

MÁSTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

ESTIMACIÓN DE MATRICES ORIGEN-DESTINO A PARTIR DE DATOS DE FLUJO VEHICULAR

Estudiante: Hurtado, Villasante, Borja

Director/a: Del Ser, Lorente, Javier

Departamento: Ingeniería de Comunicaciones

Curso académico: 2022-2023

Bilbao, 3 de Marzo de 2023

Resumen

Las matrices origen-destino son una herramienta fundamental dentro del ámbito de la movilidad urbana, ya que describen los patrones de desplazamiento de peatones, transporte público o vehículos a lo largo de un área geográfica. Contar con dicha información permite tomar decisiones más acertadas a la hora de realizar el dimensionamiento de infraestructuras de transporte, la planificación del transporte público o la planificación urbana. Dada su importancia, se han realizado una gran cantidad de estudios acerca de posibles métodos de inferencia en los últimos 50 años. Los primeros planteamientos del problema, debido a la naturaleza de los datos, se encontraban ante un problema altamente subdeterminado, lo cual exigía de la incorporación de información adicional al problema. Gran parte de estos estudios eran únicamente teóricos, o se implementaban en escenarios de tamaño muy reducido. Con el paso de los años, y la aparición de nuevos tipos de datos, la variedad de métodos de inferencia se ha ampliado considerablemente, así como su aplicabilidad a escenarios reales. Sin embargo, estos nuevos tipos de datos también tienen sus inconvenientes y el acceso a los mismos es limitado, por lo que la estimación de matrices origen-destino sigue siendo un problema abierto para muchas empresas del ámbito de la movilidad. En este proyecto, en colaboración con una empresa del ámbito de la movilidad, se teoriza e implementa un método de estimación novedoso a partir de datos convencionales, y se evalúa su funcionamiento en un escenario real.

Abstract

Origin-destination matrices are a fundamental tool in the field of urban mobility, since they describe the movement patterns of pedestrians, public transport or vehicles throughout a geographical area. Having this information makes it possible to make better decisions when sizing transport infrastructures, planning public transport or urban planning. Given its importance, a large number of studies on possible inference methods have been carried out over the last 50 years. Early approaches to the problem, due to the nature of the data, were highly underdetermined, requiring the incorporation of additional information into the problem. Most of these studies were only theoretical, or were implemented in very small scenarios. Over the years, and with the emergence of new types of data, the variety of inference methods has expanded considerably, as has their applicability to real scenarios. However, these new types of data also have their drawbacks and access to them is limited, so the estimation of origin-destination matrices remains an open problem for many companies in the mobility field. In this project, in collaboration with a company in the mobility field, a novel estimation method is theorized and implemented from conventional data, and its performance is evaluated in a real scenario.

Laburpena

Jatorri-helmuga matrizeak funtsezko tresna dira hiri-mugikortasunaren esparruan, eremu geografiko batean zehar oinezkoen, garraio publikoaren edo ibilgailuen mugimendu ereduak deskribatzen baitituzte. Informazio hau izateak erabaki egokiagoak hartzea baimentzen du garraio-azpiegituren dimentsionamendua, garraio publikoaren plangintza edo hiri-plangintza egiterakoan. Euren garrantzia dela eta, azken 50 urteetan zehar inferentzia metodo posible desberdinei buruzko ikerketa ugari egin dira. Arazoaren lehen planteamenduek, datuen izaeraren ondorioz, oso azpideterminatua zegoen arazoa aur-

kitzen zuten eta horrek informazio gehigarria eranstea eskatzen zuen. Ikasketa horietako asko teorikoak baino ez ziren edo oso tamaina txikiko agertokietan ezartzen ziren. Urteen poderioz eta datu mota berriak agertzearekin batera, inferentzia-metodoen aniztasuna nabarmen handitu da, baita egoera errealekiko aplikagarritasuna ere. Hala ere, datu mota berri hauek eragozpenak dituzte eta euren eskurapena mugatua denez, jatorri-helmuga matrizeen estimazioa arazo irekia izaten jarraitzen du mugikortasunaren arloan aritzen diren enpresa askorentzat. Proiektu honetan, mugikortasunaren arloko enpresa batekin lankidetzan, ohiko datuetatik abiatuta estimazio-metodo berritzaile bat teorizatu eta inplementatu da eta haren funtzionamendua benetako agertoki batean ebaluatu da.

Índice

1. Introducción	9
2. Contexto	12
3. Objetivos y Alcance	18
4. Beneficios	19
4.1. Beneficios técnicos	19
4.2. Beneficios sociales	19
4.3. Beneficios económicos	19
5. Estado del arte	20
5.1. Métodos tradicionales	20
5.2. Métodos basados en datos convencionales	21
5.2.1. Estimación de matrices OD estáticas	21
5.2.2. Estimación de matrices OD dinámicas	26
5.3. Métodos basados en datos modernos	27
5.3.1. Métodos basados en FCD pasivos	27
5.3.2. Métodos basados en FCD activos	28
6. Descripción de la solución propuesta	30
6.1. Descripción del escenario	30
6.2. Descripción de los datos	32
6.3. Método de resolución	33
6.3.1. Descripción matemática	33
6.3.2. Selección de puntos de origen-destino	36
6.3.3. Matriz de conectividad	38

6.3.4. Resolución del problema	40
6.3.5. Diagrama de flujo	41
6.4. Consideraciones matemáticas	42
6.4.1. Análisis del sistema de ecuaciones	42
6.4.2. Análisis de la convexidad	43
6.5. Fuentes de error	45
7. Descripción y análisis de los resultados	46
7.1. Resultados para datos sintéticos	46
7.1.1. Error por discretización espacial	52
7.1.2. Comparativa de matrices solución y reducida	54
7.2. Resultados para datos reales	56
8. Planificación	58
8.1. Equipo de trabajo	58
8.2. Fases del proyecto	58
8.3. Hitos	64
8.4. Diagrama de Gantt	64
9. Presupuesto	68
9.1. Horas internas	68
9.2. Amortizaciones	68
9.3. Gastos	69
9.4. Coste total	69
10. Conclusiones y trabajo futuro	70
10.1. Matriz OD dinámica	70
10.2. Congestión	73
10.2.1. Problema estático	73
10.2.2. Problema dinámico	74
Bibliografía	81
A. Anexo I	82

Lista de Ilustraciones

1.	Matriz origen-destino ejemplo.	10
2.	División de la ciudad de Bilbao por distritos, imagen obtenida de [72].	12
3.	Red de carreteras de la ciudad de Bilbao.	13
4.	Matriz OD Estática (izquierda) y Dinámica (derecha), imagen obtenida de [74].	14
5.	Sensores intrusivos: (a) Sensor magnético, (b) Tubo neumático y (c) Espira. Imagen obtenida de [41].	15
6.	Sensores no intrusivos: (a) Radar, (b) Cámara de vídeo y (c) Sensor infrarrojo. Imagen obtenida de [41].	15
7.	División en zonas de la ciudad de Bilbao.	31
8.	Ubicación de las espiras.	32
9.	Asociación manual de espiras a nodos OSM.	39
10.	Proceso de construcción de la matriz C	40
11.	Diagrama de flujo del método propuesto.	42
12.	Ejemplo de función convexa, imagen obtenida de [17].	44
13.	Diez conjuntos de datos sintéticos sobre las espiras.	47
14.	Matriz semilla y matrices reducidas.	48
15.	Matriz origen-destino solución para minimización de la norma y 20 puntos por zona.	49
16.	Matriz origen-destino solución para minimización de la norma y 30 puntos por zona.	49
17.	Matriz origen-destino solución para minimización de la norma y 40 puntos por zona.	50
18.	Matriz origen-destino solución para minimización de la distancia a la media y 20 puntos por zona.	50
19.	Matriz origen-destino solución para minimización de la distancia a la media y 30 puntos por zona.	51
20.	Matriz origen-destino solución para minimización de la distancia a la media y 40 puntos por zona.	51

21.	Matrices origen-destino solución para minimización de la norma con 20 y 40 puntos por zona.	52
22.	Matrices origen-destino solución para minimización de la distancia a la media con 20 y 40 puntos por zona.	52
23.	Matriz origen-destino solución media para minimización de la norma y matriz reducida.	55
24.	Matriz origen-destino solución media para minimización de la distancia a la media y matriz reducida.	56
25.	Matrices origen-destino solución medias para minimización de la distancia a la media (derecha) y norma (centro), y matriz reducida (izquierda).	56
26.	Matrices origen-destino solución medias para minimización de la norma (izquierda) y de la distancia a la media (derecha), para datos reales.	57
27.	Diagrama de Gantt (1).	65
28.	Diagrama de Gantt (2).	66
29.	Diagrama de Gantt (3).	67

Lista de Tablas

1.	Datos demográficos de los distritos de Bilbao.	30
2.	Variación de los RIC entre las matrices origen-destino solución para minimización de la norma.	53
3.	Variación de los RIC entre las matrices origen-destino solución para minimización de la distancia a la media.	54
4.	Grupo de trabajo.	58
5.	Hitos del proyecto.	64
6.	Horas internas.	68
7.	Amortizaciones.	68
8.	Gastos.	69
9.	Coste total.	69

Lista de Acrónimos

UE	Unión Europea
OMS	Organización Mundial de la Salud
OD	Origen-Destino
ZBE	Zona de Bajas Emisiones
TAZ	Zona de Análisis de Tráfico
DTA	Asignación Dinámica del Tráfico
FCD	Floating Car Data
CDR	Call Detail Record
GPS	Global Positioning System
SML	Laboratorio de Movilidad Inteligente
GR	Gravitatorio
GO	Gravedad-Oportunidad
IM	Minimización de la Información
EM	Maximización de la Entropía
ML	Máxima Verosimilitud
GLS	Mínimos Cuadrados Generalizados
BI	Interferencia Bayesiana
MCO	Mínimos Cuadrados Ordinarios
CDA	Distribución y Asignación Combinadas
PFE	Estimación del Flujo del Trayecto
AVI	Identificación de Vehículos Automática
OSRM	Open Source Routing Machine
OSM	Open Street Map
JOSM	Java OpenStreetMap
RIC	Rango InterCuartil

1. Introducción

En las últimas décadas, la movilidad urbana ha sido uno de los principales retos que afrontan las ciudades de la Unión Europea (UE). La movilidad urbana se define como el conjunto de desplazamientos, tanto de personas como de mercancías, que se dan en ciudades o zonas urbanas gracias a la red y servicios de transporte. La movilidad urbana engloba todos los desplazamientos, independientemente del medio que se utilice para ello, ya sea a pie, en transporte público, automóvil, bicicleta, etc. Son muchos los factores que afectan a los modelos de movilidad urbana, como la demografía, el uso del suelo, la disponibilidad de transporte público, el uso del automóvil y la economía local. Según [69], los principales retos a los que se enfrenta la movilidad urbana son: los accidentes, la congestión y la contaminación atmosférica y acústica. Además, en los años recientes, han surgido nuevos desafíos, como la movilidad eléctrica, la micromovilidad y el creciente impacto del reparto de mercancías de última milla.

Uno de los mayores retos es la contaminación atmosférica, tanto a nivel de la calidad del aire de ciudades y zonas urbanas, como a nivel de la emisión de gases con efecto invernadero. Muchas ciudades europeas se ven afectadas por la mala calidad del aire, excediendo los valores límite establecidos para la protección de la salud humana. Según la Agencia Europea de Medio Ambiente, en 2016, uno de cada ocho ciudadanos europeos residentes en zonas urbanas estaban expuestos a niveles de contaminantes atmosféricos que superan una o varias normas de calidad del aire de la UE. Además, hasta el 96 % de los ciudadanos urbanos de la UE estaban expuestos a niveles de uno o más contaminantes atmosféricos considerados perjudiciales para la salud por las directrices más estrictas de la Organización Mundial de la Salud (OMS) [1]. De hecho, según la OMS, la contaminación atmosférica es el mayor riesgo medioambiental para la salud en la UE, causando cada año alrededor de 400 000 muertes prematuras [71]. En 2013, la Comisión Europea estimó entre 330 000 y 940 000 millones de euros anuales los costes totales que la contaminación atmosférica tiene para la salud [31].

El transporte por carretera es una de las causas principales de la contaminación del aire y de la emisión de gases con efecto invernadero. En el caso de España, en 2017, el transporte por carretera supuso aproximadamente el 21 % de las emisiones de gases de efecto invernadero, el 41 % de las emisiones de óxidos de nitrógeno y del 11 % de las partículas [35]. Siendo mayor su contribución al deterioro de la calidad del aire en entornos metropolitanos y núcleos urbanos.

Otro de los grandes retos es la congestión del tráfico, ya que ocurre con frecuencia en zonas urbanas y en sus alrededores, afectando a la mayoría de los habitantes. Diversas investigaciones económicas indican que el coste derivado de la congestión vehicular para la sociedad es elevado, de hecho se estima que asciende hasta los 270 000 millones de euros al año en la UE [32]. Además, dichas investigaciones indican que una mayor fluidez

del tráfico en una zona urbana, aumenta las probabilidades de crecimiento económico. En [42] los autores sostienen que la transición al tráfico de flujo libre podría aumentar la productividad de los trabajadores hasta en un 30 % en regiones muy congestionadas.

Dado el impacto de la movilidad urbana tanto en la salud de los ciudadanos, como en el crecimiento económico y en el medio ambiente, el transporte sostenible está adquiriendo cada vez más peso en las agendas de desarrollo sostenible de las administraciones públicas, que consiste en el empleo de modos de transporte más limpios y sostenibles. Para dar soluciones a estos retos, surgen ámbitos tales como el de la gestión de la movilidad, en el cual se estudia la movilidad urbana para ayudar en la toma de decisiones. Ya sea mediante la visualización y el análisis de las situaciones actuales, mediante simulaciones de situaciones a futuro, mediante estudios de impacto de decisiones o del uso de nuevas tecnologías, etc.

En el ámbito de la gestión de la movilidad, un área de estudio de especial relevancia es la inferencia de los patrones de desplazamiento en zonas geográficas dadas, bien sea de peatones, transporte público o vehículos. La herramienta habitualmente empleada para describir dichos patrones de desplazamiento es la matriz origen-destino (OD), la cual contiene en cada elemento el número de viajes desde una zona de origen a una zona destino.

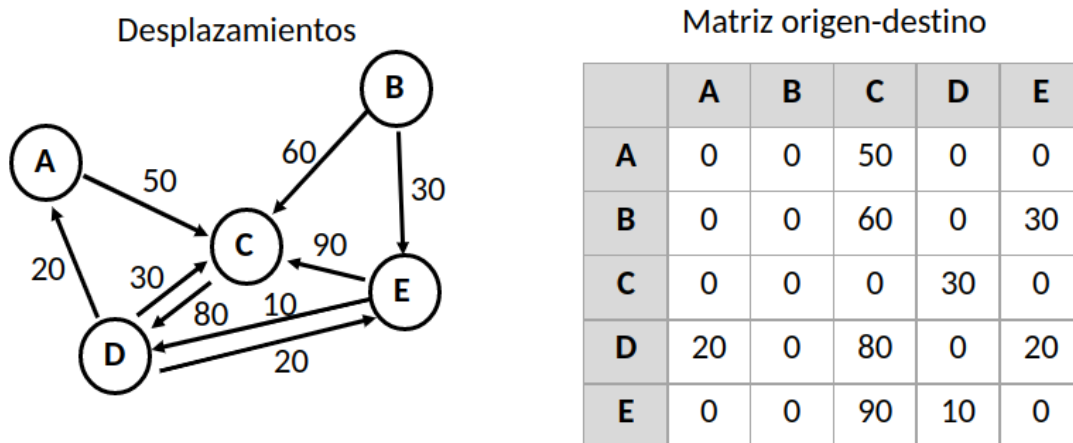


Figura 1: Matriz origen-destino ejemplo.

Las matrices OD son una representación de alto nivel de la movilidad entre las diferentes zonas que conforman un área de estudio. En el ámbito de la movilidad urbana tienen una gran variedad de aplicaciones:

- Dimensionamiento de infraestructuras de transporte: Conocer los patrones de desplazamiento puede ayudar a la hora de dimensionar una vía de transporte. Por ejemplo, a la hora de decidir el número de carriles que debe tener una nueva carretera o para adaptar una ya existente al flujo de tráfico.
- Planificación del transporte público: Conocer los patrones de desplazamiento de los peatones e incluso de los vehículos puede ayudar a la hora de ajustar la oferta de transporte público. Por ejemplo, ajustando específicamente, según la demanda, la frecuencia o el número de autobuses entre zonas.
- Planificación urbana: Conocer la cantidad y la distribución de los desplazamientos a lo largo de una zona urbana, puede ayudar en la toma de decisiones acerca de

intervenciones para la mejora de la red de carreteras. Por ejemplo, con la llegada de las zonas de bajas emisiones (ZBE), resulta útil contar con matrices origen-destino para estimar el impacto del cierre de ciertas zonas de la ciudad a vehículos contaminantes.

- Simuladores de tráfico: Generalmente los simuladores de tráfico se alimentan mediante modelos de demanda de tráfico, como las matrices origen-destino, para generar situaciones realistas de tráfico.

2. Contexto

En este capítulo del documento se introducen la terminología y los conceptos básicos del ámbito de las matrices origen-destino para la movilidad vehicular. El propósito de esta sección es ofrecer la base necesaria para facilitar la comprensión de los capítulos posteriores.

Como se ha descrito en el capítulo anterior, una matriz origen-destino contiene en la fila i y columna j el elemento m_{ij} , que representa el número de viajes desde la zona de origen i a la zona destino j . Generalmente, en el ámbito de la movilidad aplicada a vehículos, las zonas se denominan zonas de análisis de tráfico (TAZ), y se crean a partir de la división del área de estudio. Esta división se realiza en base a la aplicación de uso y a la información disponible. Es decir, por un lado, la división se realiza en función de la información que se desea extraer de la matriz origen-destino solución. Por ejemplo, a la hora de analizar los patrones de desplazamiento en una ciudad, es común emplear los distritos o los barrios como zonas de análisis de tráfico (Figura 2). Por otro lado, la división del área es dependiente de los datos de los que se dispone. Va que no es posible estimar el tráfico de una zona sin disponer de información acerca del tráfico en ella.

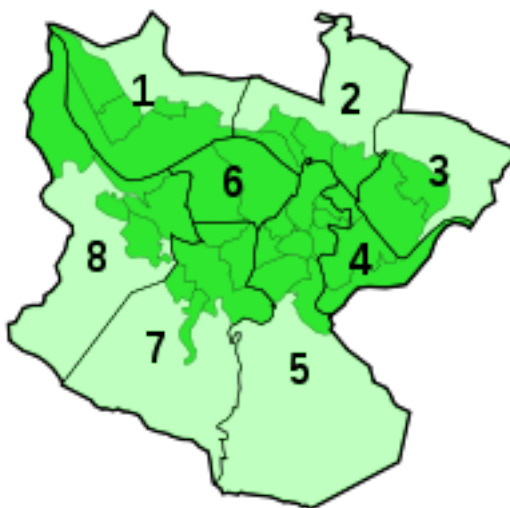


Figura 2: División de la ciudad de Bilbao por distritos, imagen obtenida de [72].

Una vez establecidas las zonas de análisis de tráfico, es necesario definir la red de carreteras del área de estudio. La red de carreteras se representa mediante una colección de puntos, denominados nodos, y una colección de pares ordenados de nodos llamados enlaces. Un nodo representa un punto concreto de la red de carreteras, por ejemplo una intersección, y se identifica mediante un número único. Un enlace representa una sección de la carretera, y se identifica mediante los números de los nodos en sus extremos. Una

vía representa una calle o parte de una carretera y se define por los enlaces que hay a lo largo de su recorrido.



Figura 3: Red de carreteras de la ciudad de Bilbao.

Dos conceptos que aparecen frecuentemente en este ámbito, y que conviene definir, son los de viaje y trayectoria (o ruta). Un viaje se define mediante: el par de puntos geográficos de origen y destino, la trayectoria escogida entre dicho par de puntos, y los tiempos de salida y llegada. Mientras que una trayectoria (o ruta), se define mediante el conjunto de enlaces que se atraviesan desde el origen al destino. Para un mismo viaje pueden existir múltiples trayectorias que unen el origen con el destino. Cada trayectoria tiene un coste asociado, ya sea en términos de tiempo, distancia, dinero o una agregación de los anteriores. Este coste es la suma de los costes de cada uno de los enlaces que componen la trayectoria.

Se denomina asignación al proceso que determina de qué manera se distribuye la demanda de viajes sobre la red de carreteras [76]. Este proceso se basa en hipótesis, acerca del comportamiento de los vehículos, a la hora de escoger la trayectoria de cada viaje. Existen diferentes métodos de asignación en función de las hipótesis que se asumen, los más comúnmente empleados son:

- **Asignación todo o nada:** Este método asume que todos los conductores perciben los costes de ruta de la misma manera y que dichos costes no dependen de la congestión del tráfico. Por lo que todos los vehículos que viajan desde un mismo punto de origen a un mismo punto destino, siguen la misma ruta, siendo esta la de menor coste.
- **Asignación proporcional:** Este método asume que no todos los conductores perciben el mismo coste para una misma ruta. Por lo tanto, para ir desde un origen a un destino, los conductores eligen entre un conjunto reducido de trayectorias cercanas a la óptima. La proporción de vehículos en cada trayectoria viene definida por

las elecciones de los conductores y por las características de la ruta. Este método tampoco tiene en cuenta la congestión a la hora de elegir la ruta de menor coste.

- **Asignación de equilibrio:** Este método asume que cada conductor trata de minimizar su coste de ruta adaptándose a las condiciones del tráfico. Es decir, teniendo en cuenta la congestión a la hora de elegir la ruta. Este método trata de satisfacer el primer principio de Wardrop [94]: las duraciones del trayecto de todas las rutas que están siendo utilizadas entre un origen y un destino, son menores o iguales a la que experimentaría un único vehículo en cualquier ruta no utilizada. Es decir, el sistema de tráfico se encuentra en equilibrio si ningún usuario puede reducir su duración del trayecto mediante un cambio de ruta.

Además, el proceso de asignación se divide en dos categorías en función del empleo de la variable temporal, ya que el tiempo influye directamente sobre la asignación. Es decir, los instantes de salida de los viajes y la congestión de la red de carreteras a lo largo del tiempo, influyen directamente sobre los tiempos de ruta, y por lo tanto sobre la elección de las mismas. Por un lado se encuentra la asignación estática del tráfico, la cual no tiene en cuenta la influencia de la variable temporal en el sistema. Este tipo de asignación es simple, computacionalmente eficiente y soporta el uso de grandes redes de transporte. Por otro lado, se encuentra la asignación dinámica del tráfico (DTA), la cual asume que la demanda de tráfico es dependiente del tiempo y por lo tanto también lo son el flujo vehicular y la congestión en los enlaces de toda la red de carreteras. En otras palabras, la variabilidad temporal supone que la asignación se realice en función del estado de la red en dicho instante. Por lo que este tipo de asignación es más representativa de la realidad, pero a su vez introduce mayor complejidad.

Otro aspecto fundamental a tener en cuenta, es que el problema de la estimación de matrices origen-destino se divide en dos categorías en función de la dependencia de la demanda de tráfico con la variable temporal:

- **Estimación de matrices estáticas:** En este tipo de problema, se analiza un intervalo temporal considerando que los flujos entre orígenes y destinos no varían con el tiempo. El resultado de este problema es una única matriz origen-destino que se ajusta a un modelo de demanda de tráfico estático.
- **Estimación de matrices dinámicas:** En este tipo de problema se analiza un intervalo temporal, teniendo en cuenta variaciones en los flujos entre zonas a lo largo del mismo, es decir, con un modelo de demanda de tráfico dinámico. Esto implica que el resultado del problema es una matriz origen-destino que evoluciona con el tiempo, o visto de otra forma, una serie temporal de matrices.

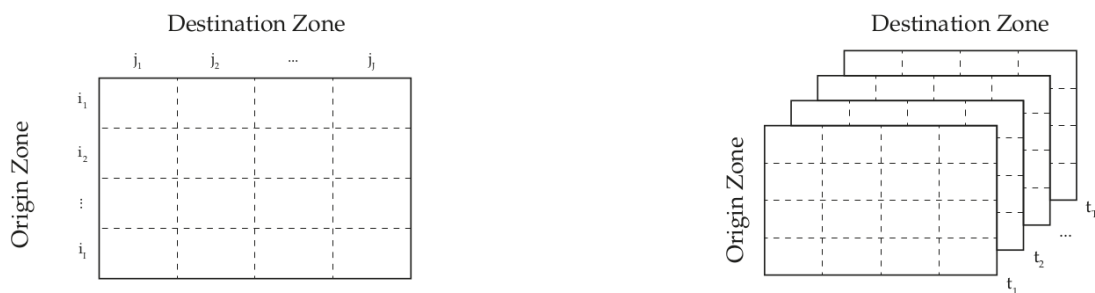


Figura 4: Matriz OD Estática (izquierda) y Dinámica (derecha), imagen obtenida de [74].

Por último, es necesario introducir los tipos de datos que se emplean para la resolución de estos problemas. Esencialmente los tipos de datos se pueden agrupar en dos categorías en función de la información que aportan: los datos que indican la cantidad (o flujo) de vehículos en un lugar dado de la red, conocidos como datos convencionales; y los datos que identifican unívocamente a los vehículos en uno o varios lugares de la red, conocidos como *Floating Car Data* (FCD). Para capturar los datos, se emplean tanto sensores instalados en la calzada u otras infraestructuras de la red de carreteras, como sensores o dispositivos a bordo del vehículo. Los sensores situados en la infraestructura de la red, se dividen en dos categorías en base a su ubicación [41]:

- **Sensores intrusivos:** Se caracterizan por estar instalados en la calzada. La mayor ventaja de este tipo de sensores es su exactitud a la hora de detectar vehículos, razón por la cual han sido ampliamente implementados a lo largo de los años. Aunque por otro lado, su coste de instalación y mantenimiento es elevado. Además, tanto la instalación y mantenimiento, como la reparación de los mismos, requiere de la interrupción del tráfico. Los principales sensores dentro de esta categoría son: los tubos neumáticos de carretera, los sensores magnéticos y los detectores de lazo inductivo o espiras.

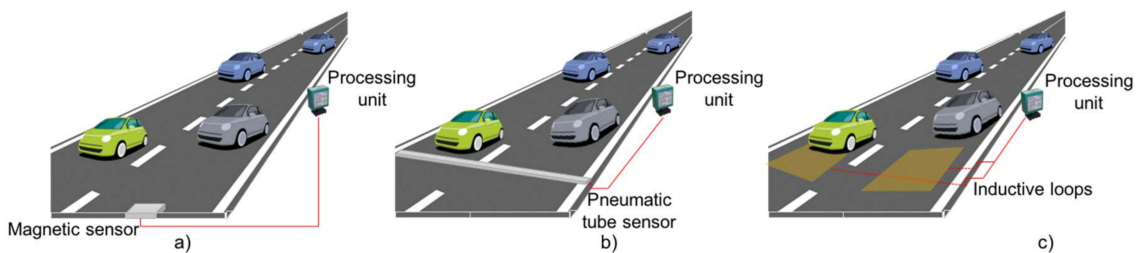


Figura 5: Sensores intrusivos: (a) Sensor magnético, (b) Tubo neumático y (c) Espira. Imagen obtenida de [41].

- **Sensores no intrusivos:** Se caracterizan por estar instalados en distintos lugares fuera de la calzada, montados sobre mástiles o puentes, o a los lados de la calzada. Son capaces de detectar el tránsito de vehículos, así como la velocidad de los mismos, o la ocupación de los carriles. Además su instalación, mantenimiento y reparación no requiere de la interrupción del tráfico. Sin embargo, son costosos y se pueden ver muy perjudicados por ciertas condiciones climáticas como: nieve, lluvia o niebla entre otras. Los principales sensores dentro de esta categoría son: las cámaras de vídeo, los sensores radar e infrarrojos.

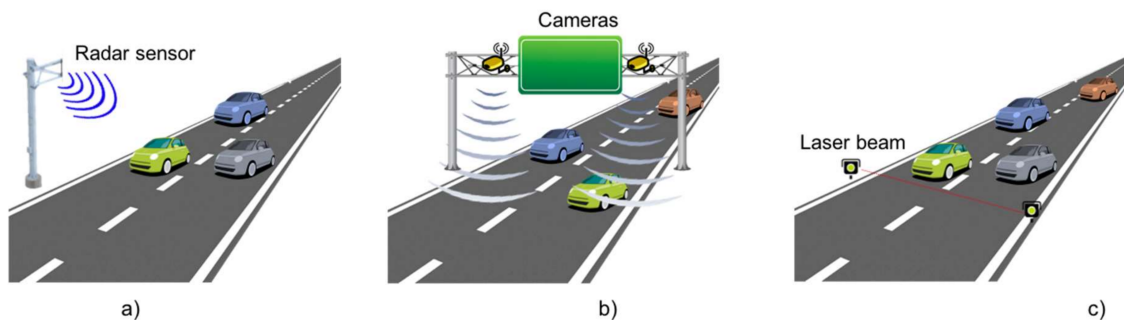


Figura 6: Sensores no intrusivos: (a) Radar, (b) Cámara de vídeo y (c) Sensor infrarrojo. Imagen obtenida de [41].

Además, estos sensores se dividen en dos categorías dependiendo del nivel de información (o datos) que proveen sobre los vehículos [39]:

- **Sensores pasivos:** Este tipo de sensores únicamente cuentan el número de vehículos sobre uno o varios carriles para determinar el flujo de tráfico (medido en vehículos por hora), es decir, ofrecen datos convencionales. Dentro de esta categoría se encuentran las espiras y los tubos neumáticos entre otros.
- **Sensores activos:** Este tipo de sensores tienen la capacidad de identificar a los vehículos. Esta identificación se realiza mediante reconocimiento de matrículas, lectura de etiquetas instaladas en los vehículos, o antenas Bluetooth. Dentro de esta categoría se encuentran las cámaras de vídeo y los sensores infrarrojos entre otros. Estos sensores ofrecen datos de tipo FCD, concretamente FCD pasivos [77], que son aquellos FCD que identifican unívocamente a los vehículos, pero únicamente en aquellos puntos de la red de carreteras donde se dispone de sensores.

Por otro lado, se encuentran los sensores o dispositivos abordo de los vehículos, como los dispositivos GPS y los teléfonos móviles. Este tipo de sensores o dispositivos capturan datos de geolocalización a lo largo del tiempo, ya sea activamente mediante el GPS o los datos de telefonía móvil (CDR¹), o pasivamente mediante ciertas aplicaciones del teléfono móvil. Estos sensores por lo tanto ofrecen datos de tipo FCD, concretamente FCD activos [77], ya que permiten la identificación unívoca de los vehículos a lo largo del tiempo, independientemente del lugar de la red de carreteras en el que se encuentren.

Una desventaja que comparten ambas categorías de FCDs, es que los vehículos capturados son únicamente una porción del tráfico total de la red. Aunque según [74], cuando se reúne una gran cantidad de datos de esta porción del tráfico en días similares, se puede inferir el estado del tráfico mediante técnicas de análisis de datos.

En cuanto a los FCDs activos, un inconveniente que comparten es que no es posible identificar el modo de transporte de los viajes. Inferirlo requiere de métodos como proyectar los puntos obtenidos a partir de los datos sobre la red de diferentes modos de transporte, para estimar cual se utiliza. Aunque esto sólo ha sido probado con éxito en distancias largas y en zonas de baja densidad. Otros estudios analizan las velocidades de desplazamiento y las comparan con las de cada modo de transporte, así como si hay una estación al principio o al final del trayecto [74].

Respecto a los datos CDR, su mayor inconveniente es que el posicionamiento del dispositivo móvil puede ser impreciso, ya que se calcula mediante triangulación con torres de telefonía móvil. Para ello se necesitan de al menos tres torres, por lo que el posicionamiento es más preciso en zonas urbanas, donde se dispone de una mayor cantidad de torres de telefonía. Aunque existen una gran variedad de aspectos relacionados con la tecnología de las antenas que pueden afectar significativamente a la precisión del posicionamiento. Generalmente, estos datos proveen información global acerca de la movilidad de los dispositivos a ciertas horas del día, pero no permiten reconstruir las trayectorias [74].

¹Un *Call Detail Record* (CDR) es el registro informático producido por una central telefónica que contiene los detalles de una llamada que ha pasado por ella. Ejemplos de los campos de un CDR típico son los números de los usuarios, la hora de inicio, la duración, el tipo de llamada (voz, SMS, etc.), resultado de la llamada (contestada, ocupada, etc.), identificador de la central, etc. Los datos CDR son baratos, ya que son producidos por los equipos de telecomunicaciones cuando los usuarios hacen llamadas telefónicas, envían o reciben mensajes, correos electrónicos o navegan por páginas web [93].

Por último, los datos de geolocalización provistos por dispositivos GPS y teléfonos móviles también tienen sus inconvenientes. En primer lugar, este tipo de datos suelen estar en manos de grandes compañías que ofrecen servicios de movilidad (en ocasiones también otro tipo de servicios en dispositivos móviles que capturan información de localización de usuarios), y normalmente sólo es posible el acceso a pequeñas partes de los mismos comprándolos a precios relativamente altos. Debido a esto, habitualmente se recurre a proveedores comerciales que compran y procesan grandes conjuntos de datos de geolocalización [2]. Además, este tipo de datos no siempre se encuentran en condiciones óptimas de utilización, lo cual implica que la limpieza y el procesamiento de los mismos sea costoso. Otro inconveniente es la variabilidad de la frecuencia de muestreo de las trazas GPS entre individuos, ya que depende del dispositivo empleado y de la política comercial de los productores y proveedores de datos. Esto se traduce en que para los casos con una baja frecuencia de muestreo, no sea posible inferir la trayectoria seguida a lo largo de un intervalo temporal. En algunos casos, esto lleva a tomar decisiones acerca de su agregación que inevitablemente sesgan el resultado.

3. Objetivos y Alcance

Este proyecto surge en colaboración con el centro de investigación y desarrollo tecnológico Tecnalía [87], más concretamente, con el Laboratorio de Movilidad Inteligente (SML) [52]. En el SML se realiza investigación y desarrollo para el análisis de la movilidad utilizando técnicas de inteligencia artificial y herramientas Big Data. Entre otras tareas, se realizan predicciones de tráfico tanto a corto como a largo plazo, se elaboran herramientas para la descripción de la movilidad tanto de forma individual como agregada y se desarrollan optimizadores para la planificación de rutas multimodales. Este proyecto pretende contribuir a la ampliación del conocimiento en materia de estimación de matrices origen-destino en el SML. El objetivo principal de esta cooperación es el desarrollo e implementación de un método para la estimación de matrices origen-destino aplicadas a la movilidad de vehículos particulares. Los datos a partir de los cuales se desarrolla el proyecto se tratan de datos de conteo de vehículos, recogidos por espiras instaladas sobre la red de carreteras de la ciudad de Bilbao.

Para alcanzar el objetivo principal se abordan los siguientes objetivos parciales:

- Estudio de los métodos de estimación de matrices origen-destino disponibles en la literatura, incidiendo en mayor profundidad en aquellos a partir de datos de conteo convencionales.
- Desarrollo teórico del método de estimación propuesto.
- Implementación del método propuesto para la ciudad de Bilbao.
- Evaluación del método y de las soluciones obtenidas a partir del mismo.
- Indicación de posibles mejoras sobre el método propuesto y planteamiento de posibles nuevas líneas de investigación a partir del mismo.

En cuanto al alcance, tal y como se describe en los Capítulos 5 y 6, el problema a resolver, a partir de los datos disponibles, no cuenta con solución única. Por lo tanto, el alcance del proyecto consiste en desarrollar un método con base teórica rigurosa, implementarlo para realizar distintas estimaciones de posibles matrices origen-destino, y evaluar la consistencia de las soluciones con la realidad.

4. Beneficios

En este capítulo se describen los beneficios que puede aportar el proyecto realizado, tanto a nivel técnico, como social y económico.

4.1. Beneficios técnicos

Los beneficios técnicos aportados por este proyecto se dividen en los aportados a nivel teórico y en los aportados a la empresa colaboradora.

A nivel teórico, a diferencia de gran parte de los métodos que se encuentran en la literatura científica, el método desarrollado es implementable sobre una red de carreteras urbana real, ofreciendo resultados consistentes con los datos sobre las espiras. Además, no requiere del empleo de software de pago ni de hardware avanzado para su implementación.

A nivel de empresa, la realización de este proyecto ha permitido aumentar el conocimiento sobre matrices origen-destino en el SML. Este proyecto deja el camino abierto a futuras investigaciones en esta línea de trabajo, y señala las direcciones en las que se deberían concentrar los futuros esfuerzos. Además, el software desarrollado queda disponible para su futuro uso o reutilización.

4.2. Beneficios sociales

Las matrices origen-destino pueden ser empleadas en la toma de decisiones a la hora de realizar planificación del transporte público, planificación urbana y dimensionamiento de infraestructuras de transporte. Todos estos aspectos repercuten directamente a nivel social, por lo que, contar con este tipo de herramientas es clave para tomar las mejores decisiones. Este proyecto se realiza con el fin de desarrollar un método capaz de generar matrices origen-destino consistentes, que posteriormente puedan ser aplicadas en este tipo de situaciones.

4.3. Beneficios económicos

El proyecto es un trabajo de investigación, por lo que no busca un beneficio económico en sí mismo. Sin embargo, su implementación en proyectos de movilidad urbana, por parte de la empresa colaboradora, puede conllevar beneficios.

5. Estado del arte

Los métodos para la estimación de matrices origen-destino han evolucionado a lo largo del tiempo, principalmente gracias a la aparición de nuevos tipos de datos. Este capítulo ofrece una revisión conjunta, en orden cronológico, de los métodos de estimación de matrices OD junto con los tipos de datos empleados para ello.

5.1. Métodos tradicionales

En los años previos a la década de los 70, no se disponía de datos de tráfico propiamente, por lo que se empleaban los llamados métodos tradicionales. Algunos de los métodos empleados son los siguientes [96]:

- Encuestas a pie de carretera: Se entrevista a conductores a la salida de zonas acordonadas, realizando habitualmente un muestreo de 1 de cada 5 conductores, aunque esto depende de la cantidad de tráfico y de los trabajadores disponibles. Este método es costoso en términos de mano de obra y supone retrasos en el tráfico.
- Encuestas domiciliarias: Se realizan encuestas por muestreo a gran escala, entrevistando a conductores en sus domicilios. No se considera un método lo suficientemente fiable como para realizar diseños detallados de esquemas de gestión del tráfico, además es costoso en mano de obra y tiempo. Aunque a diferencia del método anterior, no provoca retrasos en el tráfico.
- Métodos de señalización: Se identifica a los coches en distintos lugares de la red de carreteras, mediante pegatinas, el número de matrícula o pidiendo a los conductores que realicen señalizaciones con las luces al atravesar ciertos puntos. Estos métodos son costosos en términos de mano de obra y procesamiento, aunque tienen la ventaja de no provocar retrasos en el tráfico.
- Seguimiento de vehículos: Este método es más barato que los anteriores, aunque su aplicación se restringe a las áreas centrales de grandes ciudades.

Los métodos más frecuentemente empleados eran los que involucraban encuestas, aunque debido a su coste se realizaban estudios cada una o dos décadas. Además, para el momento en el que se recogían y procesaban todos los datos de la encuesta, las matrices OD estaban obsoletas. A pesar de ello, las matrices calculadas a partir de estos métodos han sido posteriormente empleadas, como información histórica, en métodos más recientes.

5.2. Métodos basados en datos convencionales

Para hacer frente a las desventajas de los métodos tradicionales, surgen los métodos basados en datos de tráfico convencionales. Este tipo de datos son los recogidos por sensores convencionales, generalmente de tipo pasivos e intrusivos, como las espiras. Se caracterizan por describir las condiciones de tráfico en lugares concretos de la red de carreteras. Sin embargo, no son capaces de describir los patrones de desplazamiento, por lo que el reto reside en inferir dichos patrones a partir de conteos obtenidos mediante espiras o sensores similares. Es decir, a partir de las medidas de flujo de vehículos obtenidas en varios enlaces de la red, estimar la matriz origen-destino que los genera. Este tipo de problemas se conocen como problemas inversos [86], los cuales consisten en el empleo de los resultados de unas medidas para inferir los valores de los parámetros que caracterizan un sistema.

El número de sensores empleados y su ubicación, son factores fundamentales a tener en cuenta en estos problemas. Idealmente, se deben colocar los sensores en aquellos lugares que maximicen la cobertura de la red de carreteras y la observabilidad ¹, y que minimicen la cantidad de sensores empleados, de cara a reducir el coste de la red de detección. A esto se le conoce como el problema de localización de sensores [15], y existen varios enfoques para abordarlo [47]. El rango de aplicaciones de este problema incluye, entre otras: estimación de matrices origen-destino ([100], [16]); estimación de tiempos de viaje [33]; e inferencia de flujo en enlaces ([25], [48], [65], [45]).

5.2.1. Estimación de matrices OD estáticas

En esta sección se presentan los métodos basados en datos convencionales para la estimación de matrices origen-destino estáticas. Para este tipo de problemas se considera un escenario con un número dado de zonas y una red de carreteras definida por un conjunto de enlaces L , donde $A \subseteq L$ es el subconjunto de enlaces en los que se tienen medidas. Para cada enlace $a \in A$, se puede describir el flujo vehicular V_a sobre él mediante la ecuación:

$$V_a = \sum_{ij} m_{ij} p_{ij}^a \quad (5.1)$$

Donde m_{ij} representa el número de viajes entre la zona i y la zona j , y p_{ij}^a representa la proporción de viajes de m_{ij} que atraviesan el enlace a , lo cual depende del método de asignación empleado. Según [96], contando con un número N de zonas, y asumiendo que los p_{ij}^a son conocidos, se necesita de al menos N^2 ecuaciones independientes del tipo (5.1) para definir unívocamente el conjunto completo de $[m_{ij}]$. Sin embargo, en la mayoría de los casos, la cantidad de medidas en enlaces independientes es inferior a este valor. Por ello, tanto en [96] como en [14], se llega a la conclusión de que se trata de un problema subdeterminado y que por lo tanto es necesario introducir más información. En [14] se sugiere introducir una matriz histórica u objetivo, mientras que en [96] se sugiere

¹El problema de la observabilidad consiste en determinar si un determinado subconjunto de mediciones de flujo disponibles es suficiente para estimar otro subconjunto de flujos de tráfico o el estado de una red de tráfico. Los problemas típicos de observabilidad consisten en determinar qué flujos son necesarios para estimar los flujos de pares OD, trayectos o enlaces [27].

realizar asunciones sobre el comportamiento del tráfico. A continuación se describen los métodos propuestos más relevantes para hacer frente al problema recién planteado.

5.2.1.1. Métodos basados en modelos de demanda de viajes

Este tipo de métodos de estimación asumen que el tráfico sigue ciertos patrones de comportamiento, a los que se conoce como modelos de demanda de viajes. El planteamiento en estos problemas es el siguiente, los flujos en los enlaces se expresan como una función de la matriz OD, como en (5.1), pero en este caso la matriz es función de los parámetros del modelo de demanda propuesto. De esta manera, se reduce el número de incógnitas al número de parámetros de calibración del modelo. La resolución implica estimar los parámetros del modelo postulado, de manera que se minimicen los errores entre las medidas estimadas y observadas en los enlaces. Los modelos de demanda de viajes más relevantes son los gravitatorios (GR) y los de gravedad-oportunidad (GO). Los modelos de demanda gravitatorios, se inspiran en la ley de la gravitación de Newton, la cual indica que la fuerza con la que se atraen dos objetos es proporcional al producto de sus masas e inversamente proporcional al cuadrado de la distancia. En este caso, se propone que el número de viajes entre la zona i y la zona j , es directamente proporcional a la capacidad de generar viajes de i y de atraer viajes de j , e inversamente proporcional al cuadrado de la distancia que las separa. Como se puede observar en (5.2), las capacidades de generar y atraer viajes de las zonas se pueden modelar empleando distintos parámetros. Por otro lado, los modelos de gravedad-oportunidad, se basan en el mismo fundamento que los modelos gravitatorios, pero se les añade el efecto de oportunidad de los modelos de oportunidad [82]. Estos modelos asumen que la elección de un destino por parte de un individuo depende tanto de las oportunidades del destino como de las oportunidades intermedias entre el origen y el destino.

Para comprender mejor en qué consisten estos métodos, se describe brevemente el propuesto por Low [55]. Este método emplea un modelo de demanda de viajes gravitatorio lineal y una resolución mediante mínimos cuadrados. El modelo gravitatorio propuesto por Low asume que el número de viajes m_{ij} , entre la zona i y la zona j se puede calcular mediante:

$$m_{ij} = \frac{\alpha P_i E_j}{d_{ij}^2} \quad (5.2)$$

Donde P_i representa población de la zona i , E_j representa el empleo en la zona j , d_{ij} representa la distancia o el tiempo de viaje entre las zonas i y j , y α es el parámetro de calibración. En este caso, el parámetro α desempeña el papel combinado de factor de generación y atracción de viajes. Para calibrar el valor de α , se traslada el modelo de demanda de viajes a la ecuación del flujo medido en las espiras (5.1), quedando de la forma:

$$V_a = \sum_{ij} m_{ij} p_{ij}^a = \alpha \sum_{ij} \frac{P_i E_j}{d_{ij}^2} p_{ij}^a \quad (5.3)$$

A continuación, se asume un método de asignación de viajes para dar valor a los p_{ij}^a en cada enlace. En el caso más sencillo se asume un método de todo o nada, donde:

$$p_{ij}^a = \begin{cases} 1 & \text{Localizacion}(a) \in \text{Trayectoria}(i, j) \\ 0 & \text{Localizacion}(a) \notin \text{Trayectoria}(i, j) \end{cases} \quad (5.4)$$

Tras asumir el modelo de asignación se conocen todos los parámetros, excepto α , en (5.3). Para determinar los valores de α , se emplea una técnica de resolución que minimice la diferencia entre los valores estimados y los valores observados. En el caso de Low, se minimiza el cuadrado de la suma de las diferencias, empleando técnicas de regresión lineal.

Los modelos gravitatorios se pueden presentar con distintos grados de sofisticación, lo cual lleva a dividirlos en dos categorías principalmente: los que dirigen a ecuaciones lineales en los enlaces (Low [55]) y los que dirigen a ecuaciones no lineales (Robillard [73]; Hörgberg [46]). En cuanto a los modelos de gravedad-oportunidad, se pueden encontrar algunos ejemplos de su implementación en [85]. Según [14], a la hora de calibrar los parámetros desconocidos, los modelos GO consumen más tiempo que los GR y no garantizan la fiabilidad de la matriz estimada.

5.2.1.2. Métodos de minimización de la información (IM) y maximización de la entropía (EM)

De acuerdo con el principio de máxima entropía, un sistema cerrado tiende a evolucionar hacia la configuración de mayor entropía posible, es decir, aquella que corresponde a la mayor cantidad de configuraciones microscópicas posibles. Esta configuración es también la más probable de observar en el sistema, y representa un estado de equilibrio. El concepto de información se relaciona con el de la entropía tanto en la mecánica estadística como en la teoría de la información. El concepto de la entropía también tiene aplicaciones en el ámbito del transporte. La idea principal detrás de los métodos IM y EM es encontrar la matriz origen-destino más probable compatible con las medidas de conteo (5.1). En otras palabras, explotar al máximo los datos de flujo en los enlaces para obtener la matriz origen-destino más probable. Según [96], el enfoque de minimización de la entropía se ajusta particularmente bien a este tipo de problemas, es decir, con un gran número de componentes y con aparente complejidad desorganizada. En [91] se recogen los métodos propuestos por Van Zuylen y Willumsen. En él, los autores desarrollan dos modelos, uno basado en IM y otro en EM. Los modelos emplean información de conteo de tráfico y una matriz de referencia para estimar la matriz OD más probable. Según [14] el mayor argumento en contra de estos métodos es no tener en cuenta la incertidumbre en los datos y en la matriz de referencia, lo cual puede tener un gran impacto sobre las soluciones obtenidas.

5.2.1.3. Métodos Estadísticos

Los métodos estadísticos plantean la estimación de las matrices origen-destino como problemas de optimización, donde la función objetivo depende de la técnica de inferencia estadística adoptada. Las técnicas de estimación estadística comúnmente empleadas son Máxima Verosimilitud (ML), Mínimos Cuadrados Generalizados (GLS) e Inferencia Bayesiana (BI). Para la resolución de estos problemas se asume que se dispone de datos de conteo sobre ciertos enlaces, y de información adicional sobre la demanda de viajes, por ejemplo, datos obsoletos de la demanda de tráfico entre zonas.

La estimación mediante ML calcula el conjunto de parámetros que maximiza la verosimilitud de producir los datos observados. Aplicado al ámbito de las matrices origen-destino, se traduce en maximizar la verosimilitud de observar los datos, a partir de la matriz origen-destino solución M . Los datos se componen del conjunto de mediciones de conteo sobre los enlaces, vector \bar{V} , y de un conjunto de flujos OD de referencia, vector \bar{N} . Estos dos conjuntos de datos suelen considerarse estadísticamente independientes, por lo que la verosimilitud de observar ambos conjuntos a partir de la matriz solución puede expresarse como:

$$L(\bar{N}, \bar{V} | M) = L(\bar{N} | M) L(\bar{V} | M) \quad (5.5)$$

La estimación de la matriz solución M se obtiene a partir de la maximización de la función anterior, o más convenientemente de su logaritmo natural:

$$M = \arg \max_{M \in S} \{ \ln L(\bar{N} | M) + \ln L(\bar{V} | M) \} \quad (5.6)$$

Donde S es el conjunto de matrices M factibles, que salvo excepciones, se trata del conjunto de matrices M no negativas. Para el cálculo de la expresión anterior es necesario asumir que los valores de los conjuntos \bar{V} y \bar{N} siguen una determinada función de distribución de probabilidad. Típicamente se asume una distribución normal multivariada o una distribución de Poisson. En [22] se describe con mayor detalle tanto el planteamiento mediante ML descrito, como la resolución del mismo a partir de la expresión (5.5).

Por otro lado se encuentra la estimación mediante mínimos cuadrados generalizados (GLS). Esta técnica es una generalización de la técnica de estimación por mínimos cuadrados ordinarios (MCO). GLS es adecuada para ajustar modelos lineales a conjuntos de datos que presentan heteroscedasticidad (es decir, varianza no constante) y/o autocorrelación. Generalmente los conjuntos de datos suelen presentar estas características, lo que convierte a GLS en una alternativa muy útil a la estimación por MCO. Aplicado al problema en cuestión, tiene la ventaja respecto al método anterior, de no tener que realizar suposiciones sobre la distribución de los conjuntos de datos. Para la formulación de este problema se considera \hat{M} como la matriz origen-destino solución muestreada, obtenida a partir de la agregación de la información adicional de la demanda de viajes. Para la resolución del problema se considera el siguiente sistema estocástico de ecuaciones en M :

$$\hat{M} = M + \eta \quad (5.7)$$

$$\bar{V} = A(M) + \epsilon \quad (5.8)$$

Donde η es el error de muestreo con matriz de varianza-covarianza Z . $A(M)$ representa la asignación de la matriz M , y ϵ es el error en los datos de conteo, con matriz de dispersión W . Una vez planteadas las ecuaciones anteriores, la estimación de la matriz se obtiene a partir de la siguiente expresión:

$$M = \arg \min_{M \in S} \{(\hat{M} - M)'Z^{-1}(\hat{M} - M) + (\bar{V} - A(M))'W^{-1}(\bar{V} - A(M))\} \quad (5.9)$$

Donde S es el conjunto de matrices M factibles. En [22] se describe con mayor detalle tanto el planteamiento mediante GLS descrito, como la resolución del mismo a partir de la expresión (5.9).

Por último se encuentra la estimación mediante inferencia bayesiana (BI). Este método se aplica en problemas de transporte en los que las creencias previas se combinan con las observaciones para producir las creencias posteriores. Llevado al problema en cuestión, la información a priori sobre la matriz origen-destino M se expresa como una función de probabilidad a priori $g(M)$. Los datos de conteo en los enlaces representan otra fuente de información, con probabilidad $L(\bar{V} | M)$. El teorema de Bayes permite combinar estas dos fuentes de información para dar una función de probabilidad a posteriori $f(M | \bar{V})$. Esta función describe la probabilidad de observar M condicionado a que suceda \bar{V} , matemáticamente se expresa mediante:

$$f(M | \bar{V}) = L(\bar{V} | M) g(M) \quad (5.10)$$

La estimación de la matriz origen-destino solución se obtiene a partir de la maximización del logaritmo natural de la función anterior:

$$M = \arg \max_{M \in S} \{\ln f(M | \bar{V})\} \quad (5.11)$$

En [22] se describe con mayor detalle tanto el planteamiento mediante BI descrito, como la resolución del mismo a partir de la expresión (5.11).

Dentro de los métodos estadísticos se puede realizar una división entre aquellos que tienen en cuenta la congestión en la red de carreteras y los que no. Los métodos que no tienen en cuenta la congestión separan en dos procesos independientes la asignación de rutas, es decir la elección de las proporciones p_{ij} de cada ruta, y la estimación de la matriz origen-destino. Dentro de esta categoría, Maher [58] introduce el método de BI para la estimación de matrices origen destino. El método mediante GLS se estudia por Cascetta [21], Bell [12], etc. En [13], Bell resuelve el problema GLS bajo ciertas restricciones y presenta un algoritmo, pero no hay información acerca de su implementación en una red de carreteras real. El método basado en ML se estudia por Spiess [83], Cascetta y Nguyen [23], y Hazelton [43]. En estos estudios la proporción de viajes a través de cada enlace p_{ij}^a es constante. Por otro lado, Lo [54] introduce la aleatoriedad en la elección de las proporciones y estudia los métodos BI y ML, comprobando su funcionamiento sobre un red de tamaño reducido.

Por otra parte se encuentran los estudios que tienen en cuenta la congestión, es decir, aquellos que consideran el efecto de los costes de enlace, las opciones de ruta y de las proporciones de asignación en los flujos sobre los enlaces. Generalmente en estos problemas se emplea alguna variante de la asignación en equilibrio. Algunos de los estudios que

emplean este tipo de asignación son los realizados por Nguyen [66], Yang [98], Cascetta y Postorino [24]. Dentro de los métodos que tienen en cuenta la congestión, se encuentra el planteamiento basado en distribución y asignación combinadas (CDA). El método propuesto por Erlander [36] es modificado y empleado para la estimación de matrices por varios autores. Por otro lado se encuentra el planteamiento de programación binivel. En este planteamiento el problema de nivel superior es el problema de estimación de la matriz y el problema de nivel inferior representa un problema de asignación en equilibrio. Existen varios estudios que emplean este planteamiento, como los realizados por Spiess [84], Yang et al. [99], Kim et al. [51] y Lundgren y Peterson [56]. Por último se encuentra el planteamiento basado en estimación del flujo del trayecto (PFE). Existen varios estudios a partir de este planteamiento, como los realizados por Sherali et al. [81] [79], Nie y Lee [67], y Nie et al. [68].

5.2.1.4. Otros Métodos

Los métodos descritos en las secciones anteriores son los más extendidos en cuanto a uso, en esta sección se agrupan varios métodos alternativos cuyo uso está menos generalizado.

Por un lado se encuentra el método de estimación de matrices multivehículo. Dada la falta de información sobre un único tipo de vehículo, algunos autores incluyen información adicional obtenida a partir de estudios sobre ciertos tipos de vehículos. Este método pretende eliminar la inconsistencia interna de los flujos de tráfico entre las distintas clases de vehículos. Este método ha sido estudiado por Baek et al. [8] y Wong et al. [97]. Por otro lado está el planteamiento mediante lógica difusa, la cual se emplea para modelar situaciones en las que la toma de decisiones de los usuarios es demasiado compleja como para desarrollar un modelo matemático. Este planteamiento se aplica por Teodorovic [88]. Por último, existen métodos basados en redes neuronales, como se puede ver en [40].

5.2.2. Estimación de matrices OD dinámicas

En esta sección se presentan los métodos basados en datos convencionales para la estimación de matrices origen-destino dinámicas. La expresión que resume la formulación dinámica del problema es la siguiente:

$$V_a^{\Delta T} = \sum_t \sum_{ij} p_{(ij,t)}^{(a,\Delta T)} m_{(ij,t)} \quad (5.12)$$

Donde, $V_a^{\Delta T}$ representa el flujo que atraviesa el sensor (espira) a en el intervalo temporal ΔT . $m_{(ij,t)}$ es el número de viajes entre la zona i y j , que parten de su origen en el intervalo de tiempo t . $p_{(ij,t)}^{(a,\Delta T)}$ representa la proporción de los viajes $m_{(ij,t)}$ que atraviesan la espira a durante el intervalo temporal ΔT . Las técnicas comúnmente empleadas en este tipo de problema son los filtros de Kalman [95] y la modelización Estado-Espacio [3]. La estimación dinámica se divide en las categorías online y offline, en función de su aplicación. La estimación dinámica online ofrece estimaciones en tiempo real y la offline en diferido. En general, existen pocos estudios a partir de datos convencionales, y la mayoría de los

métodos desarrollados para la estimación dinámica se aplican en redes cerradas como intersecciones, autopistas y redes de tamaño reducido.

5.2.2.1. Estimación dinámica online

La estimación dinámica online emplea mediciones en tiempo real, de un intervalo de tiempo reducido, para caracterizar los viajes entre cada origen y destino en tiempo real. Es una entrada esencial para los modelos DTA, y se utiliza para la identificación en tiempo real de los patrones de tráfico en los sistemas de control de tráfico. El primer autor en introducir un modelo estado-espacio, donde los flujos origen-destino representan el vector estado es Okutani en [70]. Posteriormente Ashok y Ben-Akiva proponen un enfoque mediante el filtro de Kalman para actualizar dinámicamente una matriz origen-destino [5]. En [50], los autores también investigan la aplicabilidad del filtro de Kalman para la estimación dinámica. En [6], Ashok y Ben-Akiva extienden el trabajo realizado en [5] y presentan una nueva formulación del problema. Para hacer frente a situaciones con falta de sensores en las entradas o salidas de la red, Li y Moor formulan un problema de optimización en [53].

5.2.2.2. Estimación dinámica offline

Este tipo de estimación se realiza en diferido, y se lleva a cabo a partir de una serie temporal de medidas, en algunos casos combinadas con información previa sobre la matriz. En [29], Chang y Tao proponen un método integrando los datos sobre los enlaces y datos sobre giros de flujos en las intersecciones, para determinar matrices OD en intersecciones utilizando un filtro de Kalman. En [80], Sherali presenta un enfoque mediante optimización paramétrica, pero con el incremento de los orígenes y destinos el problema es complejo de resolver. Otros estudios más recientes son los realizados por Tsekereis y Stathopoulos [89], y Hazelton [44].

5.3. Métodos basados en datos modernos

Posteriormente, con la llegada de las Tecnologías de la Información y la Comunicación, la variedad de datos de tráfico se amplía considerablemente. El uso de sensores activos trae consigo nueva información acerca del tráfico, ya que permiten la identificación de vehículos. Es decir, mediante estos sensores es posible detectar a un mismo vehículo en distintos lugares de la red de tráfico. Además, el uso extendido de dispositivos GPS y móviles, propicia la aparición de datos de geolocalización. Estos datos, más allá de ofrecer una medida en un instante en una localización concreta, ofrecen información espacio-temporal.

5.3.1. Métodos basados en FCD pasivos

Los FCD pasivos se caracterizan por ser datos recogidos a partir de sensores activos que permiten reconocer a los vehículos en los lugares donde están ubicados dichos sensores. Los FCD comúnmente empleados son los datos de escaneo de matrículas y los datos obtenidos a partir de tecnologías inalámbricas como bluetooth o WiFi.

En cuanto a los datos de escaneo de matrículas, Castillo et al. en [28], desarrollan un método de reconstrucción de flujos de trayectos, para la estimación de matrices, basado en programación cuadrática, a partir de datos de escaneo de matrículas y de datos de conteo sobre los enlaces. En dicho estudio se concluye que la gran cantidad de información que aportan los datos de reconocimiento de matrículas conduce a obtener mejores estimaciones. Dicho método se aplica en la red Nguyen-Dupuis y en otras redes de mayor tamaño obteniendo buenos resultados, muy superiores a los obtenidos a partir de datos de conteo únicamente. Posteriormente, en [26], Castillo et al. presentan un método para optimizar el uso de recursos de escaneo (cámaras). Por otro lado, Asakura et al. [4], emplean el sistema de identificación de vehículos (AVI), instalado aproximadamente en el año 2000 en la red de autopistas de Japón. Este sistema emplea cámaras para el reconocimiento de matrículas en distintos puntos de la red de carreteras. A partir de los datos capturados por dicho sistema, los autores consiguen implementar un modelo de mínimos cuadrados que obtiene resultados consistentes con las matrices obtenidas a través de encuestas. Cabe mencionar que debido a la falta de cámaras, en el momento en el que se realiza el estudio, el método se aplica únicamente a una porción de la red de autopistas de Han-Shin. Por otra parte, Zhou et al. formulan en [101] un modelo no lineal de mínimos cuadrados ordinarios para la estimación de matrices dinámicas en redes de carreteras urbanas. Dicho método combina datos de AVI, de conteo e información histórica sobre la demanda en una función multiobjetivo. Un estudio más reciente a partir de este tipo de datos es el llevado a cabo por Mo et al. [64], el cual propone un método híbrido entre el propuesto por [28] y [101]. Es decir, un método que combina la estimación de matrices basada en reconstrucción de flujos y basada en optimización, a partir de datos reconocimiento de matrículas. Este método se pone a prueba en la red de carreteras (simplificada) de Langfang. Los resultados muestran que el modelo propuesto puede estimar la distribución de la demanda OD con un error porcentual absoluto medio del 30%.

En cuanto a los datos de tecnologías inalámbricas, en [9] Barcelo et al. emplean datos de bluetooth para estimar matrices origen-destino en autopistas mediante un enfoque ad hoc lineal del filtro de Kalman. Los resultados del estudio indican que el método funciona correctamente para casos con y sin congestión, pero que es necesario un adecuado ajuste de las matrices de inicialización en ambos casos. Posteriormente Barcelo et al. [10] proponen una versión refinada del filtro de Kalman anterior, para calcular las matrices origen-destino tanto en redes urbanas como en autopistas, obteniendo buenos resultados en ambos casos. Este tipo de datos también son empleados en estudios realizados por Michau et al. [63], [62]. En este último se estudia la posibilidad de combinar datos de conteo y de bluetooth para estimar las matrices origen-destino de la ciudad de Brisbane. De este estudio los autores concluyen que el método es aplicable a grandes redes, y ofrecen argumentos acerca de la validez de las estimaciones, sin embargo, no se ofrece un análisis en profundidad de la eficiencia del método por falta de datos de tráfico de referencia.

5.3.2. Métodos basados en FCD activos

Los FCD activos se caracterizan por ser datos recogidos por dispositivos que capturan la geolocalización del vehículo a lo largo del tiempo. Los métodos de estimación de matrices origen-destino a partir de FCD activos emplean generalmente datos de GPS o de CDR.

En cuanto a los datos de CDR, Iqbal et al. proponen en [49], un método de estimación de matrices combinando datos CDR y datos de conteo. El método consiste en analizar los datos CDR para crear una matriz entre torres de telefonía que posteriormente mapean a

la red de carreteras para generar una matriz OD. La matriz solución se estima escalando la matriz anterior, y para determinar los factores de escala, se emplea un enfoque de optimización en conjunto con simulaciones para obtener aquellos factores que mejor concuerdan con las medidas sobre las espiras. Este método se aplica en Dhaka, Bangladesh, a partir de datos CDR de 2,87 millones de usuarios y conteos en 13 lugares clave. En [7], bachir et al. proponen una metodología para inferir flujos OD dinámicos para cada modo de transporte. Para ello analizan datos de 360 millones de trayectorias, de más de 2 millones de dispositivos, de la ciudad de París. Los modos de transporte los identifican a partir de un algoritmo de aprendizaje semi-supervisado de dos pasos. El último paso implica agrupar las zonas de la red móvil y en realizar inferencia bayesiana para generar probabilidades de transporte de las trayectorias. Por último se escalan los flujos al total de la población empleando técnicas del estado del arte actual. Según los autores las matrices obtenidas presentan una alta correlación con las matrices a partir de encuestas, aunque con diferencias absolutas razonables. Otro estudio es el realizado por Fekih et al. [37], donde se propone un método de estimación para la ciudad de Lyon. En resumen, existe una amplia literatura de métodos que emplean datos de tipo CDR, y los resultados parecen prometedores.

Por el contrario, se dispone de poca literatura acerca de los métodos a partir de datos GPS aplicados a vehículos particulares. En [38], Ge y Fukuda emplean un método de estimación a partir de un dataset de información agregada y encuestas. El método consiste en dos submodelos secuenciales de maximización de la entropía. El estudio se lleva a cabo en 23 barrios de Tokio y los autores consideran que los resultados son consistentes con sus intuiciones. Sin embargo, reconocen que los únicos viajes que pueden ser inferidos a partir del dataset agregado son aquellos que tienen que ver con el trabajo. En [75], Ros-Roca et al. concluyen que los resultados para la ciudad de Turín son prometedores, pero que pueden mejorarse. También recalcan que los datos GPS de los que disponen, adquiridos de distribuidores profesionales, por como vienen dados, no aportan suficiente información como para generar una matriz de referencia a partir de ellos.

6. Descripción de la solución propuesta

6.1. Descripción del escenario

En esta sección se presenta el escenario de estudio empleado en este proyecto. Se introducen varias características acerca del área en el que se ubica, y se describe la zonificación empleada y el porqué de la misma.

El escenario de estudio de este proyecto se sitúa en la ciudad de Bilbao. Tal y como muestra la Figura 2, la ciudad de Bilbao se divide en ocho distritos, que a su vez están compuestos por un total de 34 barrios. Previamente al cálculo de las matrices origen-destino, conviene realizar un estudio sobre las características principales del escenario, ya que conocer la naturaleza del mismo ayuda a la hora de interpretar correctamente las soluciones (estimaciones) obtenidas. En primer lugar, es necesario tener en cuenta los datos demográficos más relevantes de cada uno de los distritos y barrios, en la Tabla 1 se muestran los datos referentes a los distritos. Además, también es necesario valorar otros aspectos, como la cantidad de puestos de trabajo que alberga cada barrio y distrito. Así como, las alternativas culturales o de ocio que ofrecen. Estos últimos factores, a su vez, también dependen de la fecha y el rango temporal con los que se quiera trabajar. Es decir, factores como la época del año, las festividades o la hora del día afectan sensiblemente al volumen de tráfico. Por lo tanto, es necesario valorar todos estos aspectos en conjunto para formar una intuición del orden de magnitud del volumen de tráfico que existe entre cada uno de los distritos y barrios.

Distrito	Código Distrito	Habitantes	Superficie (km ²)	Densidad población (hab/km ²)
Deusto	1	49 193	4,96	9 917
Uribarri	2	36 181	4,19	8 635
Ocharcoaga-Churdinaga	3	26 458	3,90	6 784
Begoña	4	40 658	1,77	22 971
Ibaiondo	5	61 184	9,65	6 340
Abando	6	50 903	2,14	23 786
Recalde	7	47 245	6,96	6 788
Basurto-Zorroza	8	31 672	7,09	4 467

Tabla 1: Datos demográficos de los distritos de Bilbao.

El primer paso para la obtención de las matrices origen-destino es determinar el escenario, el cual se crea a partir de la división geográfica del área de estudio en zonas de análisis de tráfico (TAZ). Tal y como se menciona en el Capítulo 2, la división se realiza, por un lado, en base a la información que se desea extraer de la matriz origen-destino, y por otro lado, en función de los datos disponibles. Para este proyecto, la información que se desea extraer de la ciudad, es el flujo de vehículos entre los distritos que la conforman. Así como

el flujo de tráfico proveniente o con destino en el exterior de la ciudad. Sin embargo, los datos de los que se dispone (espiras) están recogidos en ubicaciones, que para una división por distritos, ofrecen una cobertura insuficiente de la red de carreteras. Es decir, para dicha división, la observabilidad a partir de los datos es mínima. Por lo tanto, se adapta la zonificación de manera que abarque el mayor área posible teniendo en cuenta la cobertura que ofrecen las espiras. Como se puede observar en la Figura 7, la zonificación elegida cuenta con cinco zonas interiores y cuatro puntos exteriores. Las zonas interiores se denominan: Deusto (TAZ 1), Uribarrí (TAZ 2), Abando (TAZ 3), Indauchu (TAZ 4) y Baseru (TAZ 5). Los puntos exteriores se representan mediante los marcadores rojos y se denominan: Exterior-Norte, Exterior-Sur, Exterior-Este y Exterior-Oeste. Los puntos exteriores se emplean para representar el flujo de tráfico proveniente o con destino en el exterior de la ciudad. La elección de los puntos exteriores se realiza escogiendo las principales carreteras de entrada y salida a las zonas interiores, procurando abarcar todos los accesos de entrada y salida de gran afluencia. De nuevo, la ubicación de las espiras, juega un papel fundamental a la hora de elegir los puntos exteriores, ya que en caso de no contar con una espira en la carretera de entrada-salida en cuestión, no es posible discriminar entre el tráfico con origen o destino en una zona interior o exterior. Por ejemplo, el punto Exterior-Oeste se emplea para representar el tráfico con origen o destino en la autopista A-8. Este tráfico es discriminable gracias a que se dispone de dos espiras (una en cada sentido) situadas en los túneles de acceso a Bilbao de San Mamés. Sin embargo, en el acceso de Enekuri (BI-604), no se dispone de espiras, por lo que no es posible diferenciar entre el tráfico con origen-destino en Deusto (TAZ 1) y el tráfico con origen-destino en el acceso de Enekuri.



Figura 7: División en zonas de la ciudad de Bilbao.

6.2. Descripción de los datos

En esta sección se introducen los datos de los que se dispone. Se describen las características principales de los mismos y se analizan los inconvenientes que acarrearán.

Los datos de los que se dispone son de tipo convencionales, es decir, datos de conteo de vehículos en determinados lugares de la red de carreteras, concretamente capturados mediante espiras. Los datos son públicos y vienen dados por parte del ayuntamiento de Bilbao. De hecho, es posible acceder a los datos en tiempo real a través de GeoBilbao¹. Las espiras ofrecen una resolución temporal cincominutal en las medidas, es decir ofrecen un dato (medido en vehículos por hora) cada cinco minutos. Aunque para el caso del problema resuelto en este proyecto, los datos de las espiras se agregan en intervalos de una hora para obtener matrices origen-destino para una hora del día en concreto. En total se dispone de 79 espiras, distribuidas a lo largo de la ciudad tal y como muestra la Figura 8. Como se ha mencionado en la sección anterior, la distribución de espiras no es la adecuada de cara a calcular matrices origen-destino por distritos o incluso por barrios. Gran parte de las espiras se encuentran ubicadas en el distrito de Abando, lo cual es lógico, ya que al tratarse del centro de la ciudad, es donde mayor afluencia de tráfico existe. Sin embargo, otros distritos están completamente desprovistos de espiras, como es el caso de Ocharcoaga y Recalde. Por otra parte se encuentran Begoña e Ibaiondo, que aunque cuentan con unas pocas espiras, están ubicadas en lugares muy concretos. Tras aplicar la zonificación introducida en la sección anterior, el número de espiras empleadas se reduce a 64.

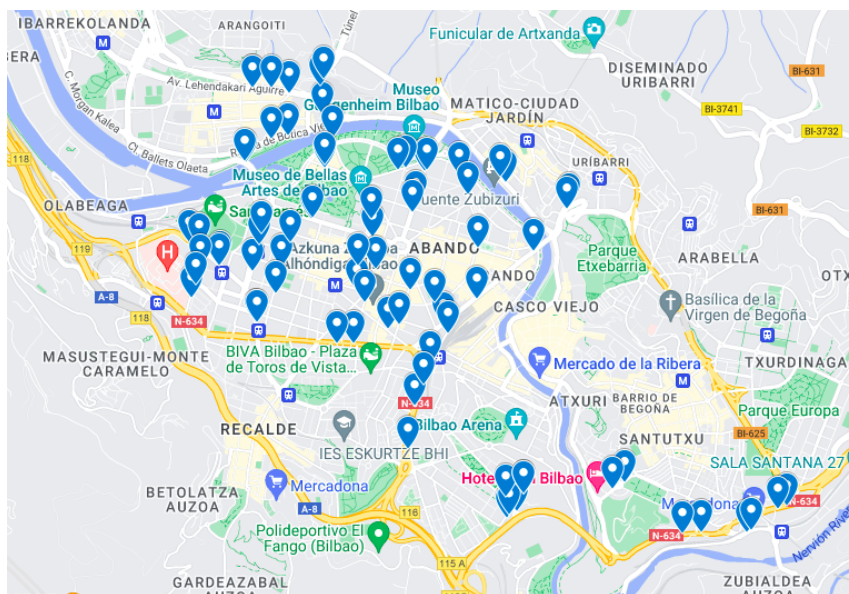


Figura 8: Ubicación de las espiras.

Otro aspecto relevante a destacar es la veracidad de los datos propiamente. Tras realizar un análisis exploratorio de los mismos, se han detectado algunas inconsistencias y valores anómalos. Por este motivo y debido a la carencia de una matriz origen-destino de referencia, el método propuesto se evalúa principalmente a partir de datos sintéticos. Este tipo de datos se introducen en el Capítulo 7. Los datos reales también se emplean en

¹GeoBilbao es un Geoportal con capas específicas de información (Mapa acústico, información en tiempo real de la ocupación de aparcamientos, de las cámaras de tráfico, de obras, de medios de transporte, del estado del tráfico, ...) [34].

la resolución, aunque con el fin de analizar si las soluciones obtenidas son consistentes con lo que a priori se espera de ellas. Por lo tanto se realiza un procesamiento de los mismos para eliminar inconsistencias y anomalías.

6.3. Método de resolución

En esta sección se presenta el método de resolución empleado en este proyecto. La sección se divide a su vez en varias secciones, en la primera se describe matemáticamente el método, y en las sucesivas se detallan en profundidad las partes fundamentales del método. Para terminar, se resume el funcionamiento del método mediante un diagrama de flujo que agrupa todos los conceptos introducidos a lo largo de la sección.

6.3.1. Descripción matemática

En esta sección se define la base matemática con la que se describe el método de resolución. Ofrece una visión general del método y a partir de ella se desarrollan las demás secciones.

Las matrices origen-destino describen el flujo de tráfico del área de estudio a nivel macroscópico. Es decir, son una representación de alto nivel de la movilidad entre las diferentes TAZ que conforman dicho área de estudio. Más concretamente, la matriz origen-destino, M , contiene en la fila i y columna j el elemento m_{ij} , el cual representa el número de viajes desde el TAZ origen i al TAZ destino j . Desde el punto de vista microscópico, cada elemento m_{ij} representa la agregación de todos los viajes individuales entre las dos TAZ. Cada uno de estos viajes, son susceptibles de pasar por una o más de las espiras que están instaladas en el área de estudio. El objetivo con este método es obtener valores de m_{ij} consistentes con las medidas obtenidas en las espiras. Es decir, por un lado cada elemento m_{ij} de la matriz OD, se puede expresar como la suma de todos los viajes individuales v , que empiezan en el punto o perteneciente al área i y acaban en el punto d perteneciente a j .

$$m_{ij} = \sum_v 1 \quad \text{donde} \quad \text{Origen}(v) \in i \quad \text{Destino}(v) \in j \quad (6.1)$$

Por otro lado, un viaje v contribuye a la medida de flujo e_k , si su trayectoria pasa por la localización de la espira k . Para representar la conectividad de cada viaje v , con cada espira k , se define la función de conectividad $C(v, k)$, que se puede expresar como:

$$C(v, k) = \begin{cases} 1 & \text{Localizacion}(k) \in \text{Trayectoria}(v) \\ 0 & \text{Localizacion}(k) \notin \text{Trayectoria}(v) \end{cases} \quad (6.2)$$

A partir de esta función, se puede calcular la medida de flujo en cada espira e_k , mediante la expresión:

$$e_k = \sum_v C(v, k) \quad (6.3)$$

En un escenario ideal, los viajes v pueden tener cualquier punto de origen y destino dentro del área de estudio. Sin embargo, en este método se propone escoger un conjunto finito de posibles orígenes y destinos que permita discretizar el problema. La hipótesis que se asume es que esta aproximación del problema, no tiene un impacto importante para resoluciones suficientemente finas. Es decir, que para suficientes puntos, pequeñas variaciones de la posición de origen-destino de un viaje no tienen impacto en el valor de m_{ij} obtenido. Esta hipótesis de trabajo debe contrastarse mediante la resolución del problema utilizando diferentes discretizaciones espaciales.

En cuanto a las trayectorias se realiza la siguiente aproximación. Se establece que entre un punto de origen o y un punto de destino d , la trayectoria es única. Es decir, se asume un método de asignación todo o nada, donde de entre todas las trayectorias posibles que unen ambos puntos se elige una sola. Concretamente se elige la que minimiza el tiempo de viaje. Esto permite transformar las sumas sobre todos los viajes posibles de las ecuaciones (6.1) y (6.3) en sumas sobre los diferentes pares de puntos origen-destino $od = (o, d)$. Las ecuaciones anteriores se convierten en las siguientes:

$$m_{ij} = \sum_{od} n_{od} \quad \text{donde } o \in i \quad d \in j \quad (6.4)$$

$$e_k = \sum_{od} C(od, k) n_{od} \quad (6.5)$$

Donde n_{od} representa el número de viajes entre un par de puntos (o, d) y $C(od, k)$ es la función de conectividad, $C(v, k)$, expresada en función del par (o, d) asociado al viaje v . El factor n_{od} , que aparece en la ecuación (6.5), surge dado que pueden existir múltiples viajes v con la misma trayectoria. Dicho factor representa el número de viajes que comparten la misma trayectoria que une los pares de puntos od . La restricción de rutado que implica que la trayectoria entre cada par (o, d) sea única, permite a su vez calcular de antemano para cada par origen-destino si su trayectoria pasa por la localización de la espira k . Es decir, que para cada par (o, d) es posible calcular su conectividad $C(od, k)$.

El sistema de ecuaciones (6.5) se puede escribir de forma matricial. Para ello se definen los siguientes vectores columna: $\bar{e} = [e_k]$, con las medidas de flujo obtenidas por la espira k ; y $\bar{n} = [n_{od}]$, que contiene en cada posición el número de vehículos que viajan entre cada par de puntos origen-destino n_{od} . Para hacer referencia a cada componente de este último vector columna solo es necesario un único índice, que se denota l . Entendiendo que existe una función biyectiva γ , que asocia a cada par od con un único l . Por ello, a partir de este momento y en lo que resta del documento, se emplea el índice l para hacer referencia a las componentes de \bar{n} . Por último, la función $C(od, k)$, que representa la conectividad, puede ser escrita en formato matricial. Concretamente, como una matriz C , en la que cada fila representa una espira, cada columna un par origen destino (o, d) , y cuyos elementos son 0 o 1. Por lo tanto, la expresión del sistema (6.5) en formato matricial es:

$$\bar{e} = C\bar{n} \quad (6.6)$$

Donde el vector \bar{n} es el conjunto de incógnitas a averiguar. A partir del cual se obtiene la matriz origen-destino solución, mediante la agregación de sus elementos en elementos m_{ij} .

El sistema de ecuaciones planteado en (6.6) se trata de un sistema lineal subdeterminado y cuenta con infinitas soluciones. Este aspecto se desarrolla con mayor detalle en la Sección 6.4.1. Para elegir una solución dentro de todas las posibles, es necesario acotar más el problema. Para ello, se introducen condiciones adicionales que brinden más información. En primer lugar, se introduce la condición de que todos los elementos del vector \bar{n} sean mayores o igual que cero.

$$n_l \geq 0, \quad l = 1, \dots, N \quad (6.7)$$

Donde N es el número de pares de puntos (o, d) . Esta condición supone que el número de viajes entre cada par de puntos origen-destino no pueda ser negativo, por lo que es una condición necesaria de cara a obtener resultados coherentes. En segundo lugar, se añade una condición de minimización para seleccionar una solución de todas las posibles. Dependiendo de la función objetivo a minimizar se alcanza una u otra solución. A continuación se exponen las dos condiciones de minimización empleadas en este método.

En primer lugar, se propone minimizar la norma L^2 del vector \bar{n} . Para un número real $p \geq 1$, se define la norma L^p del vector \bar{n} como:

$$\|\bar{n}\|_p = (|n_1|^p + |n_2|^p + \dots + |n_N|^p)^{1/p} \quad (6.8)$$

Sustituyendo en la expresión anterior por $p = 2$, se obtiene la norma L^2 :

$$\|\bar{n}\|_2 = \sqrt{\sum_{l=1}^N |n_l|^2} \quad (6.9)$$

La minimización de dicha norma, supone alcanzar la solución con el mínimo número de viajes posibles. Uniendo todas las condiciones a cumplir, el problema a resolver queda de la siguiente forma:

$$\left\{ \begin{array}{l} \text{Minimizar} \quad \|\bar{n}\|_2 \\ \text{Sujeto a} \quad n_l \geq 0, \quad l = 1, \dots, N \\ \quad \quad \quad C\bar{n} = \bar{e} \end{array} \right. \quad (6.10)$$

En segundo lugar, la condición de minimización adicional propuesta para este método es la siguiente. Se trata de minimizar la distancia de cada una de las componentes n_l del vector \bar{n} , con respecto a la media de las componentes con las que comparte zona de origen y destino.

Matemáticamente, la función a minimizar es de la forma:

$$\sum_{l=1}^N (n_l - \mathcal{A}(l))^2 \quad \text{donde} \quad \mathcal{A}(l) = \frac{1}{M} \sum_{\beta=1}^M n_{\beta} \quad \text{con} \quad \beta \sim l \quad (6.11)$$

Donde $\beta \sim l$ si y solo si el par de puntos (o, d) asociados con β y el par (o', d') asociado con l , pertenecen a las mismas zonas i (de origen) y j (de destino) correspondientemente. Es decir: $(o, o') \in i$ y $(d, d') \in j$. Y donde M es el número de pares de puntos (o, d) que comparten zona de origen y destino con el par de puntos asociados al elemento l . Empleando esta condición de minimización, se pretende que el número de viajes entre cada par de zonas origen-destino, se distribuya uniformemente entre las combinaciones de puntos origen-destino de dichas zonas. En este caso el problema a resolver queda de la siguiente forma:

$$\left\{ \begin{array}{l} \text{Minimizar} \quad \sum_{l=1}^N (n_l - \mathcal{A}(l))^2 \\ \text{Sujeto a} \quad n_l \geq 0, \quad l = 1, \dots, N \\ \quad \quad \quad C\bar{n} = \bar{e} \end{array} \right. \quad (6.12)$$

Tanto el problema planteado en (6.10) como el problema (6.12) se trata de un problema de optimización convexo [18]. Este aspecto se analiza con mayor detalle en la Sección 6.4.2. Con esta última consideración se concluye la definición matemática del método. En las siguientes secciones se detallan aspectos prácticos del método que han sido brevemente introducidos en esta sección. Como la selección de los puntos de origen-destino, la creación de la matriz de conectividad y la resolución del problema.

6.3.2. Selección de puntos de origen-destino

En esta sección se detallan todos los aspectos relativos a los puntos de origen-destino. Para qué son empleados, por qué es necesario escoger un conjunto finito, cómo se realiza la elección de sus ubicaciones, y el error asociado a la discretización de los mismos.

De cara a la resolución de los problemas planteados en (6.10) y (6.12). El primer paso es construir el sistema lineal $C\bar{n} = \bar{e}$, mediante la matriz de conectividad C , y el vector de datos en las espiras \bar{e} . Para ello, en este método se escoge un conjunto de puntos de origen-destino. Respecto al número, como se ha mencionado previamente, en un escenario ideal el número de puntos tiende a infinito, dado que cualquier ubicación es susceptible de ser un punto de origen y destino. Pero llevado a la práctica, debido a limitaciones computacionales, se discretizan en un conjunto reducido. Lo cual supone tener que elegir la ubicación de los mismos dentro del escenario.

El procedimiento de selección empleado consiste en escoger aleatoriamente las ubicaciones de los puntos. Este método, al ser automatizado, requiere de menor esfuerzo y elimina el sesgo que implica elegir los puntos manualmente. La selección se puede realizar de manera completamente aleatoria o imponiendo ciertas restricciones. Es decir, aleatoriamente pero obviando aquellas zonas con densidad de población reducida, o con escasa

oferta de servicios o empleo. Tales como barrios residenciales aislados, zonas montañosas o rurales, etc.

Otro detalle a tener en cuenta, es que a la zonificación del área de estudio se le añaden varias zonas ficticias exteriores. Estas zonas se utilizan para representar todo el tráfico con origen o destino fuera de la ciudad. Por lo que para representar dichas zonas, como se puede ver en la Figura 7, se colocan puntos de origen-destino en las principales vías de entrada y salida de la ciudad.

El hecho de escoger un conjunto finito de puntos conlleva cometer un error por discretización. Para caracterizar dicho error es necesario realizar experimentos variando tanto el número de puntos como la ubicación de los mismos. Para un número dado de puntos, el error por discretización cometido, es la diferencia entre las soluciones variando la ubicación de los puntos. Para estimar dicho error, se realiza un conjunto de experimentos con distintas ubicaciones para los puntos y se extraen las distintas matrices origen-destino solución. A partir de ese conjunto de matrices se calcula la media y la desviación estándar de cada una de las componentes m_{ij} :

$$\hat{m}_{ij} = \frac{1}{Z} \sum_{\alpha=1}^Z m_{ij\alpha} \quad (6.13)$$

$$\sigma_{m_{ij}} = \sqrt{\frac{\sum_{\alpha=1}^Z (m_{ij\alpha} - \hat{m}_{ij})^2}{Z - 1}} \quad (6.14)$$

Donde α representa el índice del experimento y Z la cantidad de experimentos realizados, o equivalentemente la cantidad de matrices origen-destino solución. El error por discretización estimado para cada una de las componentes es la desviación estándar (6.14). El error por discretización espacial total E , para un número dado de puntos, se estima mediante la agregación del error de cada componente:

$$E = \sum_i \sum_j \sigma_{m_{ij}} \quad (6.15)$$

Si el método de resolución es válido, el error por discretización debe converger a medida que el número de puntos crece. Idealmente, es decir, cuando el número de puntos tiende a infinito, la diferencia entre soluciones tiende a cero. Aunque como se ha comentado previamente, la cantidad de puntos que se pueden emplear vienen limitados por el hardware. Concretamente, el cuello de botella se encuentra en la memoria RAM de la que se dispone. Tal y como está planteado el problema, se trabaja con vectores y matrices de gran tamaño. Especialmente la matriz de conectividad, cuyo tamaño crece linealmente con el número de espiras y cuadráticamente con el número de puntos de origen-destino.

Una vez vistos los detalles sobre los puntos de origen-destino, se tiene la base para avanzar hacia el siguiente paso en la resolución del problema, que se trata de la creación de la matriz de conectividad.

6.3.3. Matriz de conectividad

En esta sección se detallan los aspectos relativos a la matriz de conectividad. Se describe tanto su estructura, como el método de construcción de la misma.

La matriz de conectividad C , como se ha mencionado previamente, tiene un número de filas igual al número de espiras y un número de columnas igual al número de pares de puntos origen-destino. Cada elemento de la matriz indica si la trayectoria que une un par de puntos pasa por la espira o no. Esta matriz se construye previamente a la resolución del problema y es dependiente de la ubicación de los puntos de origen-destino y de las espiras.

Como ya se ha descrito en la Sección 6.3.1, de todas las posibles trayectorias entre un origen y un destino se elige la que minimiza el tiempo de viaje. Esto hace posible saber por qué espiras pasan las trayectorias que unen cada par de puntos origen-destino, pudiendo pasar por ninguna, una o varias. Por lo tanto, es posible que algunas de las columnas de la matriz C estén compuestas únicamente por ceros. Es decir, que una trayectoria que une dos puntos no pase por ninguna espira, y por lo tanto no aporte información al problema. Esto supone que a la hora de resolver, el valor que se les asigna a estas componentes del vector \bar{n} , depende únicamente de la condición de minimización aplicada, sin contar con información de las espiras. Por ejemplo, en el caso de emplear la minimización de la norma L^2 , estas componentes toman un valor de cero. Por lo que en dichos casos, se elimina la columna de la matriz C y su correspondiente componente del vector \bar{n} . De esta manera se consigue ahorrar espacio en memoria RAM, lo cual permite a su vez aumentar el número de puntos origen-destino. En cuanto a las filas de la matriz C , estas deben contener al menos un elemento con valor 1. Es decir, que para cada una de las espiras exista al menos una trayectoria que pasa por ella. En caso de no cumplirse esta condición, dicha fila se elimina tanto de la matriz C como del vector de flujos \bar{e} . Es decir, la espira asociada a dicha fila se elimina.

Para la construcción de la matriz se emplea el software de rutado *Open Source Routing Machine*² (OSRM). El cual permite calcular la ruta de un vehículo entre dos puntos, dadas sus coordenadas geográficas. Por lo que el primer paso es conseguir las coordenadas geográficas (latitud,longitud) de cada uno de los puntos. La ruta calculada por el OSRM, devuelve el identificador de cada uno de los nodos OSM³ por los cuales pasa la trayectoria. Por lo que el segundo paso, es asignar a cada una de las espiras un par ordenado de nodos OSM. Es necesario que sea un par ordenado, ya que en muchas ocasiones un segmento de carretera de doble sentido viene únicamente definido por dos nodos. Por lo que el sentido de circulación del vehículo queda definido por el orden en el que atraviesa el par de nodos.

Dado que los datos de ubicación de las espiras vienen dados en coordenadas geográficas (latitud,longitud), es necesario definir un método de traducción de dichas coordenadas a nodos OSM. En este caso se emplea conjuntamente el software proporcionado por

²Open Source Routing Machine [57] es una implementación en C++ de un motor de enrutamiento de alto rendimiento para caminos más cortos en redes de carreteras. Combina sofisticados algoritmos de enrutamiento con los datos abiertos y gratuitos de redes de carreteras del proyecto OpenStreetMap (OSM).

³OpenStreetMap (OSM) [30] es un proyecto colaborativo para crear mapas editables y libres. Los mapas se crean utilizando información geográfica capturada con dispositivos GPS móviles, ortofotografías y otras fuentes libres.

Urbanite⁴, y un visualizador o editor de mapas OSM, como *Java OpenStreetMap Editor*⁵ (JOSM). El software de Urbanite permite identificar unívocamente tanto el segmento de carretera como el sentido en el que están colocadas las espiras. Y como se muestra en la Figura 9, empleando JOSM se le asocian manualmente los nodos OSM correspondientes.

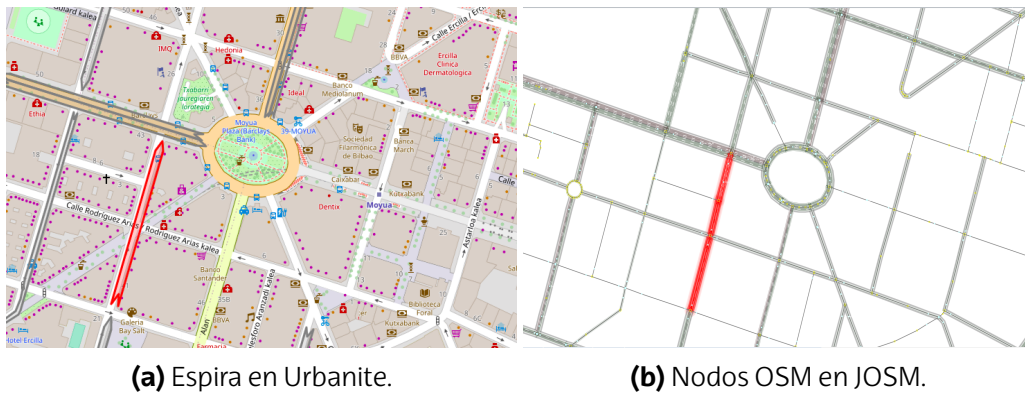


Figura 9: Asociación manual de espiras a nodos OSM.

Una vez conocidas las coordenadas geográficas de los puntos de origen-destino y teniendo la lista de pares de nodos OSM asociados a cada una de las espiras, se construye la matriz C . Para ello, se sigue el proceso ilustrado en la Figura 10 para cada uno de los pares de puntos origen-destino:

- Se emplea OSRM para calcular la trayectoria entre el par de puntos, la cual se almacena como una lista de nodos OSM.
- Los elementos de la lista se agrupan en pares de nodos, es decir, cada elemento forma un par con el elemento posterior. Formando una lista de pares de nodos asociados a la trayectoria.
- De la lista anterior se observa si algún par coincide con alguno de los pares de nodos asociados a las espiras.
- En caso de coincidir, significa que dicha trayectoria pasa por la espira o espiras en cuestión. Por lo que en la matriz C , en la fila correspondiente a la espira y en la columna correspondiente al par de puntos, se añade un 1.

⁴Urbanite [90] es un proyecto subvencionado por fondos Horizon 2020 de la Unión Europea con el acuerdo número #870338.

⁵Java OpenStreetMap Editor [78] es una aplicación de escritorio para la edición de OpenStreetMap.

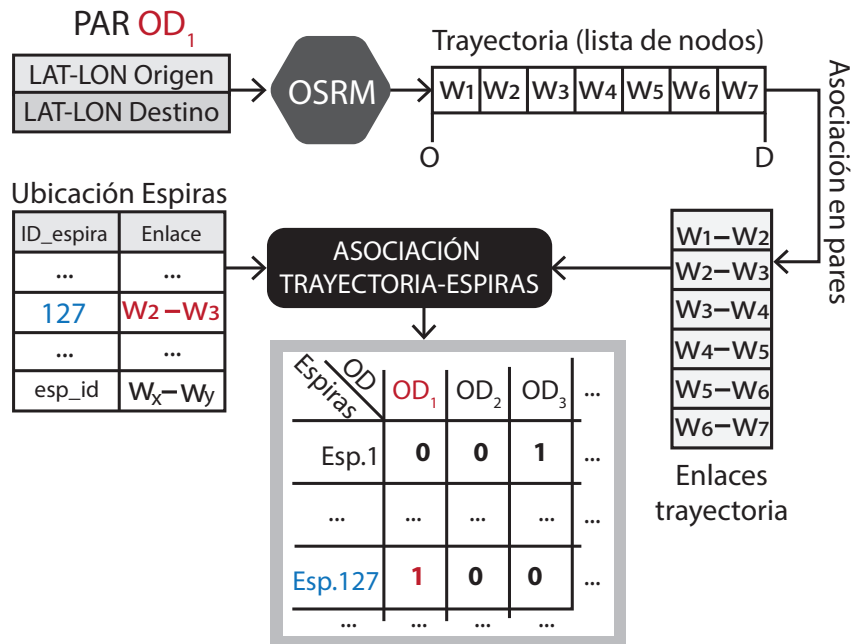


Figura 10: Proceso de construcción de la matriz C .

Una vez comentados los aspectos principales de la matriz de conectividad, queda descrito el planteamiento del método de resolución. Y es posible avanzar hacia el último apartado, donde se describe la resolución del problema.

6.3.4. Resolución del problema

En esta sección se describe la resolución del problema planteado en la Sección 6.3.1, es decir, la obtención del vector incógnita \bar{n} . Además, se detalla como a partir de dicho vector se calcula la matriz origen-destino solución.

Como se indica en la Sección 6.3.1, el problema a resolver se trata de un problema de optimización convexo. El aspecto que es susceptible a variaciones es la función de minimización, tal y como queda reflejado en (6.10) y (6.12). Además, en la Sección 6.3.1, se indica que el sistema planteado en (6.6) cuenta con múltiples soluciones. Por lo que cambiar la función de minimización implica elegir el tipo de solución. En el caso de (6.10) se escoge la solución con el menor número de vehículos posibles. En el caso de (6.12) se escoge la solución que minimiza la distancia de cada componente a la media por zona de origen y destino.

Para resolver ambos problemas se emplea el paquete de software para optimización convexa CVXOPT [60]. Esta librería ofrece distintas funciones de resolución dependiendo del tipo de problema a resolver. Concretamente, se hace uso de las funciones de *Cone Programming* [59], las cuales permiten resolver problemas con funciones de coste lineales o cuadráticas entre otras. Para la resolución de ambos problemas se emplea la función qp [61], la cual está diseñada para problemas con función de coste cuadrática. Las funciones de coste de los problemas (6.10) y (6.12) son respectivamente:

$$f_c(\bar{n}) = \sqrt{\sum_{l=1}^N |n_l|^2} \quad (6.16)$$

$$f_c(\bar{n}) = \sum_{l=1}^N (n_l - \mathcal{A}(l))^2 \quad (6.17)$$

Como se puede observar, la función de coste (6.16) no es cuadrática, por lo que se elimina la raíz para poder emplear la función de resolución *qp*. Esencialmente, eliminar la raíz no altera el objetivo de dicha función de coste, que se trata de minimizar el número de viajes. Por lo tanto, la verdadera implementación de la función (6.16), es a través de la norma L^2 al cuadrado:

$$f_c(\bar{n}) = \sum_{l=1}^N |n_l|^2 \quad (6.18)$$

En el Anexo A se desarrolla en detalle la implementación de la función *qp* para resolver ambos problemas. Una vez resueltos, se agregan los elementos del vector \bar{n} para obtener la matriz origen-destino solución. Los elementos m_{ij} de la matriz solución, se calculan mediante la suma de los elementos de \bar{n} que comparten zona de origen y destino. Matemáticamente se puede expresar mediante:

$$m_{ij} = \sum_l n_l \quad \forall l \mid (o, d) = \gamma^{-1}(l) \quad \text{con } o \in i, d \in j \quad (6.19)$$

Donde γ^{-1} es la función inversa de γ , que asocia a cada l con el par de puntos (o, d) correspondiente.

Con estas últimas consideraciones termina la descripción de la resolución del problema y con ella la descripción del método al completo.

6.3.5. Diagrama de flujo

En esta sección se reúnen los conceptos introducidos a lo largo del Capítulo 6, y se ordenan mediante un diagrama de flujo para facilitar la comprensión del método propuesto.

En la Figura 11, se resume el funcionamiento del método mediante 5 procesos. A continuación se describe brevemente cada uno de ellos, enlazando con las secciones correspondientes:

1. Zonificación: Este proceso se realiza una única vez, al comienzo, y es común a todas las resoluciones, es decir, es independiente del método de resolución. En él se divide el área de estudio en las zonas de análisis de tráfico que se requiere estudiar (Sección 6.1).

2. Selección de puntos: Es el primer paso del método de resolución y en él se seleccionan los puntos origen-destino dentro de las zonas establecidas (Sección 6.3.2).
3. Generar conectividad: Es el segundo paso del método de resolución y sirve para generar la matriz C (Sección 6.3.3). Para ello, se necesita conocer la ubicación de los puntos origen-destino y de las espiras, y se requiere de un rutador (OSRM) con un método de asignación.
4. Optimización: Una vez conocida la conectividad y contando con los datos de flujo sobre las espiras, se construye el problema matemático a resolver (Sección 6.3.1). Como hasta el momento el problema está subdeterminado (6.6), se añade más información en forma de restricciones (6.7) y de función objetivo (6.9) (6.11). El problema a resolver se trata de un problema de optimización convexo (6.10) (6.12).
5. Agregación: Una vez resuelto el problema (Sección 6.3.4), se agrega el vector solución (6.19) para obtener la matriz origen-destino solución para la zonificación establecida.

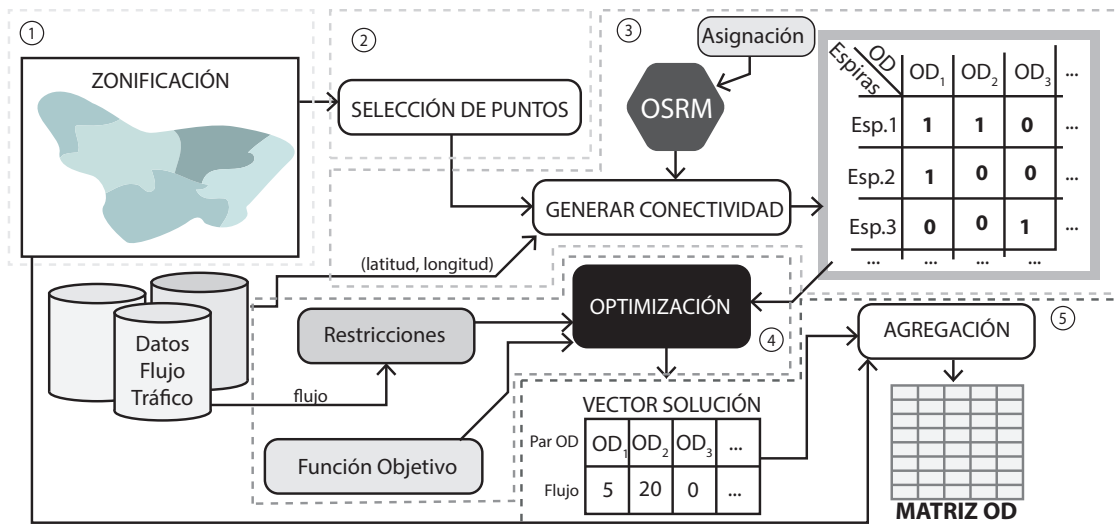


Figura 11: Diagrama de flujo del método propuesto.

6.4. Consideraciones matemáticas

En esta sección se desarrollan de manera más extensa algunas afirmaciones realizadas en la Sección 6.3.1 con respecto a la descripción matemática del método. Concretamente se estudian el sistema de ecuaciones (6.6) y los problemas de optimización convexos (6.10) y (6.12).

6.4.1. Análisis del sistema de ecuaciones

En esta sección se estudia el sistema de ecuaciones planteado en (6.6). Se analiza el tipo de sistema del que se trata y el número de soluciones que posee.

Para que el sistema de ecuaciones planteado en (6.6) tenga solución única es necesario que la matriz C sea invertible. En dicho caso los valores de \bar{n} pueden ser calculados mediante:

$$\bar{n} = C^{-1}\bar{e} \quad (6.20)$$

En este caso, la matriz C cuenta con un número mucho mayor de columnas que de filas, ya que, típicamente, el número de pares de puntos origen-destino es muy superior al de espiras. Al no tratarse de una matriz cuadrada, no es invertible. Por lo tanto el problema no cuenta con solución única.

El sistema de ecuaciones planteado en (6.6) se trata de un sistema lineal subdeterminado. Esto quiere decir que el número de ecuaciones es inferior al número de incógnitas. El número de incógnitas es igual al número de pares de puntos origen-destino. Mientras que el número de ecuaciones independientes, en el mejor de los casos, es igual al número de espiras. El número de ecuaciones independientes, depende tanto de los puntos origen-destino elegidos, como del conjunto de espiras. Un sistema lineal subdeterminado puede o bien no tener solución o tener infinitas soluciones. El primer caso se trata de un sistema inconsistente, y el segundo de un sistema consistente. El sistema planteado en (6.6), cuenta con infinitas soluciones, por lo que se trata de un sistema lineal subdeterminado consistente.

6.4.2. Análisis de la convexidad

En esta sección se describen los fundamentos de los problemas de optimización convexos. Además, se discute si los problemas (6.10) y (6.12) son de este tipo.

Un problema de optimización convexo es aquel con la siguiente forma:

$$\left\{ \begin{array}{ll} \text{Minimizar} & f_0(\bar{x}) \\ \text{Sujeto a} & f_\alpha(\bar{x}) \leq 0, \quad \alpha = 1, \dots, P \\ & A\bar{x} = \bar{b} \end{array} \right. \quad (6.21)$$

Donde $A \in R^{k \times n}$, $\bar{b} \in R^k$ y $\bar{x} \in R^n$ es el vector a optimizar. La función $f_0 : R^n \rightarrow R$ es la función objetivo y las funciones $f_\alpha : R^n \rightarrow R$ para $\alpha = 1, \dots, P$ son las funciones de restricción por desigualdad. Además se cumplen las siguientes condiciones:

- La función objetivo f_0 debe ser convexa.
- Las funciones de restricción por desigualdad f_α deben ser convexas.

La definición más básica de la convexidad de una función es la siguiente: Una función se considera convexa si el segmento de línea entre dos puntos cualesquiera de la función se encuentra por encima de la gráfica entre los dos puntos.

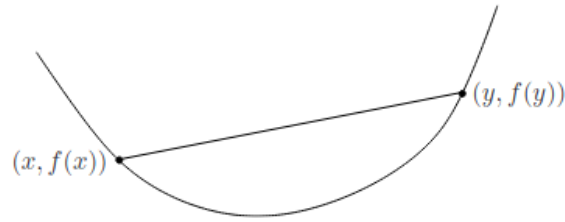


Figura 12: Ejemplo de función convexa, imagen obtenida de [17].

Matemáticamente, una función $f : R^n \rightarrow R$ es convexa, si $dom(f)$ es un conjunto convexo [19], y si para todo $x, y \in dom(f)$ y con $0 \leq \theta \leq 1$ se cumple:

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) \quad (6.22)$$

Se dice que la función f es estrictamente convexa si se cumple la inecuación (6.22) para $x \neq y$ y $0 < \theta < 1$. Para funciones afines siempre se tiene una igualdad en (6.22), por lo que todas las funciones afines (y por tanto lineales) son convexas. Las funciones afines son aquellas compuestas por la suma de una función lineal y una constante, es decir, funciones con la forma $f(x) = Ax + b$.

A continuación se analiza si el problema planteado en (6.10) cumple las condiciones planteadas en (6.21) para tratarse de un problema de optimización convexo:

- En primer lugar, la función objetivo, es decir, la norma L^2 del vector \bar{n} , se trata de una función convexa. Ya que por definición todas las normas son funciones convexas. Esto se demuestra a partir de las propiedades de homogeneidad y de la desigualdad triangular [20].
- En segundo lugar, las funciones de restricción por desigualdad son de la forma:

$$f_\alpha(\bar{n}) = -n_l, \quad l = 1, \dots, N, \quad \alpha = 1, \dots, P \quad (6.23)$$

Estas funciones, al ser lineales, cumplen las condiciones planteadas en (6.22), por lo que se tratan de funciones convexas.

- Por último, las funciones de restricción por igualdad vienen dadas por el sistema de ecuaciones $C\bar{n} = \bar{e}$. Las funciones que componen dicho sistema son afines, al igual que las del sistema $A\bar{x} = \bar{b}$.

Por otro lado, el problema planteado en (6.12) tiene una forma similar al de (6.10), pero con distinta función objetivo. En este caso, en vez de probar la convexidad del problema, se comprueba directamente la existencia de soluciones de forma empírica, es decir, resolviendo mediante el software de optimización convexo y encontrando una solución única al problema.

6.5. Fuentes de error

En esta sección se recopilan todas las fuentes de error, intrínsecas al método propuesto, identificadas.

- **Error por discretización espacial:** Este error se describe ampliamente en la Sección 6.3.2. Se trata del error cometido por el hecho de escoger un conjunto finito de puntos, de entre todos los nodos de la red, para emplearlos como puntos de origen y destino. Este error depende tanto del número de puntos escogidos como de la ubicación de los mismos.
- **Error por método de asignación:** Se trata del error cometido por el hecho de asumir un método de asignación para el tráfico. En el método propuesto, en favor de mantener la complejidad contenida, se emplea un método de asignación todo o nada. Este método, aunque es una buena aproximación, no es el más representativo de lo que ocurre en la realidad.
- **Error por condición de minimización:** Se trata del error introducido al problema por escoger una condición de minimización concreta. Tal y como se explica en la Sección 6.3.1, esta elección es necesaria de cara a tener suficiente información para resolver el problema. Aunque al mismo tiempo, se acota el espacio de soluciones a una única solución, la cual cumple una restricción impuesta que no tiene por qué corresponderse con lo que ocurre en la realidad.
- **Error por variable temporal:** Se trata del error cometido por no tener en cuenta la variable temporal. El cálculo de la matriz origen-destino estática, al no emplear la variable temporal, asume que los flujos no varían a lo largo del intervalo de estudio. Mientras que en la realidad evolucionan a lo largo del tiempo, es decir, el modelo de demanda de tráfico es dinámico.
- **Error por congestión:** Se trata del error cometido por no tener en cuenta los efectos de la congestión sobre el tráfico. Este error está relacionado con los errores por variable temporal y método de asignación, ya que, la congestión influye directamente sobre los tiempos de ruta y las decisiones de elección de ruta de los conductores. No tener en cuenta los efectos de la congestión conlleva asumir que una ruta tiene cierto coste temporal independientemente de la cantidad de usuarios, lo cual no se corresponde con la realidad. En lo relativo a la asignación, esto afecta directamente a todo tipo de asignación que tenga en cuenta el coste o la congestión a la hora de asignar rutas a los vehículos. Además, en el caso de tener en cuenta la variable temporal, la congestión afecta a los tiempos de ruta de los vehículos, y por tanto, a todo el sistema.

7. Descripción y análisis de los resultados

En este capítulo se analizan los resultados de la evaluación del método propuesto en el Capítulo 6. La evaluación se lleva a cabo tanto a partir de los datos reales descritos en la Sección 6.2, como a partir de datos sintéticos. A lo largo de la discusión de los resultados, se hace referencia a los errores planteados en la Sección 6.5.

Para evaluar el método propuesto se varían la condición de minimización, la cantidad y la ubicación de los puntos de origen-destino. Cabe recalcar que la cantidad (o número) de puntos hace referencia al número de puntos origen-destino por zona, exceptuando las zonas exteriores, las cuales se representan mediante un único punto. Por ejemplo, resolver mediante 30 puntos, hace referencia al empleo de 30 puntos en cada una de las zonas. Por otra parte, cabe mencionar que en cada resolución del problema se verifica el resultado obtenido, calculando el residuo (r) entre los datos obtenidos y los observados en las espiras, mediante:

$$r = \|\bar{e} - C\bar{n}\|_1 \quad (7.1)$$

Un resultado es solución del problema cuando r tiende a cero, es decir, cuando se cumplen las restricciones por igualdad impuestas en las espiras (6.6). Dado que las soluciones dependen de la ubicación de los puntos origen-destino, para cada número de puntos por zona (20, 30, 40), se resuelve el problema veinte veces, empleando distintas distribuciones espaciales cada vez. De esta forma, se obtiene una muestra de veinte matrices origen-destino solución del problema, para un mismo número de puntos, pero con distinta distribución espacial. Esto significa, que para un número dado de puntos (20, 30, 40), no se puede hablar de una única matriz origen-destino solución, sino de un conjunto de soluciones. Es necesario recalcar que todas las gráficas que incluyen matrices origen-destino a lo largo del capítulo, contienen los elementos de la matriz ordenados por filas y columnas a lo largo del eje horizontal. Es decir, el primer elemento en el eje horizontal de la gráfica corresponde al elemento m_{11} de la matriz, el segundo elemento corresponde al elemento m_{12} de la matriz y así sucesivamente.

7.1. Resultados para datos sintéticos

En esta sección se analizan los resultados obtenidos para datos generados artificialmente. En primer lugar, se expone el porqué de la utilización de estos datos, y el proceso de

generación de los mismos. Posteriormente se presentan, para ambas condiciones de minimización, los resultados obtenidos, y se discute acerca de ellos.

La necesidad de emplear datos sintéticos para evaluar el método surge de la falta de consistencia de los datos reales y de la carencia de una matriz origen-destino de referencia. Para evaluar el método, es necesario contar con una matriz de referencia con la cual poder comparar las soluciones obtenidas. Habitualmente se emplean matrices históricas, calculadas a partir de métodos tradicionales, como referencia para evaluar los métodos de estimación. En este caso, al no disponer de una matriz de referencia se crea una manualmente, es decir, eligiendo el número de viajes entre cada par de zonas. A esta matriz se le denomina matriz semilla, y a partir de ella, asumiendo un modelo de comportamiento de tráfico, se generan los viajes entre cada par de zonas. El modelo utilizado asume que los viajes entre zonas ocurren entre cualquier par de puntos origen-destino aleatorios, es decir, no se realiza discretización espacial. Una vez generados los viajes, y tras realizar una asignación todo o nada para determinar las trayectorias que siguen, se analiza que espiras atraviesan. Por último, se contabiliza el número de viajes que atraviesa cada una de las espiras, es decir, se crean las lecturas o datos sintéticos de las espiras. Este proceso se realiza diez veces con el fin de obtener una medida de la variabilidad del dato en cada una de las espiras. En este caso, como se puede observar en la Figura 13, las diferencias entre las medidas obtenidas para las distintas pruebas son mínimas, de hecho, para la mayoría de las espiras, los datos se solapan sobre la gráfica.

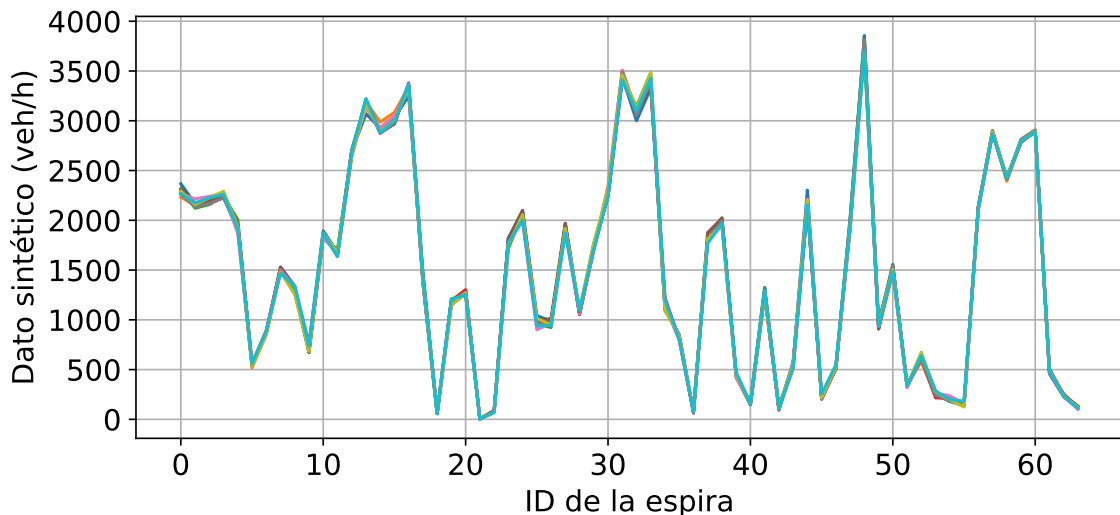


Figura 13: Diez conjuntos de datos sintéticos sobre las espiras.

En el proceso de creación de datos sintéticos, es posible que algunos de los viajes que aparecen en la matriz semilla se pierdan. Es decir, algunos de los viajes pueden tener un par de puntos origen-destino tal que no atraviesen ninguna espira, y por lo tanto el viaje se pierda a nivel de información sobre las espiras. Esto se debe a que el número de espiras con las que se cuenta es limitado. Para representar el número de viajes perdidos en cada componente de la matriz semilla se calcula la matriz reducida. Esta matriz se construye una vez generados los viajes a partir de la matriz semilla y sabiendo su correspondencia con cada una de las espiras. De manera que la matriz reducida se construye a partir de aquellos viajes que atraviesan al menos una espira. Esta matriz ofrece una visión acerca de cuantos viajes se pierden con el modelo de comportamiento de tráfico empleado. Idealmente el número de viajes perdidos debe tender a cero, ya que a la hora de realizar el proceso inverso, para calcular la matriz solución, no es posible estimar viajes que no han aportado información en las espiras. En la Figura 14, se puede observar que las matrices

reducidas obtenidas son muy similares, esto es un subproducto de lo visto en la Figura 13. Ya que, si los datos generados en las espiras son similares, los viajes perdidos también lo son y por tanto las matrices reducidas. Por otro lado, se puede observar que la cantidad de viajes perdidos con respecto a la matriz semilla no es elevada en la gran mayoría de las componentes. Las componentes de la matriz que mayor pérdida de viajes experimentan, son aquellas que hacen referencia a pares de zonas entre las que no hay espiras. Por ejemplo, entre las zonas Exterior-Norte y Exterior-Este, donde se pierden el 100 % de los viajes. Para facilitar la comprensión de la Figura 14, se unen mediante curvas los valores que toman los elementos (pares origen-destino) de cada matriz. Cabe recalcar que dichas curvas no tienen sentido más allá de su uso para facilitar la comparación visual de las matrices.

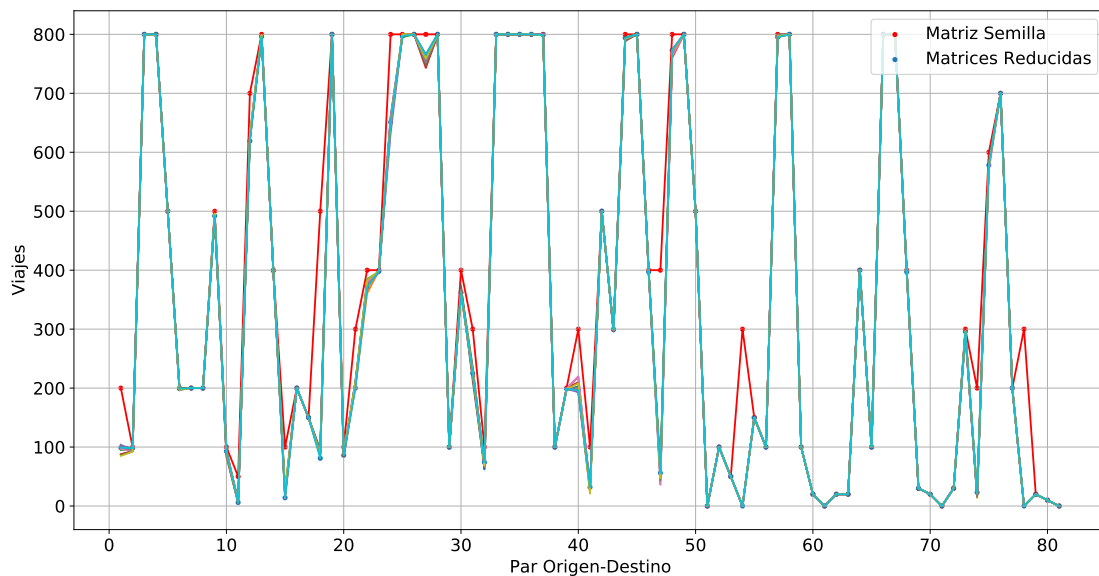


Figura 14: Matriz semilla y matrices reducidas.

Una vez introducidos los aspectos relativos a los datos sintéticos, se analizan las soluciones obtenidas para ambas condiciones de minimización y distinto número de puntos. A la hora de representar gráficamente las soluciones se emplean diagramas de cajas, ya que para cada elemento (par origen-destino) de la matriz, se cuenta con un conjunto de veinte valores (viajes). Las matrices origen-destino solución obtenidas para la minimización de la norma se muestran en las Figuras 15, 16 y 17 respectivamente. Las matrices origen-destino solución obtenidas para la minimización de la distancia a la media se muestran en las Figuras 18, 19 y 20 respectivamente. El análisis de las soluciones se divide en dos partes: en primer lugar, el estudio del error por discretización espacial; y en segundo lugar, la comparativa entre las matrices solución obtenidas y la matriz reducida.

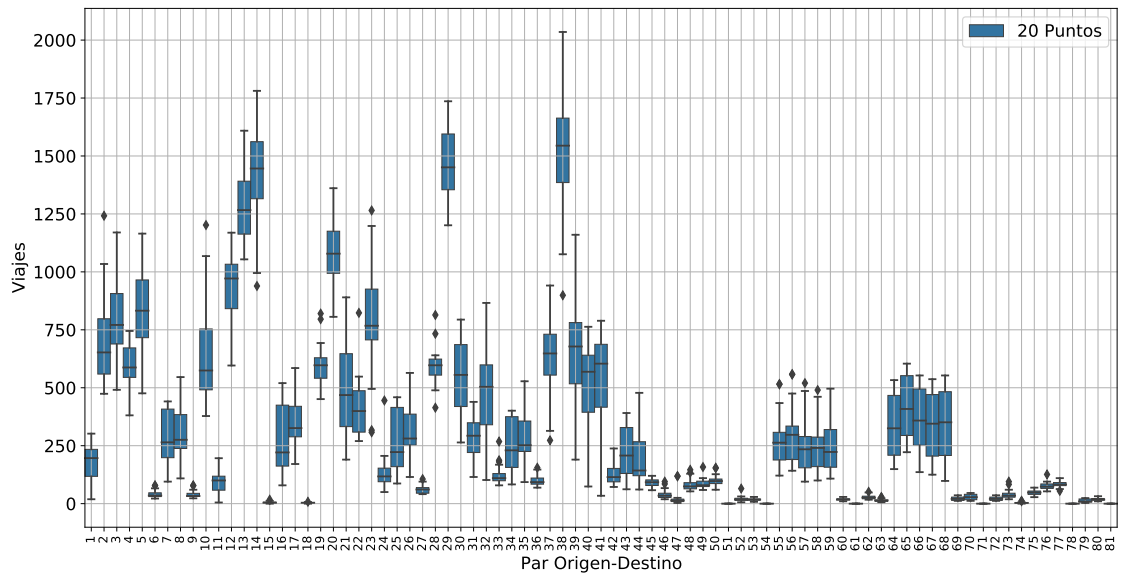


Figura 15: Matriz origen-destino solución para minimización de la norma y 20 puntos por zona.

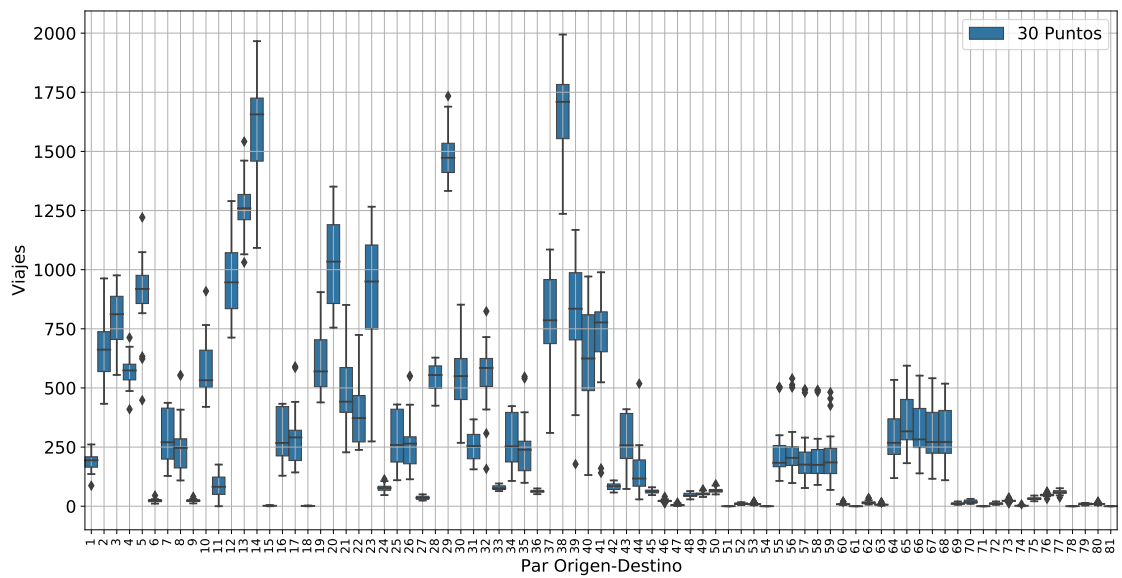


Figura 16: Matriz origen-destino solución para minimización de la norma y 30 puntos por zona.

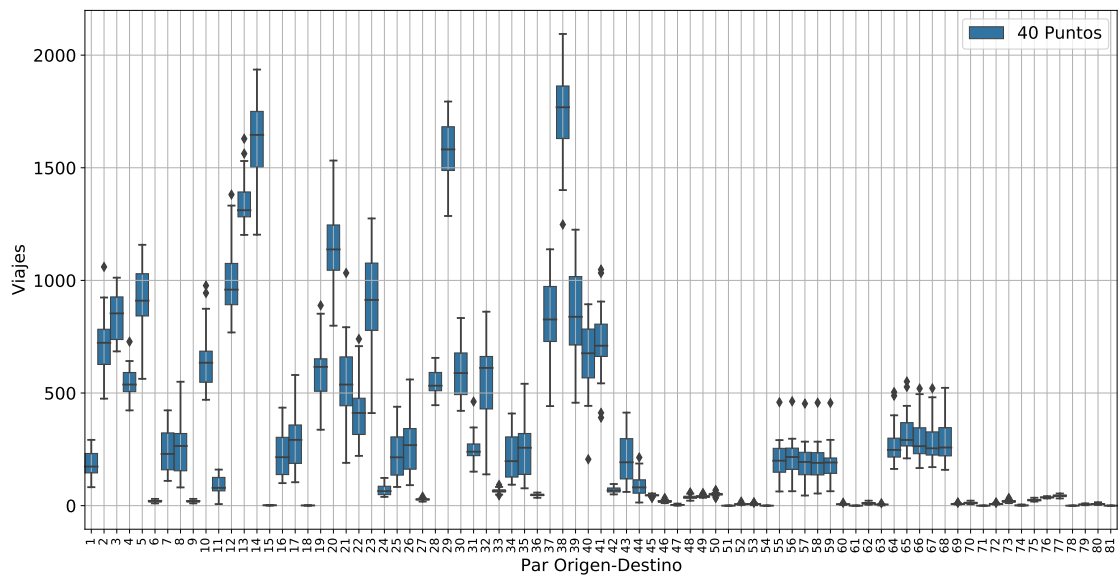


Figura 17: Matriz origen-destino solución para minimización de la norma y 40 puntos por zona.

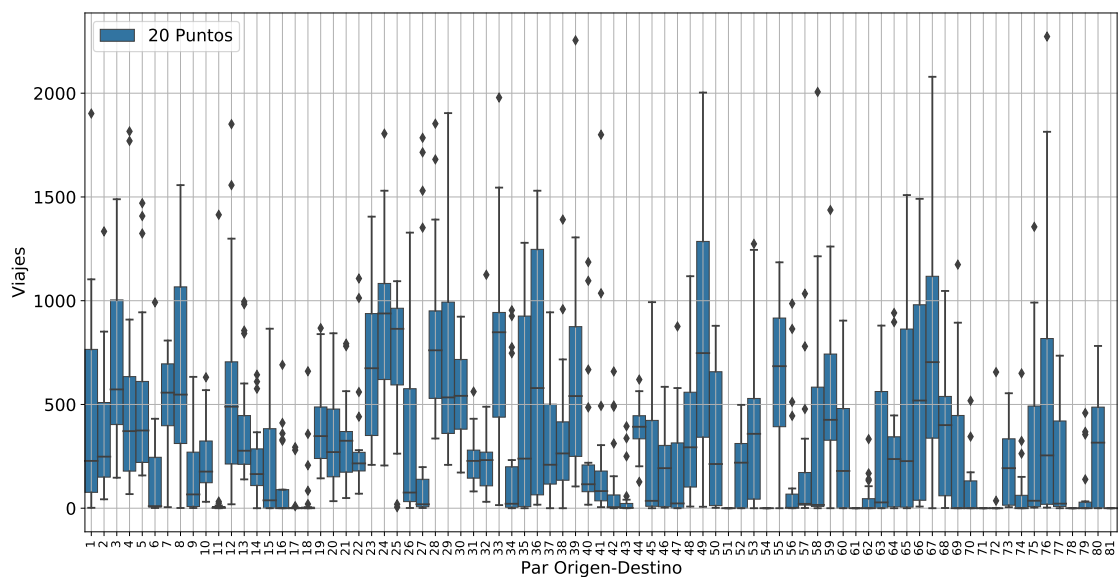


Figura 18: Matriz origen-destino solución para minimización de la distancia a la media y 20 puntos por zona.

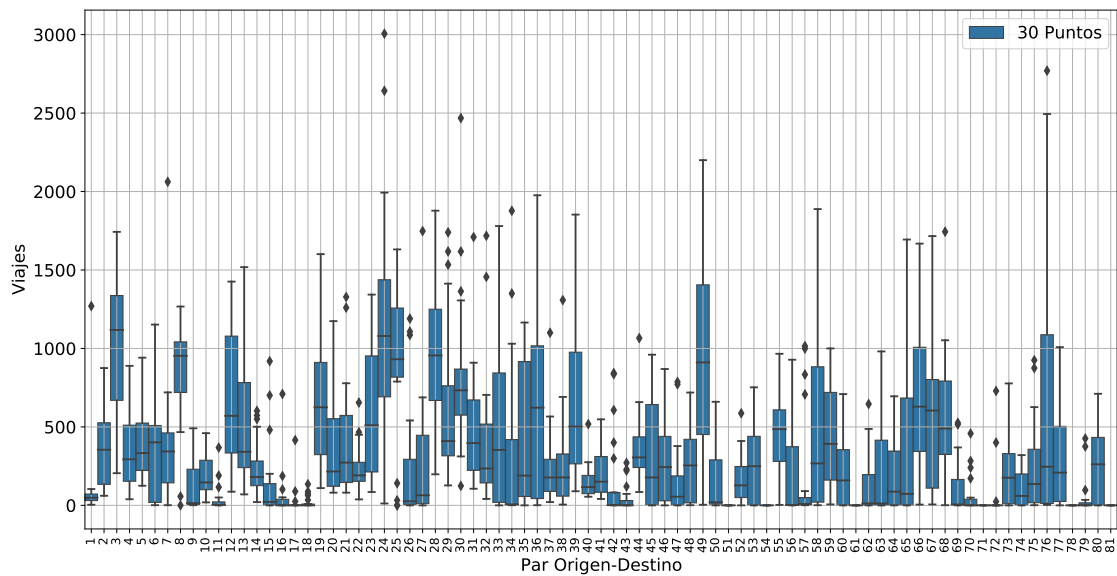


Figura 19: Matriz origen-destino solución para minimización de la distancia a la media y 30 puntos por zona.

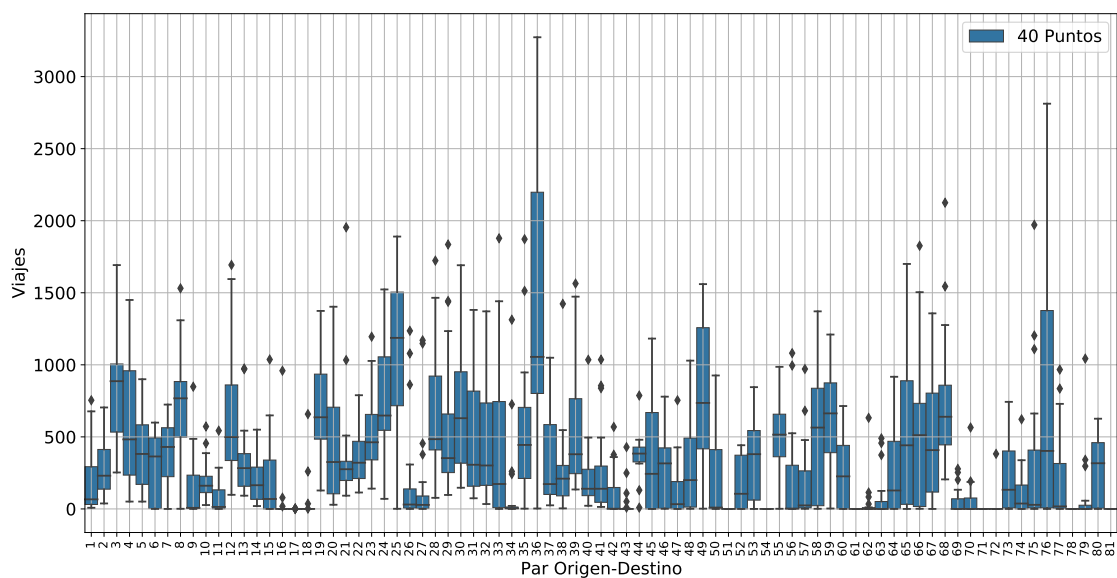


Figura 20: Matriz origen-destino solución para minimización de la distancia a la media y 40 puntos por zona.

7.1.1. Error por discretización espacial

Para caracterizar el error por discretización espacial es necesario medir las diferencias entre las soluciones. Según la hipótesis asumida, los rangos intercuartiles (RIC), de los pares origen-destino, deben disminuir (estrecharse) a medida que aumenta el número de puntos. Las figuras de las soluciones, por separado, no permiten distinguir a simple vista la veracidad de la hipótesis, por ello se introducen las Figuras 21 y 22, en las que se muestran simultáneamente las soluciones para 20 y 40 puntos.

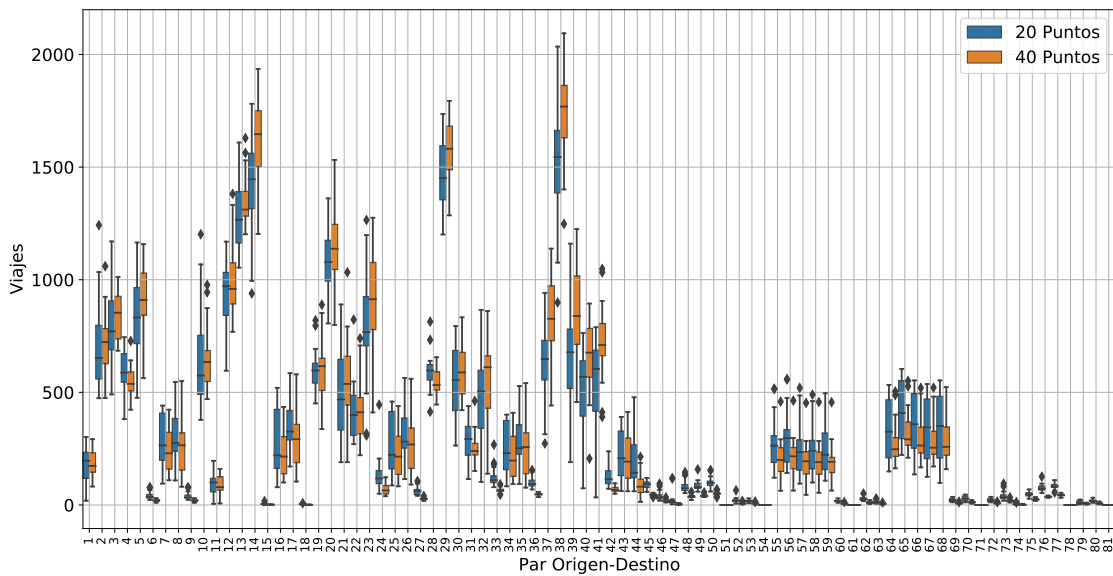


Figura 21: Matrices origen-destino solución para minimización de la norma con 20 y 40 puntos por zona.

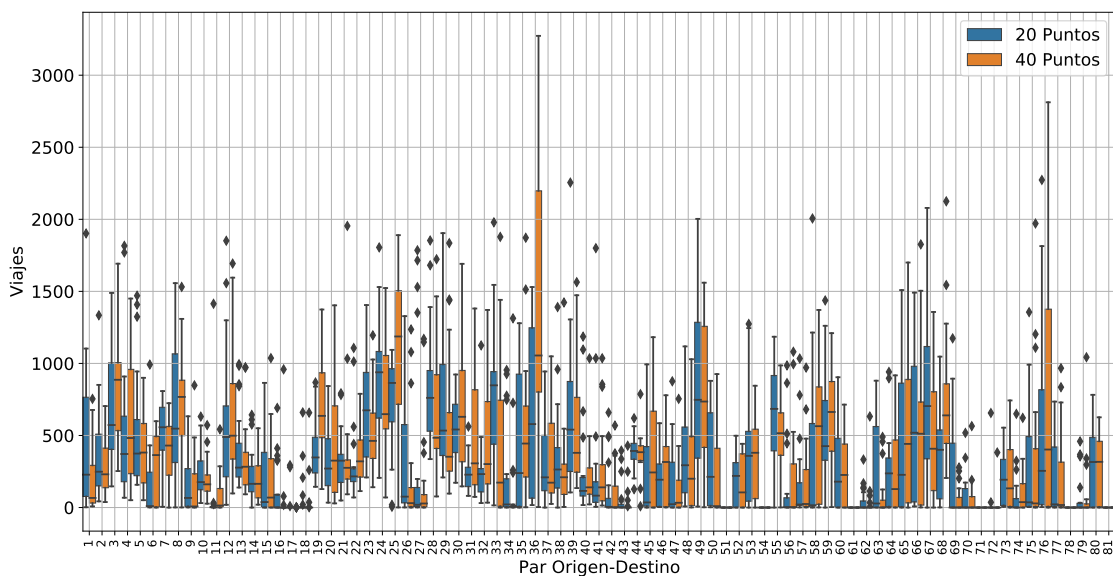


Figura 22: Matrices origen-destino solución para minimización de la distancia a la media con 20 y 40 puntos por zona.

En el caso de la Figura 21, se puede percibir una disminución en los RIC de múltiples pares origen-destino. Para caracterizar de manera más precisa el error por discretización espacial, se emplean dos métricas. Por un lado, se estima el error por discretización espacial para cada solución mediante (6.15). Por otro lado, se calcula la cantidad de pares origen-destino que disminuyen su RIC con el aumento del número de puntos.

Incremento de Puntos	RIC Minimizados	RIC Invariantes	RIC Incrementados
20 a 30	63	6	12
30 a 40	50	8	23
20 a 40	64	6	11

Tabla 2: Variación de los RIC entre las matrices origen-destino solución para minimización de la norma.

En cuanto a la estimación del error por discretización, se obtienen los siguientes valores: 7 300, para 20 puntos; 6 400, para 30 puntos; y 5 900, para 40 puntos. La variación de los RIC, en función del número de puntos, queda reflejada en la tabla 2. Esta tabla reafirma lo que indican los valores de la estimación del error, y es que parece haber una disminución global en la variabilidad de las soluciones a medida que crece el número de puntos. Por ejemplo, en la comparación entre las matrices origen-destino para 20 y 40 puntos (fila 3), el 80 % de los pares origen-destino minimizan su RIC, el 7 % permanece invariante, y el 13 % aumenta.

Para interpretar que ocurre con los pares origen-destino que no disminuyen su RIC, es necesario analizar con más detalle la tabla 2. Concretamente se analiza la tercera fila, donde se comparan las soluciones para 20 y 40 puntos, ya que a priori es donde debería ser más notable la mejora. Primero se analiza la segunda columna, la cual hace referencia a los pares origen-destino que no varían su RIC con el aumento de puntos. Para la fila 3 indica que entre 20 y 40 puntos, 6 pares-origen destino no varían su RIC. Estos 6 pares origen-destino tienen una característica en común, y es que representan pares de zonas exteriores. Cuatro de ellos representan pares de zonas exteriores consigo mismas, y por construcción del problema estos pares están forzados a tomar siempre un valor de cero viajes. Los dos pares restantes, se tratan de la misma combinación de zonas exteriores en ambos sentidos. Entre este último par de zonas no hay espiras, por lo que el número de viajes siempre es cero. Resumiendo, todos los pares que se mantienen invariantes, es debido a que siempre toman un valor de cero viajes independientemente de la cantidad y la ubicación de los puntos origen-destino. Esto explica por qué el RIC no disminuye a medida que aumenta el número de puntos.

Por otro lado, se analiza la tercera columna, es decir aquellos pares origen-destino para los cuales el RIC crece con el aumento de los puntos. En este caso ocurre para 11 pares de los 81 que conforman la matriz origen-destino. Analizando los pares a los que corresponden las subidas, se observa que cuatro tienen como destino la zona Exterior-Oeste, tres tienen como destino Deusto (TAZ 1), y cuatro tienen como origen Abando (TAZ 3). Otro aspecto a destacar es que dentro de estos once casos, no hay ningún par con origen en una zona exterior, todos los casos tienen como origen una de las zonas interiores.

Es complicado determinar a que se deben estos aumentos, aunque una de las causas puede ser que aún no se ha alcanzado un régimen de puntos para el cual se empiecen a observar mejoras significativas en el error por discretización espacial. Es decir, puede que para estas cantidades de puntos siga siendo relevante la ubicación de los mismos a la hora de obtener soluciones.

En el caso de la Figura 22, a diferencia del caso anterior, no se aprecia una mejora en los RIC a nivel global. En cuanto a la estimación del error por discretización, se obtienen los siguientes valores: 24 400, para 20 puntos; 24 600, para 30 puntos; y 24 300, para 40 puntos. Estos valores indican que la variabilidad entre las soluciones, para un mismo número de puntos, es mayor que en el caso anterior, y que el aumento del número de puntos, dentro

del régimen empleado, no reduce significativamente el error.

La variación de los RIC en función del número de puntos queda reflejada en la tabla 3. De nuevo se analiza la tercera fila, la correspondiente a la comparación de las soluciones para 20 y 40 puntos. En este caso, el 49 % de los pares origen-destino minimizan su RIC, el 9 % lo mantiene y el 42 % lo aumenta. En este caso, los pares que se mantienen invariantes son los mismos que en el caso anterior, pero añadiendo un nuevo par que une dos zonas exteriores. Por otro lado, no se identifica ningún patrón en los pares que aumentan su RIC.

Estos resultados sugieren que la condición de minimización empleada es más sensible a la ubicación de los puntos, al menos dentro del régimen de puntos con el que se trabaja. Además, los resultados refuerzan la idea de que el aumento en el número de puntos no es suficiente como para apreciar una reducción significativa en el error. Cabe mencionar, que cada una de las cinco zonas, en media, cuenta con 840 nodos (posibles puntos). Mientras que a la hora de resolver, en el mejor de los casos, se emplean 40 puntos por zona, lo cual en total representa el 4.8 % de los nodos. Por ello, puede que para estas cantidades de puntos siga siendo relevante la ubicación de los mismos a la hora de obtener soluciones, y por lo tanto el error no muestre una clara convergencia.

Incremento de Puntos	RIC Minimizados	RIC Invariantes	RIC Incrementados
20 a 30	36	8	37
30 a 40	35	7	39
20 a 40	40	7	34

Tabla 3: Variación de los RIC entre las matrices origen-destino solución para minimización de la distancia a la media.

7.1.2. Comparativa de matrices solución y reducida

En esta sección se comparan las matrices origen-destino solución obtenidas a partir de ambas condiciones de minimización con la matriz reducida. Para realizar la comparación se emplea la matriz reducida porque representa la cantidad de viajes de la matriz semilla que aporta información sobre las espiras. En el proceso inverso, es decir a la hora de calcular una solución a partir de los datos, la solución más cercana a la matriz semilla es la matriz reducida. Ya que la información perdida sobre las espiras en el proceso directo, no puede ser recuperada al calcular una solución en el proceso inverso. Las matrices solución elegidas son las calculadas a partir de 40 puntos, ya que, teóricamente son aquellas que menor error por discretización espacial introducen. Para poder representar gráficamente las veinte soluciones para cada condición de minimización, se calcula la matriz solución media. Es decir, una matriz cuyos elementos son la media de los veinte elementos correspondientes de cada una de las soluciones. La matriz reducida utilizada es una cualquiera de las diez generadas, ya que, como se ve en la Figura 14, son muy similares.

Las Figuras 23 y 24 muestran por separado las comparativas entre las soluciones con ambas condiciones de minimización y la matriz reducida. En la Figura 23 se observa que para la condición de minimización de la norma, la solución obtenida no se asemeja a la matriz reducida. Esta solución representa, dentro del espacio de soluciones, aquella que minimiza el número de viajes. Concretamente se emplean 25 000 viajes, mientras que la matriz reducida posee 28 000. A cambio, no se consigue obtener una solución que sea representativa del modelo de demanda de tráfico asumido en la matriz semilla. En la Figura 24, se observa que la solución obtenida presenta una mayor similitud con la matriz reducida

en comparación con el caso anterior. Esto es debido a que la condición de minimización es equivalente a la suposición del comportamiento del tráfico empleada en el proceso de generación de viajes sintéticos. La suposición es que los viajes parten de un lugar aleatorio de la zona de origen y terminan en un lugar aleatorio de la zona de destino. Este proceso realizado para un gran número de viajes hace que los viajes se distribuyen homogéneamente entre los puntos de origen y destino de las zonas. La condición de minimización impuesta, al tratar de reducir la distancia de cada una de las componentes del vector $\bar{\pi}$ con respecto a la media de las componentes con las que comparte zona de origen y destino, fuerza también a que la distribución de viajes sea homogénea entre los puntos. Sin embargo, el realizar la discretización espacial, hace que los viajes se distribuyan entre el conjunto de puntos escogidos y no entre todas las posibles combinaciones de la red. De esta forma se pierde parte de la afinidad con el modelo de demanda de tráfico asumido en la generación sintética, lo cual se traduce en diferencias entre la matriz solución media y la reducida.

Por último, en la Figura 25, se muestran de manera conjunta las matrices solución media, para ambas condiciones de minimización, y la matriz reducida. Las filas corresponden a las zonas de origen y las columnas a las zonas destino. La figura incluye un mapa de calor para poder analizar visualmente las matrices. Analizando la figura únicamente en términos del mapa de calor, se observa de nuevo que la solución para la condición de minimización de la distancia a la media tiene mayor semejanza con la matriz reducida.

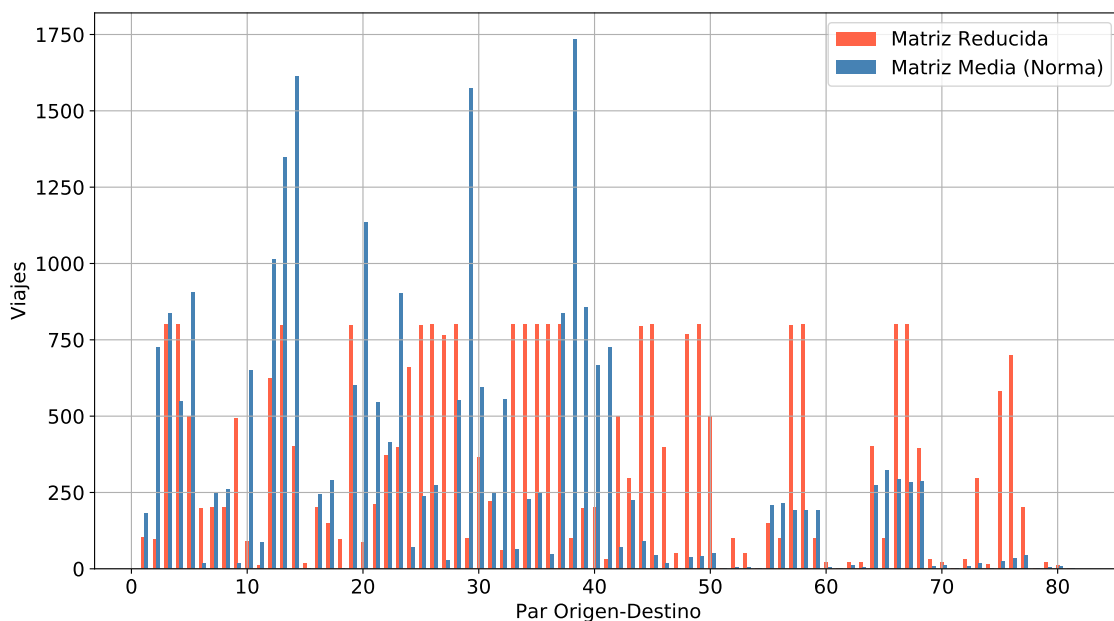


Figura 23: Matriz origen-destino solución media para minimización de la norma y matriz reducida.

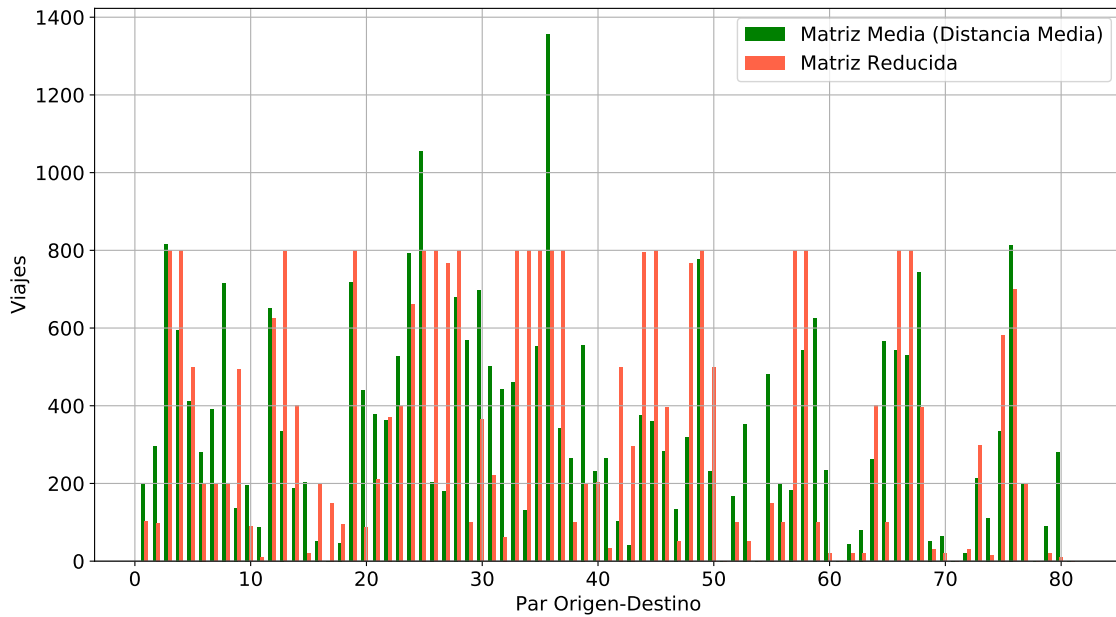


Figura 24: Matriz origen-destino solución media para minimización de la distancia a la media y matriz reducida.

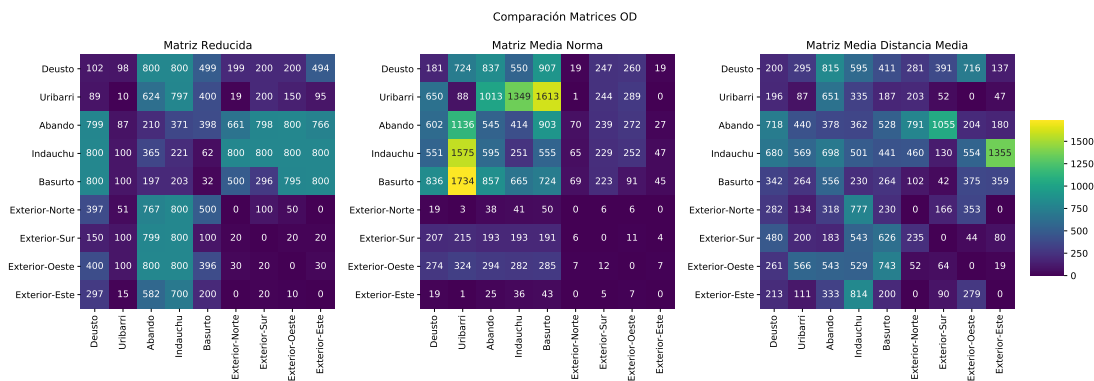


Figura 25: Matrices origen-destino solución medias para minimización de la distancia a la media (derecha) y norma (centro), y matriz reducida (izquierda).

7.2. Resultados para datos reales

En esta sección se muestran los resultados obtenidos a partir de los datos reales y se realiza un breve análisis cualitativo de los mismos. Las limitaciones de los datos explicadas en la Sección 6.2 del Capítulo 6, afectan directamente sobre las soluciones. Por lo tanto, el objetivo es ofrecer argumentos acerca de la validez o no validez de las mismas.

Las soluciones que se analizan corresponden al intervalo horario de entre las 15:00 y las 16:00 de un día laborable. Para representar ambas soluciones se calculan las dos matrices solución media, una para cada condición de minimización, y en ambos casos con 40 puntos por zona. En este caso, la solución para la norma emplea aproximadamente 22 000 viajes, mientras que la solución para la media emplea 25 000. Como se puede ver en la Figura 26, en ambas soluciones los viajes entre zonas exteriores son prácticamente nulos, lo cual puede tener sentido dados los puntos exteriores elegidos. Por otra parte, en ambas soluciones la cantidad de viajes desde zonas interiores a exteriores es reducido en

comparación con los viajes entre zonas interiores, siendo esto más acentuado en el caso para la reducción de la norma. Este aspecto probablemente es poco consistente con lo que ocurre en la realidad, ya que, en esta franja horaria pueden darse muchos desplazamientos hacia zonas exteriores, por ejemplo, por parte de trabajadores o estudiantes que terminan la jornada. Lo mismo sucede con los viajes desde zonas exteriores hacia zonas interiores, el valor es muy reducido con lo que a priori se puede esperar, o al menos demasiado desigual con respecto al número de viajes entre zonas interiores. En ambas matrices solución, la mayor cantidad de viajes suceden entre zonas interiores. En algunos casos, los viajes intrazona parecen demasiado elevados, como son los casos de Deusto, Indauchu y Abando. Donde a priori, parece razonable que para realizar un desplazamiento tan corto, la mayoría de las personas no empleen vehículos. Este razonamiento también es aplicable para los viajes entre zonas adyacentes como Basurto-Indauchu o Indauchu-Abando. En resumen, las soluciones tienden a concentrar los viajes entre zonas interiores, y a obviar los que involucran zonas exteriores. Lo cual en general parece no ser correcto, y aún menos en la franja horaria empleada.

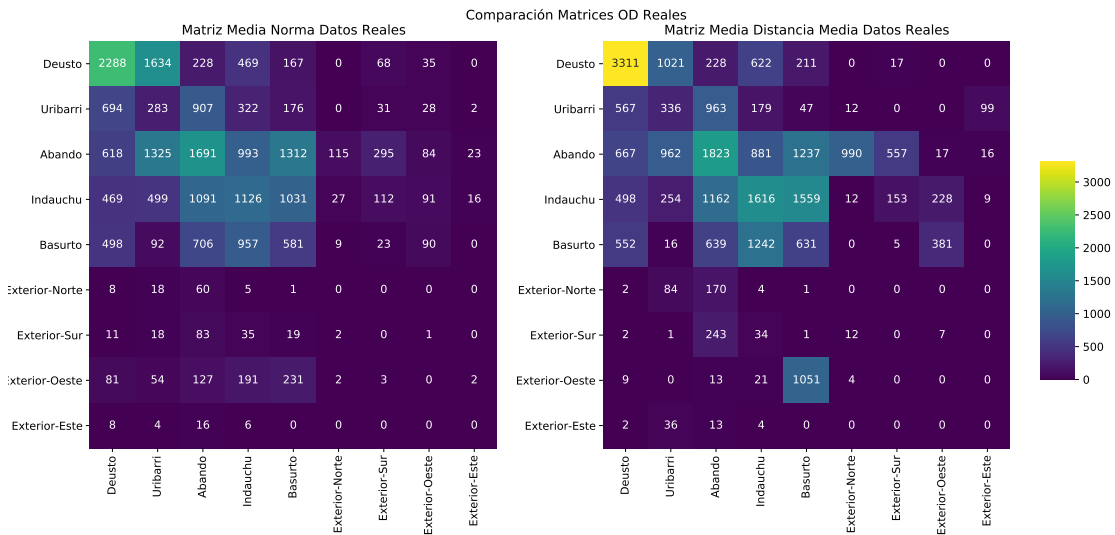


Figura 26: Matrices origen-destino solución medias para minimización de la norma (izquierda) y de la distancia a la media (derecha), para datos reales.

8. Planificación

En este capítulo se presenta la planificación del proyecto realizado, se describe el equipo de trabajo, y las fases e hitos que conforman el proyecto.

8.1. Equipo de trabajo

En este proyecto el equipo de trabajo está formado por cuatro personas. Por un lado, el director del TFM, encargado de marcar los objetivos y asesorar al alumno. Por otro lado, el tutor de empresa y el colaborador por parte del SML, los cuales, a través de su experiencia, se encargan de orientar al alumno dentro del ámbito de trabajo en concreto. Por último, el alumno, encargado de desarrollar, implementar y redactar el trabajo con la supervisión de los anteriores.

Identificador	Nombre y Apellidos	Cargo
D.1	Javier Del Ser	Director del TFM
T.1	Ibai Laña	Tutor de empresa
C.1	Iñaki Olabarrieta	Colaborador de empresa
A.1	Borja Hurtado	Alumno (Desarrollador del proyecto)

Tabla 4: Grupo de trabajo.

8.2. Fases del proyecto

A continuación, se presentan las fases del proyecto a través de los paquetes de trabajo que se han llevado a cabo, junto con las tareas que componen cada uno de ellos, y los hitos.

Paquete de trabajo 1: Propuesta de proyecto (02/02/2022 - 04/02/2022)

T.1.1 Presentación del proyecto

- *Descripción:* Presentación de la propuesta de proyecto ante los miembros del SML para su posterior discusión y reformulación.
- *Fecha de inicio:* 02/02/2022
- *Fecha final:* 02/02/2022
- *Recursos técnicos:* Ordenador, Open Office, proyector y sala de reuniones.

- *Recursos humanos*: Director (2h), Tutor de Empresa (2h), Colaborador de Empresa (2h) y Alumno (2h).

T.1.2 Definición de objetivos y alcance

- *Descripción*: Definición de los objetivos a cumplir y del alcance del proyecto.
- *Fecha de inicio*: 03/02/2022
- *Fecha final*: 04/02/2022
- *Recursos técnicos*: Ordenador, Open Office, proyector y sala de reuniones.
- *Recursos humanos*: Director (2h), Tutor de Empresa (2h), Colaborador de Empresa (2h) y Alumno (2h).

Hito (H.1): Definición del proyecto.

Paquete de trabajo 2: Formación teórica y práctica (07/02/2022 – 04/03/2022)

T.2.1 Formación teórica

- *Descripción*: Formación acerca del estado del arte de la movilidad vehicular y de la estimación de matrices origen-destino.
- *Fecha de inicio*: 07/02/2022
- *Fecha final*: 15/02/2022
- *Recursos técnicos*: Ordenador.
- *Recursos humanos*: Alumno (35h).

T.2.2 Formación práctica

- *Descripción*: Formación acerca de herramientas software, rutadores y simuladores como: OSMR, JOSM, SUMO y Docker.
- *Fecha de inicio*: 16/02/2022
- *Fecha final*: 04/03/2022
- *Recursos técnicos*: Ordenador.
- *Recursos humanos*: Alumno (65h).

Hito (H.2): Contextualización del proyecto.

Paquete de trabajo 3: Análisis y limpieza del dataset (07/03/2022 – 11/03/2022)

T.3.1 Análisis exploratorio del dataset

- *Descripción*: Análisis del conjunto de datos de espiras de la ciudad de Bilbao.
- *Fecha de inicio*: 07/03/2022

- Fecha final: 08/03/2022
- Recursos técnicos: Ordenador.
- Recursos humanos: Alumno (10h).

T.3.2 Limpieza del dataset

- Descripción: Limpieza del conjunto de datos, tratamiento de valores anómalos e inconsistencias.
- Fecha de inicio: 09/03/2022
- Fecha final: 11/03/2022
- Recursos técnicos: Ordenador.
- Recursos humanos: Alumno (18h).

Hito (H.3): Tratamiento de datos.

Paquete de Trabajo 4: Determinar el escenario de estudio (14/03/2022 - 17/03/2022)

T.4.1 Determinación de la zonificación

- Descripción: Determinación de la zonificación del área de estudio, en función de la información que se desea obtener y los datos disponibles.
- Fecha de inicio: 14/03/2022
- Fecha final: 14/03/2022
- Recursos técnicos: Ordenador.
- Recursos humanos: Alumno (4h).

T.4.2 Creación y obtención de los archivos necesarios

- Descripción: Creación de los archivos, en formato geojson, correspondientes a la zonificación. Obtención de los archivos OSM correspondientes a la red de carreteras del área de estudio.
- Fecha de inicio: 15/03/2022
- Fecha final: 17/03/2022
- Recursos técnicos: Ordenador.
- Recursos humanos: Alumno (18h).

Hito (H.4): Tratamiento de la zona de estudio.

Paquete de Trabajo 5: Enfoque inicial de resolución del problema (18/03/2022 - 18/04/2022)

T.5.1 Planteamiento teórico de la resolución

- Descripción: Planteamiento teórico inicial del método de resolución, mediante el uso del simulador SUMO, teniendo en cuenta la variable temporal y con resolución mediante algoritmo genético.

- *Fecha de inicio:* 18/03/2022
- *Fecha final:* 18/03/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

T.5.2 Implementación del método

- *Descripción:* Implementación del método de resolución inicial.
- *Fecha de inicio:* 21/03/2022
- *Fecha final:* 15/04/2022
- *Recursos técnicos:* Ordenador.
- *Recursos humanos:* Alumno (120h).

T.5.3 Análisis del funcionamiento del método

- *Descripción:* Análisis y discusión sobre el funcionamiento del método de resolución inicial.
- *Fecha de inicio:* 18/04/2022
- *Fecha final:* 18/04/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

Paquete de Trabajo 6: Enfoque intermedio de resolución del problema (19/04/2022 - 21/06/2022)

T.6.1 Replanteamiento teórico de la resolución

- *Descripción:* Replanteamiento teórico del método, mediante el uso del rutador OSRM, sin tener en cuenta la variable temporal y con resolución mediante algoritmo genético.
- *Fecha de inicio:* 19/04/2022
- *Fecha final:* 19/04/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

T.6.2 Implementación del método

- *Descripción:* Implementación del método de resolución intermedia.
- *Fecha de inicio:* 20/04/2022
- *Fecha final:* 20/06/2022
- *Recursos técnicos:* Ordenador.

- *Recursos humanos:* Alumno (264h).

T.6.3 Análisis del funcionamiento del método

- *Descripción:* Análisis y discusión sobre el funcionamiento del método de resolución intermedio.
- *Fecha de inicio:* 21/06/2022
- *Fecha final:* 21/06/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

Paquete de Trabajo 7: Enfoque final de resolución del problema (22/06/2022 - 01/11/2022)

T.7.1 Planteamiento teórico final de la resolución

- *Descripción:* Replanteamiento teórico del método, mediante el uso del rutador OSRM, sin tener en cuenta la variable temporal y con resolución mediante optimización convexa.
- *Fecha de inicio:* 22/06/2022
- *Fecha final:* 22/06/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

T.7.2 Implementación del método

- *Descripción:* Implementación del método de resolución final.
- *Fecha de inicio:* 23/06/2022
- *Fecha final:* 31/10/2022
- *Recursos técnicos:* Ordenador.
- *Recursos humanos:* Alumno (288h).

T.7.3 Análisis del funcionamiento del método

- *Descripción:* Análisis y discusión sobre el funcionamiento del método de resolución final.
- *Fecha de inicio:* 01/11/2022
- *Fecha final:* 01/11/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

Hito (H.5): Desarrollo del método de resolución.

**Paquete de Trabajo 8: Experimentos para método de resolución final
(02/11/2022 - 29/11/2022)**

T.8.1 Experimentos para datos reales

- *Descripción:* Experimentación, a partir de datos reales, empleando distinto número de puntos origen-destino por zona y distintas ubicaciones de los mismos.
- *Fecha de inicio:* 02/11/2022
- *Fecha final:* 10/11/2022
- *Recursos técnicos:* Ordenador.
- *Recursos humanos:* Alumno (42h).

T.8.2 Análisis resultados para datos reales

- *Descripción:* Caracterización del error por discretización espacial y análisis cualitativo de los resultados para los datos reales.
- *Fecha de inicio:* 11/11/2022
- *Fecha final:* 11/11/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

T.8.3 Experimentos para datos sintéticos

- *Descripción:* Experimentación, a partir de datos sintéticos, empleando distinto número de puntos origen-destino por zona y distintas ubicaciones de los mismos.
- *Fecha de inicio:* 14/11/2022
- *Fecha final:* 28/11/2022
- *Recursos técnicos:* Ordenador.
- *Recursos humanos:* Alumno (66h).

T.8.4 Análisis de resultados para datos sintéticos

- *Descripción:* Caracterización del error por discretización espacial y análisis cualitativo de los resultados para los datos sintéticos.
- *Fecha de inicio:* 29/11/2022
- *Fecha final:* 29/11/2022
- *Recursos técnicos:* Ordenador y sala de reuniones.
- *Recursos humanos:* Director (5h), Tutor de Empresa (5h), Colaborador de Empresa (5h) y Alumno (5h).

Hito (H.6): Evaluación del método de resolución.

Paquete de Trabajo 9: Redacción de la memoria (30/11/2022 - 01/03/2023)

T.9.1 Redacción del documento

- *Descripción:* Redacción de la memoria del proyecto.
- *Fecha de inicio:* 30/11/2022
- *Fecha final:* 01/03/2023
- *Recursos técnicos:* Ordenador.
- *Recursos humanos:* Director (5h), Tutor de Empresa (10h), Colaborador de Empresa (10h) y Alumno (256h).

Hito (H.7): Fin de proyecto.

8.3. Hitos

En esta sección se muestran los hitos del trabajo junto con su descripción y fecha.

Hito	Descripción	Fecha
H.1	Definición del proyecto	07/02/2022
H.2	Contextualización del proyecto	07/03/2022
H.3	Tratamiento de datos	14/03/2022
H.4	Tratamiento de la zona de estudio	18/03/2022
H.5	Desarrollo del método de resolución	02/11/2022
H.6	Evaluación del método de resolución	30/11/2022
H.7	Fin de proyecto	02/03/2023

Tabla 5: Hitos del proyecto.

8.4. Diagrama de Gantt

En esta sección se muestra el diagrama de Gantt del proyecto, repartido en tres figuras.

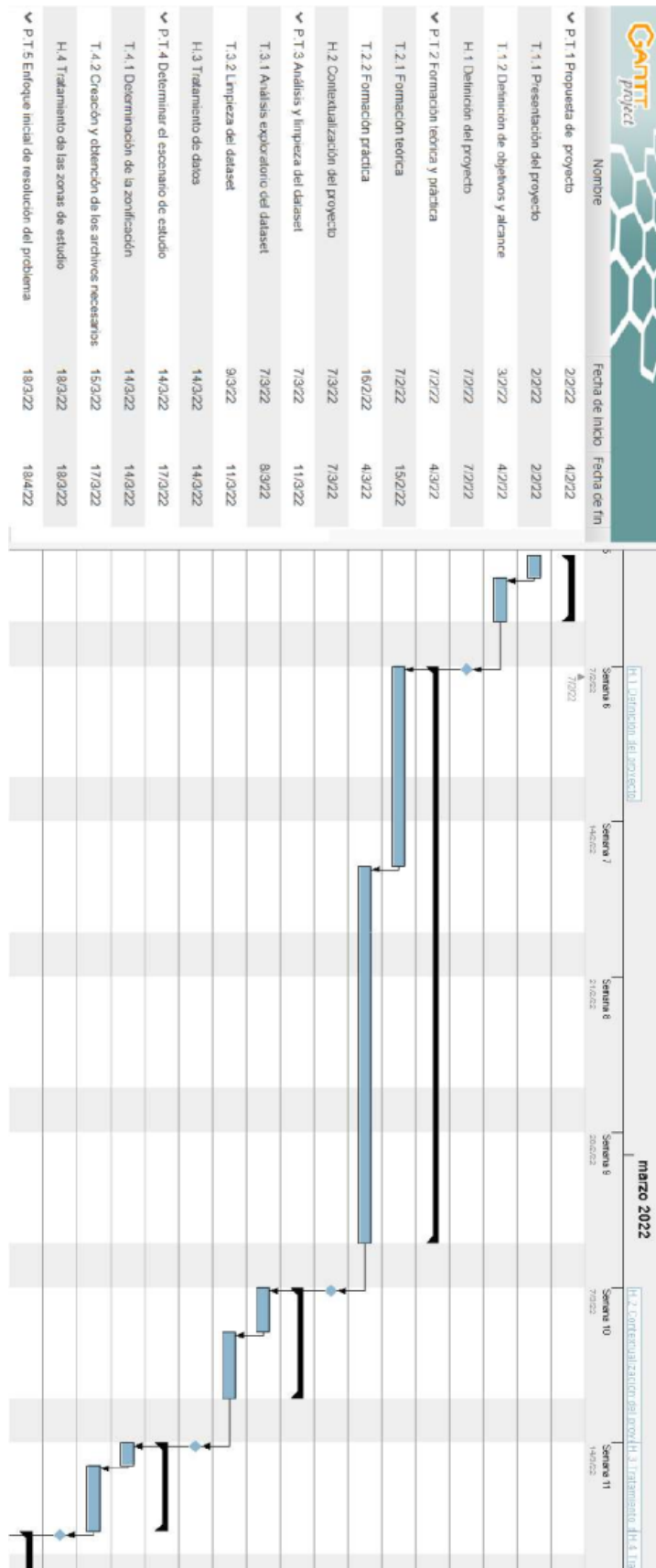


Figura 27: Diagrama de Gantt (1).

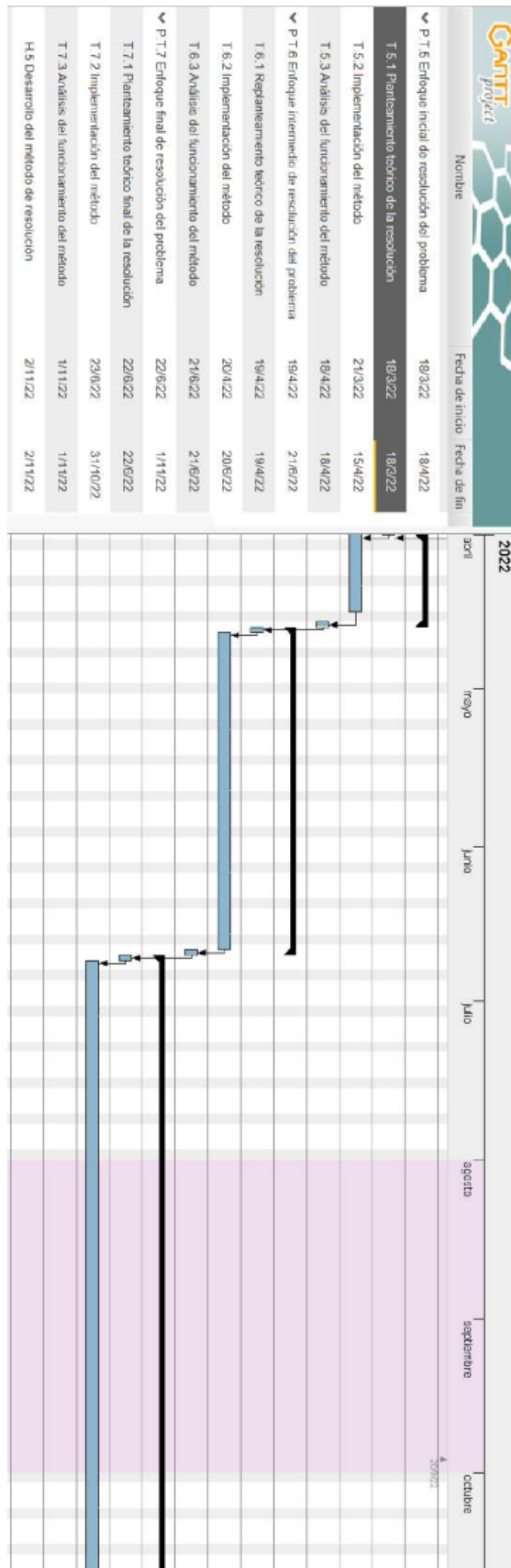


Figura 28: Diagrama de Gantt (2).

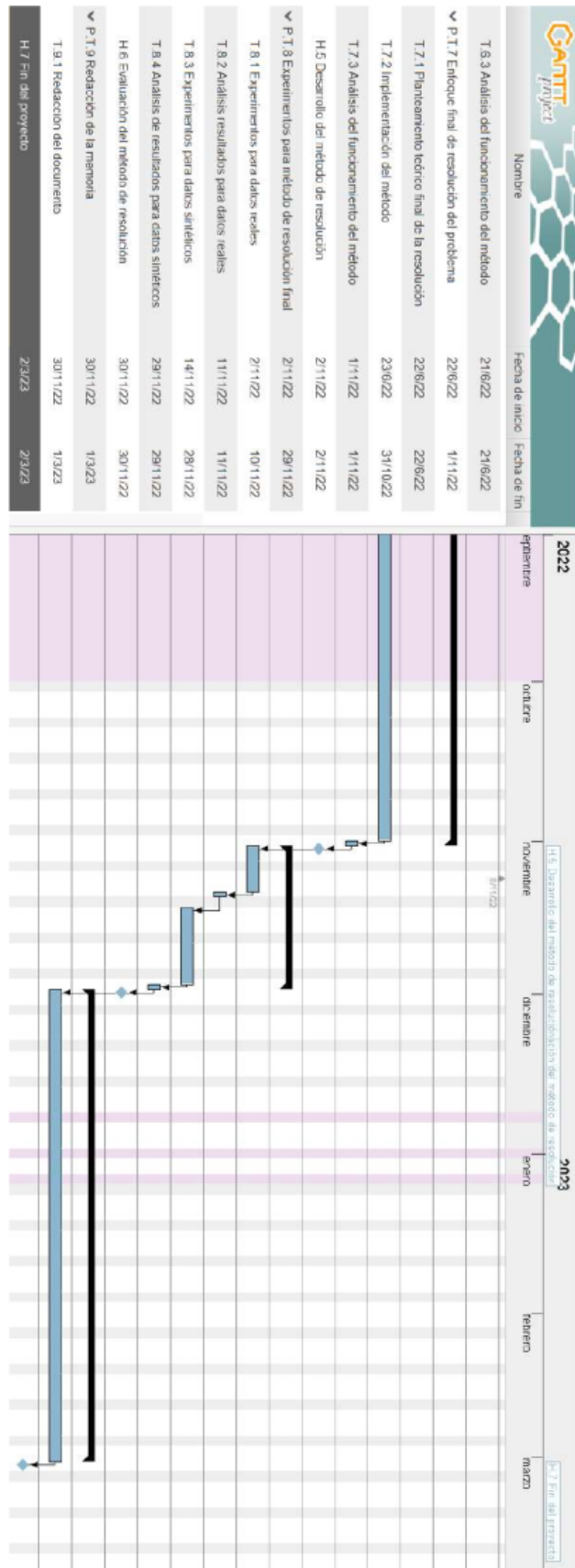


Figura 29: Diagrama de Gantt (3).

9. Presupuesto

En este capítulo se muestra el presupuesto del proyecto, el cual se desglosa en horas internas, amortizaciones y gastos.

9.1. Horas internas

En esta sección se muestra la partida de horas internas, donde se representan las horas dedicadas por parte de los trabajadores para realizar las tareas del proyecto y el coste que esto conlleva. Conociendo el coste horario y el número de horas invertidas en el proyecto, se obtiene el coste total de las horas internas de los trabajadores.

Concepto	Nº de horas	Coste horario (€/h)	Total (€)
Director del TFM (D.1)	49	150	7 350
Tutor de empresa (T.1)	54	95	5 130
Colaborador de empresa (C.1)	54	150	8 100
Alumno (A.1)	1 230	25	30 750
Subtotal			51 330

Tabla 6: Horas internas.

9.2. Amortizaciones

En esta sección se muestra la partida de amortizaciones, donde se recogen los costes de la participación de los activos fijos que se emplean a lo largo del proyecto.

Concepto	Coste (€)	Vida útil (h)	Nº de horas	Total (€)
Ordenador de A.1	800	25 000	1 230	39,36
Ordenador-1 del SML	1 200	30 000	250	10
Ordenador-2 del SML	1 200	30 000	250	10
Ordenador-3 del SML	1 200	30 000	250	10
Ordenador-4 del SML	1 200	30 000	250	10
Subtotal				69,36

Tabla 7: Amortizaciones.

9.3. Gastos

En esta sección se muestra la partida de gastos, donde se recogen los costes asociados a aspectos únicamente empleados para la realización de este proyecto.

Concepto	Coste (€)
Material de oficina	25
Gasolina	75
Subtotal	100

Tabla 8: Gastos.

9.4. Coste total

En esta sección se reúnen los costes asociados a cada una de las partidas para obtener el coste total del proyecto.

Concepto	Coste (€)
Horas internas	51 330
Amortizaciones	69,36
Gastos	100
Subtotal	51 499,36
Imprevistos	10 %
Total	56 649,30

Tabla 9: Coste total.

10. Conclusiones y trabajo futuro

En este capítulo se describen las conclusiones que se extraen del proyecto y se proponen algunos aspectos a desarrollar e implementar como continuación al trabajo realizado.

En este proyecto se diseña e implementa un procedimiento novedoso para la estimación de matrices origen-destino consistentes con las medidas de los sensores de flujo instaladas sobre la red de carreteras. En la formulación del problema se identifica la necesidad de incorporar una condición de minimización añadida para que el sistema de ecuaciones planteado en (6.6) tenga solución única. El método utiliza dicha condición para añadir información sobre el comportamiento o las características del tráfico dentro de la ciudad. Se demuestra que la elección de dicha condición de minimización tiene un impacto drástico en la solución obtenida. Un aspecto interesante para estudiar a futuro es la posibilidad de incorporar nueva información al problema, de cara a encontrar una solución que tenga más sentido en un escenario realista.

En cuanto a la resolución para datos reales, no se consigue ofrecer una solución con sentido, en gran parte debido a la falta de veracidad de los datos y a las inconsistencias que muestran entre sí. Debido a estas limitaciones en los datos, la batería de pruebas se realiza empleando datos sintéticos, lo cual permite la comparación de la solución con un dato de referencia. Los experimentos se llevan a cabo variando el número de puntos y las localizaciones de los mismos, para poder evaluar el error por discretización del método. Los resultados de estos experimentos no son concluyentes, debido a las limitaciones en el hardware del que se dispone. Por otra parte, contar con una mayor cantidad de espiras, sobre todo en aquellas zonas donde no se dispone de ellas, hubiera permitido aplicar el método propuesto en un área geográfica más amplia, aumentando su posible utilidad.

Como complemento a la formulación e implementación del método propuesto, se proponen a continuación varias mejoras para extenderlo y hacer frente a algunas de las fuentes de error presentes en él.

10.1. Matriz OD dinámica

En esta sección se teoriza acerca de la incorporación de la variable temporal al problema. Se presenta la descripción matemática de un posible método, sin entrar en consideraciones prácticas acerca de su implementación. Incluir la variable temporal significa resolver el problema dinámico, es decir, la solución es una serie temporal de matrices OD (o matriz dinámica), que se ajustan a un modelo de demanda de tráfico dinámico.

En primer lugar, se define el tiempo total de estudio como T , el cual se divide a su vez en intervalos temporales ΔT . Teóricamente, para captar cualquier variación en los flujos

de tráfico a lo largo del tiempo, la duración de los intervalos debe tender a cero. Aunque llevado a la práctica la duración de los intervalos queda fijada por la resolución temporal de las medidas que ofrecen los sensores empleados.

Lo que se propone para la resolución del problema dinámico es el empleo del método estático para cada uno de los intervalos ΔT , teniendo en cuenta la duración de los viajes. Es decir, para el intervalo ΔT , se tienen en cuenta aquellos viajes que generan mediciones en las espiras dentro de dicho intervalo, y viajes de intervalos anteriores, que debido a la duración de sus rutas, generan datos sobre las espiras en este intervalo temporal. Una aproximación que se realiza, con el fin de contener la complejidad, es suponer que todos los viajes que ocurren en el intervalo ΔT , inician su ruta al comienzo de dicho intervalo. Por ejemplo, en el intervalo ΔT , entre los instantes de tiempo $t = 0$ y $t = 15$ minutos, todos los viajes parten en el instante $t = 0$ minutos. De manera que cuando ΔT tiende a cero, la aproximación es más representativa de lo que ocurre en la realidad. Aunque de nuevo, llevado a la práctica, el valor de ΔT queda limitado por la resolución temporal de los datos en las espiras.

El problema, al igual que en el caso estático, se plantea como un problema de optimización convexo. Los únicos aspectos que varían respecto a los sistemas planteados en (6.10) y (6.12), son la forma y cantidad de funciones de restricción por igualdad. En este caso, las funciones de restricción por igualdad a satisfacer son:

$$\bar{e}^{(t)} = \sum_{k=1}^{\infty} C^{(k\Delta T)} \bar{n}^{(t-(k-1)\Delta T)} \quad (10.1)$$

Donde:

- $\bar{e}^{(t)}$ representa las medidas de flujo en las espiras en el intervalo temporal entre t y $t + \Delta T$.
- $C^{(k\Delta T)}$ es la matriz de conectividad para los viajes que, desde su origen, requieren de un tiempo comprendido entre $(k-1)\Delta T$ y $k\Delta T$, para atravesar cierta espira.
- $\bar{n}^{(t-(k-1)\Delta T)}$ representa el vector \bar{n} para el intervalo de tiempo entre $t - (k-1)\Delta T$ y $t - (k-2)\Delta T$. Para $k = 1$, se obtiene el vector $\bar{n}^{(t)}$, que hace referencia al vector \bar{n} incógnita en el instante t .

Para facilitar la comprensión de la expresión anterior, se introduce a continuación un ejemplo práctico llevado al escenario de estudio de este proyecto. En primer lugar, se fija la duración de los intervalos ΔT a 15 minutos. En segundo lugar, se asume que el tiempo máximo de ruta entre dos puntos origen-destino cualesquiera del área de estudio (Bilbao) es de 30 minutos. Esto implica que desde que un vehículo sale de su origen hasta que atraviesa una espira cualquiera, puede haber una duración máxima de 30 minutos. Con estas condiciones, las funciones de restricción por igualdad planteadas en (10.1) quedan de la forma:

$$\bar{e}^{(t)} = C^{(\Delta T)} \bar{n}^{(t)} + C^{(2\Delta T)} \bar{n}^{(t-\Delta T)} \quad (10.2)$$

Donde:

- $C^{(\Delta T)}$ es la matriz de conectividad para los viajes que desde su punto de origen, requieren de un tiempo comprendido entre 0 y 15 minutos, para atravesar cierta espira.
- $\bar{n}^{(t)}$ es el vector \bar{n} incógnita en el instante t . A partir del cual se construye la matriz OD para el intervalo entre t y $t + \Delta T$.
- $C^{(2\Delta T)}$ es la matriz de conectividad para los viajes que desde su punto de origen, requieren de un tiempo comprendido entre 15 y 30 minutos, para atravesar cierta espira.
- $\bar{n}^{(t-\Delta T)}$ es el vector \bar{n} del instante $t - \Delta T$. A partir del cual se construye la matriz OD para el intervalo entre $t - \Delta T$ y t .

A priori, puede parecer que el sistema planteado en (10.2), cuenta con el doble de incógnitas que un problema estático al uso, ya que para cada intervalo ΔT cuenta con dos vectores \bar{n} independientes. Realmente no es así, ya que el vector $\bar{n}^{(t-\Delta T)}$ se puede obtener recurriendo al paso temporal anterior, mediante el sistema:

$$\bar{e}^{(t-\Delta T)} = C^{(\Delta T)}\bar{n}^{(t-\Delta T)} + C^{(2\Delta T)}\bar{n}^{(t-2\Delta T)} \quad (10.3)$$

Aquí de nuevo, el cálculo de $\bar{n}^{(t-\Delta T)}$ requiere de conocer el valor del vector \bar{n} que le precede temporalmente, es decir, $\bar{n}^{(t-2\Delta T)}$. Este inconveniente se traslada hasta el primer paso temporal del tiempo total de estudio T , donde es necesario asumir información adicional para el problema. Por ejemplo, si el tiempo total de estudio T , comprende desde las 4 hasta las 12 de la mañana, es razonable asumir que el flujo desde las 3:45 a las 4:00 es cero. Expresado matemáticamente:

$$\bar{e}^{(4:00)} = C^{(\Delta T)}\bar{n}^{(4:00)} + C^{(2\Delta T)}\bar{n}^{(3:45)} \rightarrow 0 \quad (10.4)$$

Por otro lado, si el tiempo total de estudio T , no incluye ningún intervalo ΔT donde sea coherente realizar la suposición anterior, otra condición de contorno asumible es que el flujo en el intervalo $(k - 1)\Delta T$ es igual al flujo en el intervalo $k\Delta T$. Expresado matemáticamente:

$$\bar{e}^{(t)} = C^{(\Delta T)}\bar{n}^{(t)} + C^{(2\Delta T)}\bar{n}^{(t-\Delta T)} \rightarrow \bar{n}^{(t)} \quad (10.5)$$

En los dos ejemplos anteriores se introduce la información adicional en forma de condición de contorno previa al comienzo temporal del problema, aunque también se puede introducir en forma de información posterior al final del problema. De hecho, no es necesario que se trate de una condición de contorno, se puede introducir información adicional en cualquier intervalo temporal del problema. Por otro parte, el ejemplo anterior solo

requiere de asumir información adicional acerca del valor de un vector \bar{n} . Esto es producto de la suposición de que el tiempo máximo de ruta entre dos puntos origen-destino cualesquiera del área de estudio es de 30 minutos y el intervalo ΔT es de 15 minutos. Sin embargo, a medida que aumenta el cociente entre el tiempo máximo de viaje y la duración del intervalo ΔT , aumenta la cantidad de información que es necesario introducir.

10.2. Congestión

En esta sección se discute acerca de la incorporación de la congestión al problema, tanto para la versión estática planteada en la Sección 6.3, como para la versión dinámica planteada en la Sección 10.1.

10.2.1. Problema estático

El problema estático asume que los flujos entre orígenes y destinos no varían a lo largo del tiempo de estudio, no tiene en cuenta los tiempos de ruta ni el efecto de la congestión sobre los mismos. Sin embargo, para este tipo de problema, es posible tener en cuenta la congestión a través del método de asignación. Es decir, tener en cuenta la capacidad de los enlaces a la hora de asignar trayectorias a los viajes, ya que de lo contrario, se asume que los enlaces tienen capacidad infinita. De esta forma, y aunque la capacidad de los enlaces no repercute sobre los tiempos de viaje, se busca realizar una asignación que no implique sobrepasar la capacidad máxima de los enlaces.

Llevado al caso del método propuesto en la Sección 6.3, se plantea un nuevo método de asignación que introduce la congestión, como alternativa a la asignación todo o nada. El nuevo método sigue un principio similar al de la asignación proporcional, en el cual se elige una trayectoria entre un conjunto de trayectorias cercanas a la óptima. En este caso el conjunto entre el cual se elige consiste en el formado por la trayectoria más corta y todas las trayectorias entre origen y destino que atraviesan alguna de las espira. La elección del conjunto se realiza de esta manera debido a cuestiones del funcionamiento de OSRM. El conjunto de trayectorias no tiene por qué incluir únicamente las más cercanas a la óptima, aunque para evitar inconsistencias, es necesario filtrar las trayectorias con coste muy superior a la más corta. La asignación de la proporción de viajes para cada una de las trayectorias del conjunto se realiza, a priori, en función del coste temporal. Concretamente, de forma inversamente proporcional al coste temporal, es decir, se asigna una mayor proporción de viajes a aquellas trayectorias con menor coste temporal, lo cual es más representativo de la realidad. Llevado a la práctica, esto se traduce en realizar modificaciones sobre la matriz de conectividad C . En este nuevo planteamiento del problema, la matriz pasa de estar compuesta por ceros y unos a estar compuesta por ceros y números decimales entre cero y uno.

Una vez construida la matriz de conectividad, se resuelve el problema tal y como se plantea en la Sección 6.3. A partir del vector \bar{n} solución y conociendo las proporciones de asignación de cada trayectoria, es posible calcular el flujo en cada uno de los enlaces. Una vez calculado, se compara el flujo medido con la capacidad máxima teórica de cada enlace de la red, para identificar los lugares en los que hay congestión. Una vez identificados, se reajustan las proporciones de asignación de las trayectorias con el fin de minimizar la congestión en dichos enlaces. Es decir, el método propuesto, trata de realizar un proceso iterativo que adapte las proporciones de asignación con el objetivo de

minimizar la cantidad de enlaces con congestión. Este proceso puede ser implementado, por ejemplo, mediante un algoritmo genético.

10.2.2. Problema dinámico

En el problema dinámico, se asume que los flujos entre orígenes y destinos varían a lo largo del intervalo de estudio, y se tienen en cuenta los tiempos de ruta a la hora de resolver el problema. Por lo que el efecto de la congestión, se puede introducir tanto en el proceso de asignación, como a la hora de cuantificar los retrasos en los tiempos de ruta.

Gracias a la variable temporal, el método de asignación con congestión, es más representativo de la realidad en el problema dinámico que en el estático. Ya que, en este último, se calcula el flujo de vehículos que atraviesan un enlace a lo largo de todo el tiempo de estudio, pero no es posible determinar cómo se distribuye dicho tráfico a lo largo del tiempo. En cambio, en el problema dinámico es posible captar las variaciones en el flujo y por tanto en la congestión de los enlaces a lo largo del tiempo.

En cuanto al modelado de los retrasos en los tiempos de ruta, por lo general, lo más habitual es el empleo de simuladores de tráfico como SUMO [11], Vissim [92], etc. Gracias a su capacidad para reproducir situaciones de tráfico realistas, es posible obtener una buena aproximación de los retrasos sufridos en los desplazamientos debido a la congestión. Esto llevado al problema dinámico, teorizado en la Sección 10.1, se traduce en una traslación de trayectorias a distintas matrices de conectividad $C^{(k\Delta T)}$.

Bibliografía

- [1] European Environment Agency. Outdoor air quality in urban areas. <https://www.eea.europa.eu/airs/2018/environment-and-health/outdoor-air-quality-urban-areas>, November 2018.
- [2] Constantinos Antoniou, Loukas Dimitriou, and Francisco Pereira. *Mobility patterns, big data and transport analytics: tools and applications for modeling*. Elsevier, 2018.
- [3] Masanao Aoki. *State space modeling of time series*. Springer Science & Business Media, 2013.
- [4] Yasuo Asakura, Eiji Hato, and Masuo Kashiwadani. Origin-destination matrices estimation model using automatic vehicle identification data and its application to the han-shin expressway network. *Transportation*, 27:419–438, 2000.
- [5] Kalidas Ashok. Dynamic origin-destination matrix estimation and prediction for real-time traffic management system. In *12th International Symposium on Transportation and Traffic Theory, 1993*, pages 465–484, 1993.
- [6] Kalidas Ashok and Moshe E Ben-Akiva. Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows. *Transportation science*, 34(1):21–36, 2000.
- [7] Danya Bachir, Ghazaleh Khodabandelou, Vincent Gauthier, Mounim El Yacoubi, and Jakob Puchinger. Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101:254–275, 2019.
- [8] Seungkil Baek, Hyunmyung Kim, and Yongtaek Lim. Multiple-vehicle origin–destination matrix estimation from traffic counts using genetic algorithm. *Journal of Transportation Engineering*, 130(3):339–347, 2004.
- [9] Jaume Barcelö, Lidin Montero, Laura Marqués, and Carlos Carmona. Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. *Transportation research record*, 2175(1):19–27, 2010.
- [10] Jaime Barceló Bugada, Lidia Montero Mercadé, Manuel Bullejos, Oriol Serch, and Carlos Carmona Bautista. Dynamic od matrix estimation exploiting bluetooth data in urban networkss. In *Proceedings of the International Conference*, pages 116–121, 2012.
- [11] Michael Behrisch and Robert Hilbrich. Simulation of urban mobility. <https://www.eclipse.org/sumo/>, February 2023.

- [12] Michael GH Bell. The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research Part B: Methodological*, 25(1):13–22, 1991.
- [13] Michael GH Bell. The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research Part B: Methodological*, 25(1):13–22, 1991.
- [14] Sharminnda Bera and KV Rao. Estimation of origin-destination matrix from traffic counts: the state of the art. *European Transport\ Trasporti Europei*, (49):2–23, 2011.
- [15] Lucio Bianco, Giuseppe Confessore, and Pierfrancesco Reverberi. A network based model for traffic sensor location with implications on o/d matrix estimates. *Transportation Science*, 35(1):50–60, 2001.
- [16] Lucio Bianco, Giuseppe Confessore, and Pierfrancesco Reverberi. A network based model for traffic sensor location with implications on o/d matrix estimates. *Transportation Science*, 35(1):50–60, 2001.
- [17] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*, chapter Convex functions, page 67. Cambridge University Press, 2004.
- [18] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*, chapter Convex optimization problems, pages 127–146. Cambridge University Press, 2004.
- [19] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*, chapter Convex sets, pages 23–25. Cambridge University Press, 2004.
- [20] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*, chapter Convex functions, pages 71–73. Cambridge University Press, 2004.
- [21] Ennio Cascetta. Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B: Methodological*, 18(4-5):289–299, 1984.
- [22] Ennio Cascetta and Sang Nguyen. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological*, 22(6):437–455, 1988.
- [23] Ennio Cascetta and Sang Nguyen. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological*, 22(6):437–455, 1988.
- [24] Ennio Cascetta and Maria Nadia Postorino. Fixed point approaches to the estimation of o/d matrices using traffic counts on congested networks. *Transportation science*, 35(2):134–147, 2001.
- [25] Enrique Castillo, Antonio J Conejo, Rosa Eva Pruneda, and Cristina Solares. Observability in linear systems of equations and inequalities: Applications. *Computers & operations research*, 34(6):1708–1720, 2007.
- [26] Enrique Castillo, Inmaculada Gallego, José María Menéndez, and Ana Rivas. Optimal use of plate-scanning resources for route flow estimation in traffic networks. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):380–391, 2010.
- [27] Enrique Castillo, Pilar Jimenez, Jose Maria Menendez, and Antonio J Conejo. The observability problem in traffic models: algebraic and topological methods. *IEEE Transactions on Intelligent Transportation Systems*, 9(2):275–287, 2008.

- [28] Enrique Castillo, José María Menéndez, and Pilar Jiménez. Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B: Methodological*, 42(5):455–481, 2008.
- [29] Gang-Len Chang and Xianding Tao. An integrated model for estimating time-varying network origin–destination distributions. *Transportation Research Part A: Policy and Practice*, 33(5):381–399, 1999.
- [30] Steve Coast. Openstreetmap. <https://www.openstreetmap.org/about>, February 2023.
- [31] European Commission. Executive summary of the impact assessment. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52013SC0532&from=EN>, December 2013.
- [32] European Commission, Directorate-General for Mobility, Transport, L Wijngaarden, A Schroten, H Essen, D Sutter, and E Andrew. *Sustainable transport infrastructure charging and internalisation of transport externalities : executive summary*. Publications Office, 2019.
- [33] Adam Danczyk and Henry X Liu. A mixed-integer linear program for optimizing sensor locations along freeway corridors. *Transportation Research Part B: Methodological*, 45(1):208–217, 2011.
- [34] Ayuntamiento de Bilbao. Geobilbao. <https://www.geobilbao.eus/geobilbao/>, Febrero 2023.
- [35] Ministerio de transportes movilidad y agenda urbana. Transporte, menor consumo y emisiones, y uso de energías alternativas. <https://www.mitma.gob.es/ministerio/proyectos-singulares/cumbre-del-clima/transporte>, 2019.
- [36] Sven Erlander, Sang Nguyen, and Neil Frederick Stewart. On the calibration of the combined distribution-assignment model. *Transportation Research Part B: Methodological*, 13(3):259–267, 1979.
- [37] Mariem Fekih, Tom Bellemans, Zbigniew Smoreda, Patrick Bonnel, Angelo Furno, and Stéphane Galland. A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of lyon region (france). *Transportation*, 48:1671–1702, 2021.
- [38] Qian Ge and Daisuke Fukuda. Updating origin–destination matrices with aggregated data of gps traces. *Transportation Research Part C: Emerging Technologies*, 69:291–312, 2016.
- [39] Monica Gentili and Pitu B Mirchandani. Locating active sensors on traffic networks. *Annals of Operations Research*, 136:229–257, 2005.
- [40] Zhejun Gong. Estimating the urban od matrix: A neural network approach. *European Journal of operational research*, 106(1):108–115, 1998.
- [41] Juan Guerrero-Ibáñez, Sherali Zeadally, and Juan Contreras-Castillo. Sensor technologies for intelligent transportation systems. *Sensors*, 18(4), 2018.
- [42] David T. Hartgen and Milton Gregory Fields. Gridlock and growth: The effect of traffic congestion on regional economic performance. 2009.

- [43] Martin L Hazelton. Estimation of origin–destination matrices from link flows on uncongested networks. *Transportation Research Part B: Methodological*, 34(7):549–566, 2000.
- [44] Martin L Hazelton. Statistical inference for time varying origin–destination matrices. *Transportation Research Part B: Methodological*, 42(6):542–552, 2008.
- [45] Sheng-xue He. A graphical approach to identify sensor locations for link flow inference. *Transportation Research Part B: Methodological*, 51:65–76, 2013.
- [46] Per Högberg. Estimation of parameters in models for traffic prediction: a non-linear regression approach. *Transportation Research*, 10(4):263–265, 1976.
- [47] Shou-Ren Hu and Han-Tsung Liou. A generalized sensor location model for the estimation of network origin–destination matrices. *Transportation Research Part C: Emerging Technologies*, 40:93–110, 2014.
- [48] Shou-Ren Hu, Srinivas Peeta, and Chun-Hsiao Chu. Identification of vehicle sensor locations for link-based network traffic applications. *Transportation Research Part B: Methodological*, 43(8-9):873–894, 2009.
- [49] Md Shahadat Iqbal, Charisma F Choudhury, Pu Wang, and Marta C González. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74, 2014.
- [50] Pushkin Kachroo, Kaan Ozbay, and Arvind Narayanan. Investigating the use of kalman filtering approaches for dynamic origin-destination trip table estimation. In *Proceedings IEEE SOUTHEASTCON 97. Engineering the New Century*, pages 138–142. IEEE, 1997.
- [51] Hyunmyung Kim, Seungkirl Baek, and Yongtaek Lim. Origin-destination matrices estimated with a genetic algorithm from link traffic counts. *Transportation Research Record*, 1771(1):156–163, 2001.
- [52] Smart Mobility Lab. Soluciones de analítica para la movilidad. <https://sml.tecnalia.com/index.html>, Febrero 2023.
- [53] Baibing Li and Bart De Moor. Dynamic identification of origin–destination matrices in the presence of incomplete observations. *Transportation Research Part B: Methodological*, 36(1):37–57, 2002.
- [54] HP Lo, N Zhang, and William HK Lam. Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. *Transportation Research Part B: Methodological*, 30(4):309–324, 1996.
- [55] Dana E Low. New approach to transportation systems modeling. *Traffic quarterly*, 26(3), 1972.
- [56] Jan T Lundgren and Anders Peterson. A heuristic for the bilevel origin–destination-matrix estimation problem. *Transportation Research Part B: Methodological*, 42(4):339–354, 2008.
- [57] Dennis Luxen. Open source routing machine. <https://project-osrm.org/>, February 2023.
- [58] Michael J Maher. Inferences on trip matrices from observations on link volumes: a bayesian statistical approach. *Transportation Research Part B: Methodological*, 17(6):435–447, 1983.

- [59] Joachim Dahl Martin Andersen and Lieven Vandenberghe. Cone programming. <https://cvxopt.org/userguide/coneprog.html#cone-programming>, March 2022.
- [60] Joachim Dahl Martin Andersen and Lieven Vandenberghe. Python software for convex optimization. <https://cvxopt.org/>, March 2022.
- [61] Joachim Dahl Martin Andersen and Lieven Vandenberghe. Quadratic programming. <https://cvxopt.org/userguide/coneprog.html#quadratic-programming>, March 2022.
- [62] Gabriel Michau, Nelly Pustelnik, Pierre Borgnat, Patrice Abry, Ashish Bhaskar, and Edward Chung. Combining traffic counts and bluetooth data for link-origin-destination matrix estimation in large urban networks: The brisbane case study. *arXiv preprint arXiv:1907.07495*, 2019.
- [63] Gabriel Etienne Michau, Alfredo Nantes, Edward Chung, Patrice Abry, and Pierre Borgnat. Retrieving dynamic origin-destination matrices from bluetooth data. In *Transportation Research Board (TRB) 93rd Annual Meeting Compendium of Papers*, pages 1–11. Transportation Research Board (TRB), 2014.
- [64] Baichuan Mo, Ruimin Li, and Jingchen Dai. Estimating dynamic origin–destination demand: A hybrid framework using license plate recognition data. *Computer-Aided Civil and Infrastructure Engineering*, 35(7):734–752, 2020.
- [65] ManWo Ng. Synergistic sensor location for link flow inference without path enumeration: A node-based approach. *Transportation Research Part B: Methodological*, 46(6):781–788, 2012.
- [66] S Nguyen. Estimating an od matrix from network data: A network equilibrium approach, publication 87. *Centre de Recherche sur les Transports, Universite de Montreal*, 1977.
- [67] Yu Nie and Der-Horng Lee. Uncoupled method for equilibrium-based linear path flow estimator for origin-destination trip matrices. *Transportation Research Record*, 1783(1):72–79, 2002.
- [68] Yu Nie, Hui-Min Zhang, and WW Recker. Inferring origin–destination trip matrices with a decoupled gls path flow estimator. *Transportation Research Part B: Methodological*, 39(6):497–518, 2005.
- [69] European Court of Auditors. Sustainable urban mobility in the european union. <https://op.europa.eu/webpub/eca/special-reports/urban-mobility-6-2020/en/index.html>, June 2020.
- [70] Iwao Okutani. The kalman filtering approaches in some transportation and traffic problems. *Transportation and traffic theory*, 1987.
- [71] World Health Organization. *Ambient air pollution: a global assessment of exposure and burden of disease*. World Health Organization, 2016.
- [72] Fernando Pascullo. Distritos de bilbao numerados. https://commons.wikimedia.org/wiki/File:Distritos_Bilbao_numerados.svg, Febrero 2023.
- [73] Pierre Robillard. Estimating the od matrix from observed link volumes. *Transportation research*, 9(2-3):123–128, 1975.

- [74] Xavier Ros Roca. *Dynamic OD matrix estimation exploiting ICT traffic measurements*. PhD thesis, UPC, Departament d'Estadística i Investigació Operativa, Oct 2021.
- [75] Xavier Ros-Roca, Lidia Montero, Jaume Barceló, Klaus Nökel, and Guido Gentile. A practical approach to assignment-free dynamic origin–destination matrix estimation problem. *Transportation Research Part C: Emerging Technologies*, 134:103477, 2022.
- [76] Krishna Saw, BK Katti, and G Joshi. Literature review of traffic assignment: static and dynamic. *International Journal of Transportation Engineering*, 2(4):339–347, 2015.
- [77] Ralf-Peter Schäfer, Kai-Uwe Thiessenhusen, Elmar Brockfeld, and Peter Wagner. A traffic information system by means of real-time floating-car data. 2002.
- [78] Immanuel Scholz. Java openstreetmap editor. <https://josm.openstreetmap.de/>, February 2023.
- [79] Hanif D Sherali, Arvind Narayanan, and R Sivanandan. Estimation of origin–destination trip-tables based on a partial set of traffic link volumes. *Transportation Research Part B: Methodological*, 37(9):815–836, 2003.
- [80] Hanif D Sherali and Taehyung Park. Estimation of dynamic origin–destination trip tables for a general network. *Transportation Research Part B: Methodological*, 35(3):217–235, 2001.
- [81] Hanif D Sherali, RAGA Sivanandan, and Antoine G Hobeika. A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes. *Transportation Research Part B: Methodological*, 28(3):213–233, 1994.
- [82] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.
- [83] Heinz Spiess. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological*, 21(5):395–412, 1987.
- [84] Heinz Spiess. A gradient approach for the od matrix adjustment problem. *Transportation Research Part B: Methodological*, 24(1):1–12, 1990.
- [85] OZ Tamin and LG Willumsen. Transport demand model estimation from traffic counts. *Transportation*, 16:3–26, 1989.
- [86] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.
- [87] TecNALIA. Centro de investigación y desarrollo tecnológico. <https://www.tecnalia.com/>, Febrero 2023.
- [88] Dus an Teodorovic . Fuzzy logic systems for transportation engineering: the state of the art. *Transportation Research Part A: Policy and Practice*, 33(5):337–364, 1999.
- [89] Theodore Tsekeris and Antony Stathopoulos. Real-time dynamic origin-destination matrix adjustment with simulated and actual link flows in urban networks. *Transportation Research Record*, 1857(1):117–127, 2003.
- [90] Urbanite. <https://www.urbanite-project.eu/>, Febrero 2023.

- [91] Henk J Van Zuylen and Luis G Willumsen. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 14(3):281–293, 1980.
- [92] PTV VISSIM. Multimodal traffic simulation software. <https://www.ptvgroup.com/en/solutionsproducts/ptv-vissim/>, February 2023.
- [93] Huayong Wang, Francesco Calabrese, Giusy Di Lorenzo, and Carlo Ratti. Transportation mode inference from anonymized and aggregated mobile phone call detail records. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 318–323, 2010.
- [94] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers, Part II*, pages 325–362, 1952.
- [95] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.
- [96] Luis G Willumsen. Estimation of an od matrix from traffic counts—a review. 1978.
- [97] Sze Chun Wong, CO Tong, Ka-lo Wong, WH K Lam, Hong Kam Lo, Huajie Yang, and HP Lo. Estimation of multiclass origin–destination matrices from traffic counts. *Journal of urban planning and development*, 131(1):19–29, 2005.
- [98] Hai Yang, Yasunori Iida, and Tsuna Sasaki. The equilibrium-based origin–destination matrix estimation problem. *Transportation Research Part B: Methodological*, 28(1):23–33, 1994.
- [99] Hai Yang, Tsuna Sasaki, Yasunori Iida, and Yasuo Asakura. Estimation of origin–destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological*, 26(6):417–434, 1992.
- [100] Hai Yang and Jing Zhou. Optimal traffic counting locations for origin–destination matrix estimation. *Transportation Research Part B: Methodological*, 32(2):109–126, 1998.
- [101] Xuesong Zhou and Hani S Mahmassani. Dynamic origin–destination demand estimation using automatic vehicle identification data. *IEEE Transactions on intelligent transportation systems*, 7(1):105–114, 2006.

A. Anexo I

En este anexo se describe en detalle la implementación de las funciones de *Cone Programming* introducidas en la Sección 6.3.4. Concretamente, se desarrolla como resolver los problemas (6.10) y (6.12) mediante la función qp .

La función qp se emplea para la resolución de problemas con función de coste cuadrática de la forma:

$$\left\{ \begin{array}{ll} \text{Minimizar} & (1/2) \bar{x}^t P \bar{x} + \bar{q}^t \\ \text{Sujeto a} & G \bar{x} \leq \bar{h} \\ & A \bar{x} = \bar{b} \end{array} \right. \quad (\text{A.1})$$

Donde las desigualdades son desigualdades vectoriales por componentes.

Para poder resolver los problemas planteados en (6.10) y (6.12) mediante la función qp , se debe asignar a cada una de las variables que aparecen en (A.1) el valor adecuado. Concretamente, las asignaciones en común a realizar en ambos problemas son las siguientes:

$$\left\{ \begin{array}{l} \bar{q} = \bar{0} \\ G = -I \\ \bar{h} = \bar{0} \end{array} \right. \quad (\text{A.2})$$

De manera que el problema planteado en (A.1) queda de la forma:

$$\left\{ \begin{array}{ll} \text{Minimizar} & (1/2) \bar{x}^t P \bar{x} \\ \text{Sujeto a} & \bar{x} \geq \bar{0} \\ & A \bar{x} = \bar{b} \end{array} \right. \quad (\text{A.3})$$

Una vez realizadas las asignaciones en común para ambos problemas, es necesario configurar la matriz P para modelar la condición de minimización de cada caso. En el caso del problema (6.10), la matriz P es la matriz identidad I . En el caso del problema (6.12), los elementos P_{uv} de la matriz toman los siguientes valores:

$$P_{uv} = \begin{cases} (N(u) - 1)/N(u) & u = v \\ (-1/N(u)) & u \neq v \quad u \sim v \\ 0 & u \neq v \quad u \not\sim v \end{cases} \quad (\text{A.4})$$

Donde:

$$\begin{cases} u \sim v \iff (o, d) = \gamma^{-1}(u) \sim (o', d') = \gamma^{-1}(v) \\ (o, d) \sim (o', d') \iff \begin{cases} (o, o') \in i \\ (d, d') \in j \end{cases} \end{cases} \quad (\text{A.5})$$

Donde i es la zona de origen, j la zona destino, y $N(u)$ representa el número de elementos en cada fila u que cumple $u \sim v$. Es decir, el valor de N para cada fila de P , depende de la cantidad de elementos en dicha fila que cumplen $u \sim v$.