaluar cualquier red de recarga a partir del comportamiento individual de los conductores y de resumir su rendimiento en tres métricas que capturan la insatisfacción del usuario (F_1), el desuso de las estaciones (F_2) y el coste de la infraestructura (F_3). Partiendo de ese simulador, las tareas se han centrado ahora en construir la capa que decide qué redes conviene desplegar, pasando así de medir escenarios a optimizarlos.

La tarea central desarrollada consiste en formular la localización y el dimensionado de la infraestructura como un problema de optimización multiobjetivo y resolverlo con el algoritmo evolutivo NSGA-II, sobre la librería *Metaheuristics.jl* de Julia. El optimizador y el simulador trabajan por separado y se comunican mediante un vector de funciones objetivo, de modo que el algoritmo propone redes de recarga y el simulador le devuelve la calidad de cada una.

Para construir el sistema, el trabajo se organizó en varias tareas. En primer lugar, una capa de agregación que transforma las métricas locales del simulador en tres objetivos globales comparables por el algoritmo, la insatisfacción global de los conductores (J_1), el desuso ponderado de las estaciones (J_2) y el coste total de la infraestructura (J_3), que se minimizan de forma simultánea para obtener un frente de Pareto de configuraciones no dominadas. A continuación, una heurística que reduce las miles de ubicaciones posibles del mapa a un conjunto manejable de candidatas. Después, una representación genética que codifica cada solución como un individuo del algoritmo. Y por último, el control de las restricciones del problema y una evaluación robusta que tiene en cuenta el carácter estocástico del simulador. Cada una de estas tareas planteó sus propios problemas, que se detallan más adelante.

2 Valoración de las tareas desarrolladas con los conocimientos y competencias adquiridas en relación con los estudios universitarios

Durante las prácticas he puesto en práctica competencias del Grado en Ciencia de Datos que en la etapa anterior apenas había utilizado, sobre todo las relacionadas con la optimización y el aprendizaje automático. Implementar

un algoritmo evolutivo multiobjetivo como NSGA-II supone aplicar conceptos teóricos de optimización como la dominancia entre soluciones o los operadores de cruce y mutación.

La formulación del problema se ha apoyado además en la teoría de grafos, utilizando la centralidad de intermediación para construir soluciones, y en la estadística, necesaria para evaluar de forma robusta el simulador. Todo ello se ha programado en Julia, cuidando la eficiencia y organizando la lógica del optimizador en un módulo independiente. De este modo, teoría de grafos, programación y optimización se han combinado en un mismo desarrollo.

Además, una gran parte del trabajo ha estado orientada al análisis de datos para interpretar correctamente los resultados. La Figura 1 muestra un ejemplo de los resultados obtenidos por una estrategia, donde se observa cómo evolucionan las proporciones de tipos de cargadores a lo largo de las generaciones, y cómo la búsqueda tiende a descartar los cargadores rápidos más caros para reducir el coste.

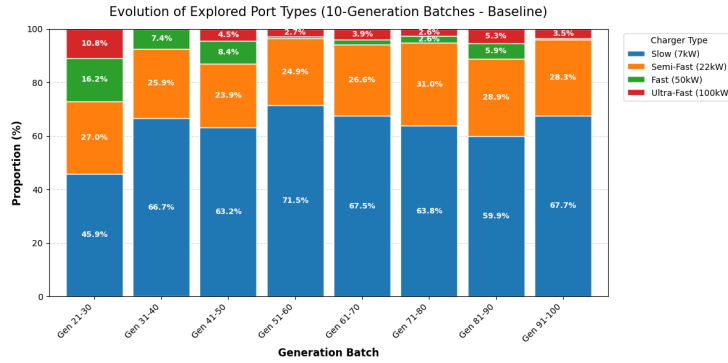


Figure 1: Evolución de los tipos de puerto explorados durante las generaciones.

En general, en estas prácticas hemos cerrado el ciclo completo de un proyecto de optimización, desde la formulación matemática del problema hasta la interpretación de sus soluciones, pasando por la implementación del algoritmo y su comunicación con un simulador basado en agentes. Este tipo de trabajo recoge problemas complejos y multidisciplinares que se abordan a lo largo de la carrera.

3 Relación de los problemas planteados y de los procedimientos seguidos para su resolución

El problema principal fue acotar un espacio de búsqueda combinatorio, en el que hay que decidir tanto la ubicación como el dimensionado de las estaciones, evaluando cada candidato con un simulador estocástico y computacionalmente

costoso. Para acotar ese espacio, diseñé una heurística de selección de ubicaciones candidatas que reduce las miles de posiciones del mapa a unas pocas decenas, quedándose con las cercanas a las vías principales del grafo, ordenándolas por su centralidad en la red y exigiendo una distancia mínima de separación entre ellas. Este conjunto incorpora además las estaciones ya existentes del registro de la DGT, de manera que la búsqueda trabaja solo sobre ubicaciones razonables.

Una dificultad añadida fue codificar decisiones discretas, como el número y la potencia de los puertos, con operadores genéticos continuos. Cada individuo se representa como un cromosoma continuo que se traduce en una configuración discreta de puertos de distinta potencia, cada nivel con su tarifa asociada, según la decodificación que ilustra la Figura 2. Opté por esta codificación real asumiendo como contrapartida que algunas operaciones genéticas resultan silenciosas y no modifican la red decodificada, una ineficiencia que quedó anotada como trabajo futuro.

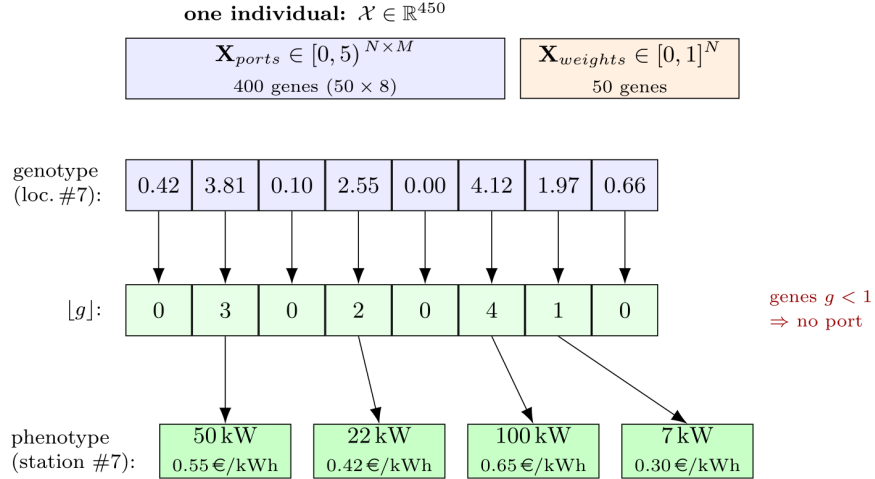


Figure 2: Decodificación del cromosoma de un individuo para una ubicación candidata.

A ello se sumaba que los operadores genéticos no respetan por sí mismos las restricciones duras y generan redes que superan el límite máximo de estaciones activas, lo que obliga a podar las configuraciones no factibles conservando las ubicaciones de mayor prioridad. Para resolverlo desarrollé y comparé varias estrategias de factibilidad. Las dos principales, **Baseline** y **Fixed Crossover**, mantienen factible a la descendencia durante la reproducción y se diferencian en el origen de la prioridad de cada ubicación. La primera coevoluciona esas prioridades junto con la red, dejando que la búsqueda descubra qué ubicaciones

importan, mientras que la segunda las fija para proteger desde el inicio las localizaciones más centrales. Junto a estas implementé también operadores de reparación (**Random** y **Deterministic**), que restauran la factibilidad una vez generada la descendencia. Aunque en el trabajo de análisis final los reparadores no fueron incluidos por falta de tiempo, sirvieron de contraste, ya que en pruebas preliminares sus soluciones no llegaban a aprovechar del todo las ventajas de NSGA-II. Disponer de varias estrategias comparables permitió comprobar si construir la factibilidad dentro de la reproducción compensaba frente a repararla después.

El carácter estocástico y el coste del simulador condicionaron el procedimiento de evaluación. Cada individuo se evaluó con varias réplicas independientes agregadas por la mediana, con semillas derivadas de la huella del cromosoma para asegurar la reproducibilidad, y una caché de duplicados que llegó a ahorrar entre un 16% y un 19% de las llamadas al simulador en los experimentos finales. Aun así, cada estrategia necesitó del orden de 80.000 segundos para completar 100 generaciones, lo que obligó a mantener poblaciones pequeñas y a apoyarse en el paralelismo del servidor. Para comparar las estrategias de forma objetiva recurrí a varios procedimientos estadísticos, analizando la convergencia mediante el hipervolumen, contrastando las diferencias con tests no paramétricos de Mann-Whitney U y estudiando la evolución generacional de las medianas de cada objetivo, junto con un análisis de sensibilidad al factor de penalización λ_{crit} . En la Figura 3, mediante comparación de hipervolumenes, se aprecia que **Baseline** llega a explorar un espacio de soluciones más amplio, mientras que **Fixed Crossover**, al proteger desde el inicio las ubicaciones más centrales, se mantiene más defensiva con la insatisfacción (J_1) y cubre un rango de compromisos más limitado.

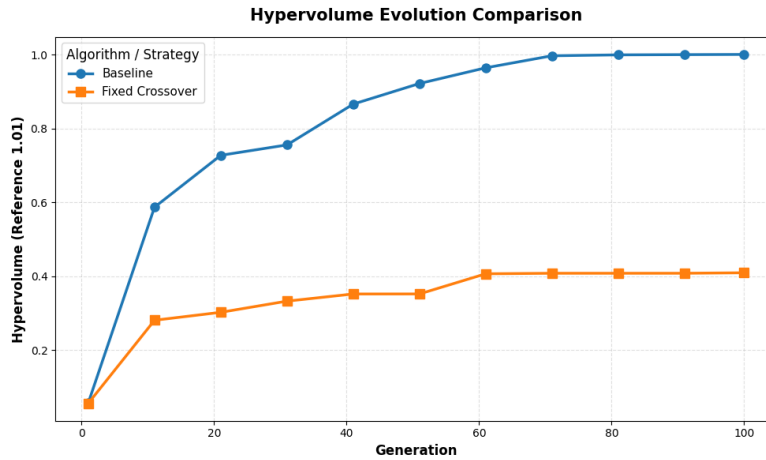


Figure 3: Comparación de la evolución del hipervolumen entre las estrategias **Baseline** y **Fixed Crossover**.

4 Identificación de las aportaciones que, en materia de aprendizaje han supuesto las prácticas

De todo lo aprendido en estas prácticas, destaco varias aportaciones para mi formación.

- La implementación práctica de un algoritmo de optimización evolutiva multiobjetivo, desde el diseño de las soluciones hasta el tratamiento de sus restricciones.
- El acoplamiento de un optimizador con un simulador costoso y estocástico, y las técnicas necesarias para evaluarlo de forma robusta y reproducible.
- La interpretación de un frente de Pareto como herramienta de soporte a la decisión, del que extraer escenarios de diseño según la prioridad del planificador.
- El diseño y la ejecución de un análisis experimental riguroso para comparar y validar las soluciones obtenidas.

Ha sido de especial interés analizar el frente de Pareto. De él se obtuvieron escenarios con características específicas, que favorecen objetivos concretos o buscan un equilibrio entre ellos. La Figura 4 muestra el escenario más balanceado obtenido, el punto de rodilla, en el que una única estación de alta potencia logra reducir y estabilizar la insatisfacción del usuario (J_1), mientras el desuso (J_2) y el coste (J_3) se mantienen bajos. Poder interpretar los resultados con este detalle y comprobar que las soluciones siguen una lógica coherente ha sido una de las partes más valiosas del trabajo, y respalda que el comportamiento del sistema es realista.

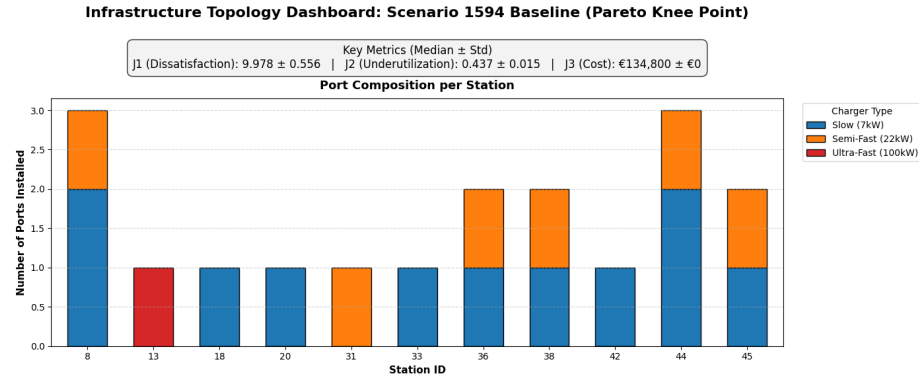


Figure 4: Ejemplo de configuración extraída del frente de Pareto, solución de compromiso o punto de rodilla.

5 Evaluación de las prácticas y sugerencias de mejora

Mi experiencia en la Cátedra ENIA-UPV ha vuelto a ser muy positiva, tanto en lo personal como en lo profesional. Poder completar el proyecto que había iniciado, pasando de un simulador que evalúa redes de recarga a un optimizador que las diseña, ha resultado especialmente satisfactorio, ya que me ha permitido ver el ciclo completo de una herramienta de soporte a la decisión aplicada a un problema real de movilidad y sostenibilidad. Este trabajo ha constituido, además, el núcleo de mi Trabajo de Fin de Grado.

El trato recibido por parte de mis tutores y compañeros ha sido de nuevo excelente, y valoro especialmente la autonomía que se me ha concedido para proponer y evaluar mis propias soluciones, como el desarrollo y la comparación de las distintas estrategias de factibilidad. Esa libertad ha reforzado mi sentido de la responsabilidad y mi exigencia con la calidad del resultado.

En cuanto a posibles mejoras, más orientadas a la evolución técnica del proyecto que a la organización de las prácticas, quedan abiertas varias líneas naturales de continuación. Sería conveniente adoptar operadores genéticos enteros que eliminen las operaciones silenciosas de la codificación actual, explorar modelos sustitutos que aproximen las funciones objetivo para reducir el elevado coste del simulador, e incorporar un modelo base no evolutivo, como una búsqueda aleatoria o una colocación voraz por centralidad, que permita aislar cuánto del rendimiento se debe realmente a NSGA-II. Como balance final, considero que estas prácticas han supuesto una aportación decisiva en mi formación, y me gustaría seguir colaborando en proyectos de investigación de esta naturaleza en el futuro.