



**Secretaría de Estado de Infraestructuras,  
Transporte y Vivienda**

**Subdirección General de Planificación  
de Infraestructuras y Transporte**

**Estudio de la Movilidad  
Interprovincial de Viajeros  
aplicando la Tecnología Big Data  
- Proyecto piloto -**

**Informe metodológico**

Diciembre de 2019

## Resumen

Este documento constituye el informe metodológico del Estudio de la Movilidad Interprovincial de Viajeros aplicando la Tecnología Big Data que el Ministerio de Fomento realizó durante el año 2018. Se trata de un proyecto piloto y experimental que pretende abrir camino al empleo de las nuevas tecnologías en sustitución de la metodología clásica de encuestas directas a los ciudadanos. Dado que es la primera vez que esta tecnología es utilizada para obtener flujos de movilidad interprovincial a nivel nacional, la metodología empleada, el alcance de sus resultados y la bondad de estos deberá irse perfeccionando en la medida que los datos utilizados sean cada vez más completos y precisos y se avance en la definición de algoritmos más sofisticados para extraer la información deseada. En definitiva, se considera este estudio como un punto de apoyo o de partida para ir avanzando en la aplicación de la tecnología big data para la definición y cualificación de la demanda de transporte.

El estudio ha utilizado como fuente de datos principal los registros de las antenas de telefonía móvil una vez anonimizados, cumpliendo con las normativas vigentes de protección de datos. A partir de ellos y su fusión con otras fuentes que incluyen datos entre otros de la red y de la oferta y la demanda de transporte, se han generado distintos indicadores-valores de demanda de transporte para un periodo estival (julio/agosto) y un periodo no estival (octubre).

Los resultados principales del proyecto son matrices de viajes y matrices de etapas. La información de la demanda se ha segmentado por modo de transporte (carretera, ferrocarril, aéreo, marítimo, con un análisis parcial y muy experimental del autobús a nivel de las concesiones del Ministerio de Fomento para Origen/destino entre CC.AA.), periodo del día, distancia del viaje, residencia de los viajeros (residentes en España y residentes en el extranjero) y motivo del viaje.

En este documento se indican las bases de datos recopiladas y utilizadas en el proyecto y la metodología empleada.

# Índice

<b>1. INTRODUCCIÓN</b> .....	<b>1</b>
1.1 ANTECEDENTES.....	1
1.2 OBJETO Y ALCANCE DEL ESTUDIO .....	2
1.2.1 <i>Definiciones</i> .....	2
1.2.2 <i>Objeto y alcance</i> .....	2
<b>2. DATOS</b> .....	<b>6</b>
2.1 TELEFONÍA MÓVIL .....	6
2.2 USOS DEL SUELO .....	7
2.3 MARCO MUESTRAL.....	7
2.4 TRANSPORTE POR CARRETERA.....	7
2.4.1 <i>Generación de rutas por carretera</i> .....	8
2.4.2 <i>Datos de aforo</i> .....	8
2.4.3 <i>Datos sobre el modo autobús</i> .....	9
2.5 TRANSPORTE FERROVIARIO.....	10
2.6 TRANSPORTE AÉREO.....	12
2.7 TRANSPORTE MARÍTIMO .....	12
<b>3. METODOLOGÍA</b> .....	<b>14</b>
3.1 EXTRACCIÓN Y ANONIMIZACIÓN DE LOS REGISTROS DE TELEFONÍA MÓVIL.....	14
3.2 PRE-PROCESADO Y LIMPIEZA DE LOS DATOS .....	14
3.2.1 <i>Limpieza y depuración de los registros anonimizados</i> .....	15
3.2.2 <i>Limpieza y depuración de los datos de red</i> .....	15
3.3 CONSTRUCCIÓN DE LA MUESTRA POTENCIAL .....	15
3.4 EXTRACCIÓN DE DIARIOS DE ACTIVIDADES Y VIAJES .....	15
3.4.1 <i>Identificación de estancias y actividades</i> .....	16
3.4.2 <i>Caracterización de viajes y etapas. Identificación de modo y ruta</i> .....	17
3.5 SELECCIÓN DE LA MUESTRA EFECTIVA .....	18
3.6 ELEVACIÓN DE LA MUESTRA AL TOTAL DE LA POBLACIÓN.....	19
3.6.1 <i>Elevación de la muestra de residentes en España</i> .....	19
3.6.2 <i>Elevación de la muestra de no residentes</i> .....	19
3.7 VALIDACIÓN DE RESULTADOS.....	19
3.8 AJUSTE DE RESULTADOS.....	20
3.9 GENERACIÓN DE RESULTADOS.....	22
<b>4. RESULTADOS DEL TRABAJO</b> .....	<b>23</b>

## Figuras

<i>Figura 1. Días objeto de estudio</i>	3
<i>Figura 2. Zonificación de estudio</i>	3
<i>Figura 3. Usos del suelo en la ciudad de Madrid</i>	7
<i>Figura 4. Ejemplo de conjunto de rutas alternativas generadas para un viaje con origen en la provincia de Madrid y destino la provincia de Barcelona. Cada uno de los colores muestra una ruta distinta (6 rutas en total)</i>	8
<i>Figura 5. Mapa de aforos permanentes seleccionados para el estudio</i>	9
<i>Figura 6. Esquema sobre la información de espacio-temporal generada para el caso del autobús. Para cada expedición se dispone de la parada (Pi) y de la hora de paso (ti).</i>	10
<i>Figura 7. Ejemplo de información principal sobre red - concesiones de autobuses</i>	10
<i>Figura 9. Ejemplo de la información de circulaciones</i>	11
<i>Figura 10. Información de rutas ferroviarias entre las estaciones de Madrid Puerta de Atocha y Sants (oferta para el día 16/07/2017)</i>	12
<i>Figura 11. Red de aeropuertos considerada en el estudio</i>	13
<i>Figura 12. Red de puertos considerada en el estudio</i>	13
<i>Figura 13. Comparativa de los resultados para los modos ferrocarril y aéreo antes y después del proceso de ajuste para un día tipo. En la fila superior se muestran los resultados antes del ajuste. En la fila inferior se muestran los resultados obtenidos tras el ajuste. En el modo ferroviario se han eliminado los pares O/D con servicios de cercanías</i>	22

## Tablas

<i>Tabla 1. Listado de nacionalidades consideradas en el estudio</i>	5
<i>Tabla 2. Especificación de la matriz de viajes/etapas</i>	24

# 1. Introducción

## 1.1 Antecedentes

La planificación y gestión de los sistemas de transporte requiere del conocimiento de la movilidad y de información acerca de la demanda de viajes. Los métodos tradicionales de recogida de datos basados en encuestas, buscan y analizan la información detallada que es proporcionada por los ciudadanos sobre su movilidad. Sin embargo, las encuestas de movilidad presentan una serie de limitaciones importantes. Su realización consume mucho tiempo y recursos, por lo que la muestra analizada finalmente es limitada, tanto en el número de personas entrevistadas como en su ámbito geográfico y temporal, lo que repercute en la calidad de la información obtenida. Por el mismo motivo, a menudo, no se dispone de información sobre eventos especiales que influyen en la movilidad (fines de semana, periodos vacacionales, eventos relevantes, etc.). Finalmente, es común que los encuestados simplifiquen sus respuestas, por lo que la información puede adquirir sesgos y convertirse en imprecisa o incorrecta.

En este sentido, el Ministerio de Fomento, a través de la Secretaría de Estado de Infraestructuras, Transporte y Vivienda, en su labor de optimizar la planificación y la gestión de las infraestructuras y el transporte, viene realizando de manera periódica estudios de demanda. Dichos estudios se consideran esenciales para evaluar el impacto de las actuaciones de infraestructuras, de servicios o normativas sobre la evolución del sistema de transportes, y constituyen una información de gran valor para los distintos tipos de usuarios, tanto públicos como privados. Uno de estos estudios lo constituye los análisis de la movilidad de viajeros a nivel nacional. En el pasado, dicho estudio se realizaba a través de las denominadas encuestas Movilia, las cuales suponían un importante esfuerzo económico, técnico y humano. La última encuesta realizada con ese objeto, Movilia 2007, data de hace ya más de diez años.

Por otro lado, y en los últimos años, la proliferación de nuevas fuentes de datos geolocalizados, entre ellos los procedentes de dispositivos móviles, ha abierto la posibilidad de explorar nuevas formas de estudiar la movilidad y obtener información de demanda de transporte en un plazo de tiempo reducido y a un coste significativamente menor que el de los métodos tradicionales, eliminando o mitigando algunas de las principales limitaciones de las encuestas de movilidad. De esta forma, los datos procedentes de la red de telefonía móvil, previamente anonimizados y cumpliendo con las normativas vigentes de protección de datos, que fueron generados originalmente a otros efectos (facturación o gestión de la red), resultan particularmente interesantes para este propósito, gracias a la posibilidad de obtener muestras de gran tamaño y con una significativa resolución espacio-temporal de prácticamente todos los segmentos de población.

Teniendo presentes estas oportunidades, en 2017, desde el Ministerio de Fomento se promovió la realización del proyecto **‘Estudio de la Movilidad Interprovincial de Viajeros aplicando la Tecnología Big Data’**, utilizando como fuente principal de datos los registros anonimizados generados por los terminales móviles al conectarse a las antenas de las redes de telefonía y que, tras concurso público, fue realizado durante 2018 por la UTE de la empresas Kineo Mobility Analytics y Orange España.

## 1.2 Objeto y alcance del estudio

### 1.2.1 Definiciones

En el proyecto se emplean las siguientes definiciones de ‘estancia’, ‘actividad’, ‘etapa’ y ‘viaje’:

- **Estancia:** permanencia de un individuo durante cierto tiempo en un lugar determinado.
- **Actividad:** estancia originada por una interacción o conjunto de interacciones con el entorno (incluyendo interacciones otras personas, con un determinado servicio, etc.) que motivan que el individuo se desplace hasta allí.
- **Etapas:** desplazamiento entre las localizaciones de dos estancias consecutivas realizado en un modo de transporte.
- **Viaje:** desplazamiento entre las localizaciones de dos actividades consecutivas.

Teniendo en cuenta estas definiciones, un viaje tiene un propósito principal determinado por la actividad en el origen, la actividad en el destino o la combinación de ambas. La definición de actividad depende del propósito del estudio y de la unidad de análisis espacial que se tome como referencia (por ejemplo, ¿a partir de qué distancia entre emplazamientos se considera que la visita a varios puntos de atractivo turístico constituye una única actividad o un conjunto de actividades?). Así, una vez fijado un criterio y una unidad de análisis espacial para las estancias y actividades, es posible construir un diario de actividades y viajes para un determinado período de tiempo, y subdividir los viajes en las distintas etapas que componen los viajes multietapa, ya sean multimodales (por ejemplo, un viaje compuesto por una primera etapa en coche hasta el aeropuerto, una segunda etapa en avión y una última etapa en tren hasta el destino final) o monomodales (por ejemplo, un viaje en coche con una parada intermedia para repostar, o un viaje con dos etapas en ferrocarril, con un transbordo).

### 1.2.2 Objeto y alcance

El objeto del proyecto es caracterizar la **movilidad interprovincial** de viajeros a escala nacional, con el siguiente alcance:

- **Período de estudio:** se analizan dos periodos del año, en concreto el periodo de julio-agosto de 2017 y el periodo de octubre de 2017. Los días objeto de estudio se muestran en la Figura 1. El estudio se ha realizado considerando agrupaciones de días para analizar días promedio (lunes promedio, martes promedio, miércoles promedio, etc.) así como días específicos, para analizar la movilidad en días con un comportamiento de la movilidad particular y distinto del promedio del periodo.
- **Zonificación:** para caracterizar el origen y destino de los viajes se ha generado una zonificación de 59 zonas que corresponden con la división a nivel provincial, Baleares y Canarias a nivel de isla, dos zonas adicionales correspondientes a las ciudades autónomas de Ceuta y Melilla. La Figura 2 muestra la zonificación empleada.

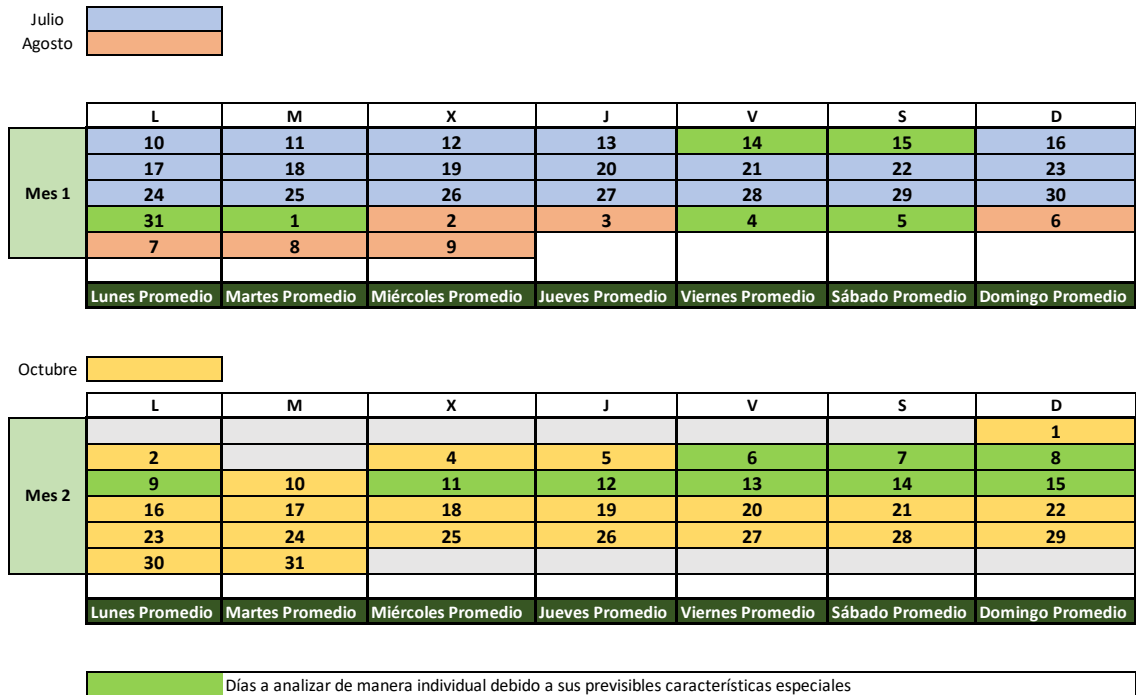


Figura 1. Días objeto de estudio

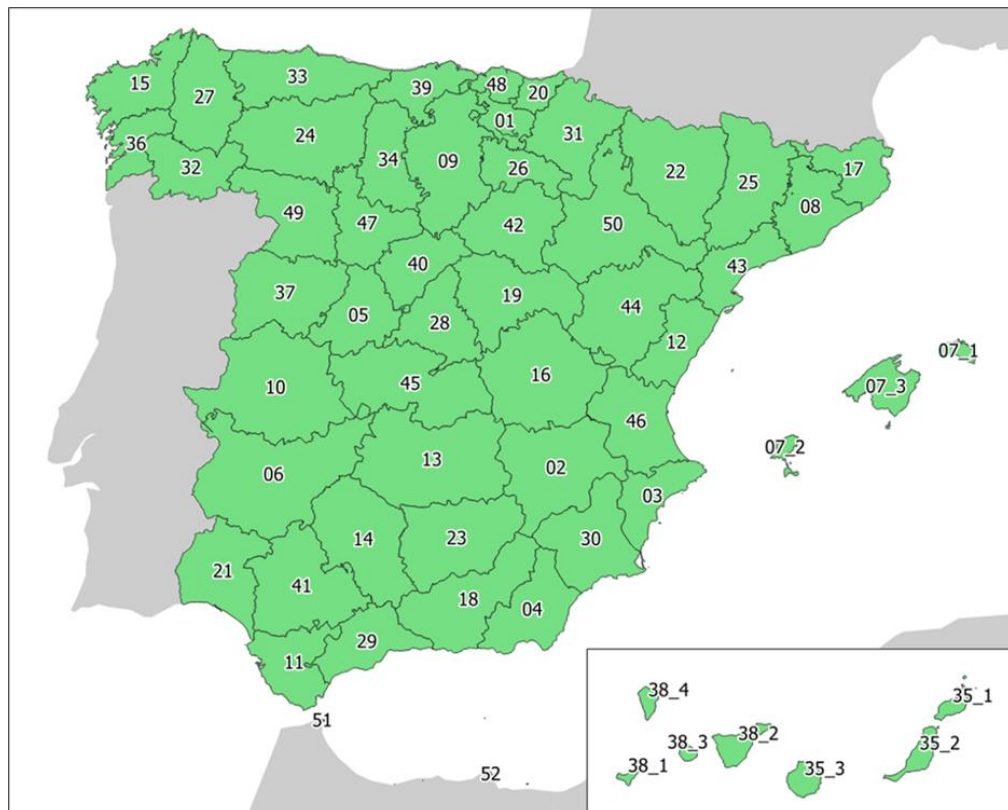


Figura 2. Zonificación de estudio



- **Viajes y etapas a analizar:** se analizan los viajes entre las zonas indicadas en el punto anterior. En los modos ‘carretera’ y ‘ferrocarril’, solo se analizan los viajes cuyos itinerarios superan los 50 km de longitud. Adicionalmente, para las provincias de Madrid, Barcelona, Alicante y Vizcaya se analizan también los viajes de entre 10 y 50 km dado que en un análisis previo eran las provincias con mayor número de viajes entre esas distancias. Con respecto a las etapas asociadas a los viajes, se han analizado todas aquellas etapas de carácter interprovincial asociadas a los viajes relevantes, en función de la distancia del viaje que las origina. Se ha detectado que, en algunos casos, en un viaje con etapas continuadas en un mismo modo, cuando no existen registros claros en la terminal de conexión, al final el algoritmo utilizado identifica la etapa como única en dicho modo entre O/D.
- **Tipos de matrices de movilidad obtenidas.** Se han obtenido dos tipos de matrices de movilidad: matrices de viajes y matrices de etapas.
- **Segmentación por modos de transporte:** la demanda de viajes se ha segmentado por los modos carretera, ferrocarril, marítimo y aéreo.

Para la carretera se ha realizado un detalle adicional en función de los O/D interprovinciales:

- **O/D intra CC.AA.:** en las matrices elaboradas, en la columna Modo aparece “carretera” que representa el volumen total de movimientos en este modo para cada par O/D.
- **O/D inter CC.AA.:** en las matrices elaboradas, en la columna Modo pueden aparecer:
  - **“autobús”** indica que se trata de un O/D que cuenta con servicios de concesión nacional de autobús y se contabilizan los movimientos realizados exclusivamente en dichas concesiones.
  - **“privado”**, representa para cada uno de los O/D inter CC.AA., la movilidad total que se produce por carretera a la que se le ha restado, cuando existe, la correspondiente a las concesiones nacionales, anteriormente descrita.

Es decir, en los O/D inter CC.AA. el volumen total de movimientos por carretera será la suma de “autobús” más “privado” para cada par O/D.

Los viajes aéreos con inicio y final en un helipuerto no han podido ser claramente distinguidos dado el pequeño volumen de personas transportado en cada servicio y especialmente esto ocurre si existe coincidencia geográfica de las terminales con otros modos, por ejemplo: marítimas y aéreas.

Los viajes realizados por vehículos pesados distintos del autobús (por ejemplo, transporte de mercancías) han sido identificados y descartados de la muestra.

Es muy importante señalar que, en la matriz de viajes, para los viajes multimodales, se proporciona como modo de transporte el modo que corresponde a la etapa de mayor longitud dentro del viaje.

- **Franjas horarias:** la información de demanda se ha segmentado en distintas franjas horarias:
  - 00:00 – 06:00
  - 06:00 – 10:00
  - 10:00 – 17:00
  - 17:00 – 00:00

Como criterio de hora del viaje se ha considerado la hora de inicio del viaje. Es decir, un viaje con inicio a las 16:30 y finalización a las 18:45 ha sido imputado a la franja horaria de 10:00 a 17:00.

- **Marco muestral:** como marco muestral, para los residentes en España, se ha considerado la población residente en España mayor a 10 años reportada por el Padrón de Habitantes de 2017. Para el caso de los no residentes en España, se ha utilizado como marco muestral la información de la encuesta FRONTUR del INE de los periodos de julio, agosto y octubre de 2017.
- **Segmentación por lugar de residencia:** la información se proporciona segmentada según el lugar de residencia de los viajeros, distinguiendo entre ‘residentes’ y ‘no residentes’. Dentro del grupo de ‘no residentes’, el lugar de residencia se facilita por país o grupo de países. Se ha empleado para ello la segmentación empleada por el INE en la encuesta FRONTUR, salvo en el caso de los residentes en Suiza, los cuales, debido al reducido tamaño de la muestra de los datos de telefonía, han sido incluidos en el grupo ‘resto de Europa’. La *Tabla 1* muestra la lista de nacionalidades consideradas. Los viajes y etapas considerados para los no residentes son aquellos que realizan en territorio nacional independientemente del modo o lugar de acceso. Su análisis no siempre ha sido fácil de realizar pues las señales de posicionamiento y movimiento captadas por las antenas, en ciertas ocasiones, no responden a una movilidad lógica.

*Tabla 1. Listado de nacionalidades consideradas en el estudio*

Nacionalidades			
1	Alemania	9	Países Nórdicos
2	Bélgica	10	Rusia
3	Francia	11	Resto de Europa
4	Irlanda	12	EE.UU.
5	Italia	13	Resto de América
6	Holanda	14	Resto del mundo
7	Portugal	15	España
8	Reino Unido	-	-

- **Caracterización de actividades:** para el caso de los movimientos de residentes en España las actividades asociadas al origen y destino de los viajes se clasifican según las siguientes categorías:
  - *Casa:* lugar de residencia habitual, inferido del análisis anonimizado de los registros de antenas de telefonía
  - *Trabajo/Estudio:* lugar donde se permanece una jornada completa (6-8 horas)
  - *Frecuente:* lugar donde se realiza una actividad de manera recurrente distinta de casa y trabajo. Esta actividad puede estar asociada a actividades de ocio o a trabajos a tiempo parcial.
  - *Esporádica:* lugar donde se realiza una actividad no recurrente.

Esta cualificación de la demanda inferida utilizando la tecnología big data no aporta una información tan detallada y precisa como la obtenida utilizando la metodología clásica de encuestas directas a los ciudadanos.

## 2. Datos

En esta sección se describen las distintas fuentes de datos recopiladas y los procesos y análisis realizados sobre las mismas para obtener la información de la red, oferta y demanda de transporte necesaria para la ejecución del proyecto.

### 2.1 Telefonía móvil

La principal fuente de datos la constituyen los registros de telefonía móvil procedentes de la red de Orange España con algo más de 14 millones de líneas móviles, lo que supone un 27,3% de cuota de mercado (fuente: CNMC, julio de 2017). La elevada cuota de mercado permite que los datos, una vez anonimizados, con el adecuado procesamiento y análisis, puedan utilizarse para inferir estadísticas representativas a nivel del universo poblacional de España.

Los datos de partida pueden clasificarse en dos categorías:

- **Datos de eventos registrados**<sup>1</sup>: datos anonimizados asociados a los registros de conexión del dispositivo con la red de telefonía móvil. Estos registros están constituidos esencialmente por lo que se denomina CDRs (Call Detail Records), que proporcionan un registro cada vez que el dispositivo interactúa con la red (uso de voz o datos). A estos registros se les unen otros datos que incluyen determinados eventos pasivos (actualización periódica de la posición del dispositivo, cambios de áreas de cobertura, etc.), dando lugar a unos “CDRs enriquecidos” con mayor granularidad temporal. Estos datos proporcionan una granularidad temporal elevada (especialmente en el caso de los ‘smartphones’, que en España son utilizados por más del 90% de los usuarios de telefonía móvil), lo que permite determinar con cierto nivel de detalle la localización del dispositivo. En cuanto a la granularidad espacial, se dispone de información de localización a nivel de celda de telefonía, lo que supone una precisión espacial de decenas o cientos de metros en ciudad y hasta varios kilómetros en zonas rurales. Es precisamente esta menor precisión en las zonas menos habitadas la que origina cierta dificultad en distinguir sobre modos terrestres (carretera y ferrocarril) cuando su trazado está relativamente próximo y sus usuarios son atendidos o captados por las mismas antenas. Este hecho también se produce en las zonas o corredores con movimientos de corto recorrido en los que los trazados estén relativamente próximos y las señales temporales registradas no permitan definir con precisión el modo empleado. En general, para un día medio, se generan de alrededor de 1.000 millones de registros.
- **Datos de la topología de la red de telefonía móvil**: datos sobre la red de telefonía, incluyendo la localización de las torres de comunicación y la orientación de las antenas.

---

<sup>1</sup> Es importante señalar que los datos de eventos registrados son anonimizados y se procesan en un entorno seguro en la infraestructura del operador móvil para generar información agregada y por tanto anonimizada, con el fin de cumplir con la legislación en materia de protección de datos.

## 2.2 Usos del suelo

Se han utilizado también datos de usos del suelo, para mejorar la caracterización y la localización espacial de las actividades identificadas a partir de los datos de telefonía móvil. En la Figura 3 se muestra un ejemplo de la información de usos del suelo para el área de Madrid.

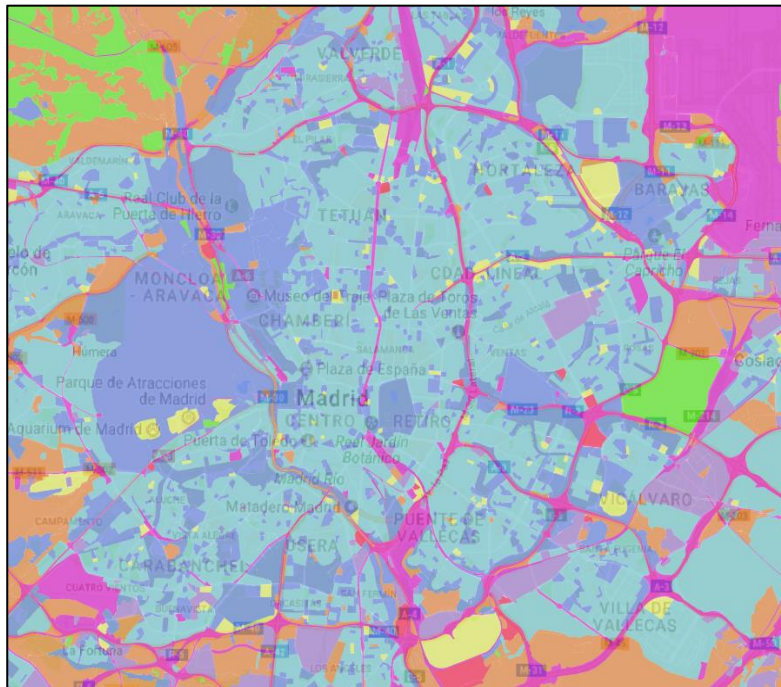


Figura 3. Usos del suelo en la ciudad de Madrid

## 2.3 Marco muestral

Para los procesos de elevación de la muestra se han utilizado datos procedentes del Padrón Municipal de Habitantes, para la población residente en España, y datos procedentes de la encuesta FRONTUR realizada por el INE, para la población extranjera. En concreto se han empleado datos del Padrón de Habitantes del año 2017, considerando la población residente en España con edad superior a 10 años; y datos de la encuesta FRONTUR de los periodos de julio, agosto y octubre de 2017, considerando la información de extranjeros por nacionalidad y tipo de visitante (excursionistas y turistas).

## 2.4 Transporte por carretera

En el caso del modo carretera, se han recopilado principalmente tres tipos de datos:

- información acerca de rutas por carretera,
- información de aforos permanentes y
- información asociada al modo autobús.

### 2.4.1 Generación de rutas por carretera

Los procesos de identificación de modo y ruta de los viajes por carretera requieren disponer del conjunto de rutas alternativas existentes entre todos los pares origen-destino objeto de estudio. El conjunto de rutas posibles se ha generado a partir de la información facilitada por distintas APIs de planificadores de viajes y fuentes de open data como OpenStreetMap ([www.openstreetmap.org](http://www.openstreetmap.org)). Aplicando diferentes parametrizaciones (tipo de vía, hora del día, etc.), se genera un conjunto de rutas suficientemente amplio que permita capturar las posibles rutas seleccionadas por los viajeros al realizar el viaje desde su origen a su destino.

Se han empleado alrededor de 17 millones de rutas, que se han utilizado en los procesos de identificación de modo y ruta. En la Figura 4 se muestra un ejemplo de las rutas generadas para un viaje con origen en la provincia de Madrid y destino la provincia de Barcelona.

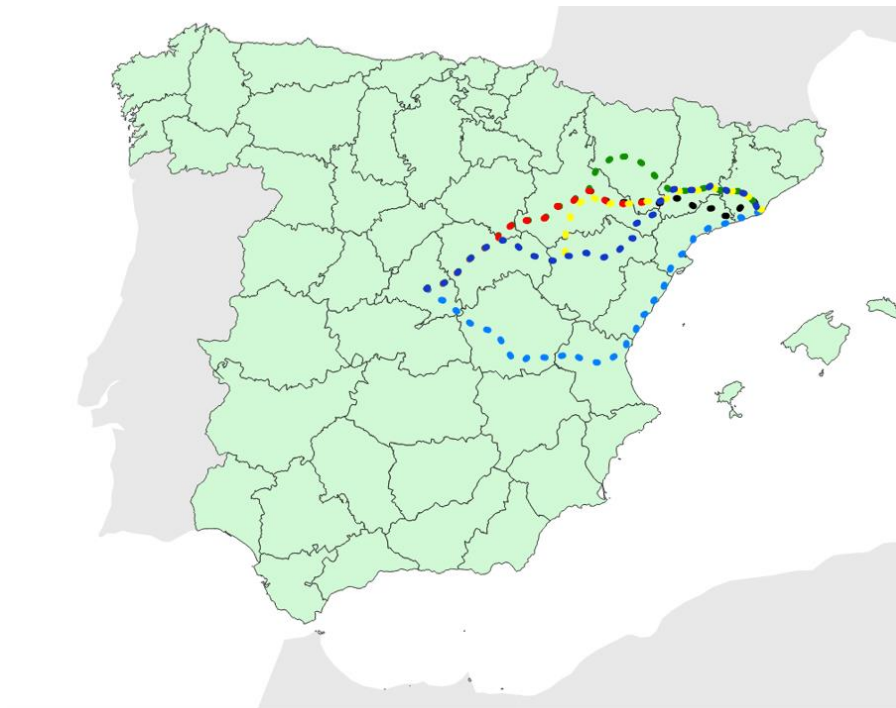


Figura 4. Ejemplo de conjunto de rutas alternativas generadas para un viaje con origen en la provincia de Madrid y destino la provincia de Barcelona. Cada uno de los colores muestra una ruta distinta (6 rutas en total)

### 2.4.2 Datos de aforo

Se han recopilado datos de aforo de todas las estaciones permanentes a nivel nacional para todos los días objeto de estudio. Para cada estación se dispone del identificador del aforo, su ubicación, el identificador de calzada y carril, y el volumen total de vehículos ligeros y pesados que han atravesado el aforo cada 15 minutos. En la Figura 5 se muestran los puntos de aforo permanentes recopilados en el proyecto. Los datos de aforo, junto con otros datos de demanda en otros modos, se han utilizado para la toma de decisiones con respecto a la identificación de los días singulares objeto de estudio.



Es importante señalar en este punto que los datos de aforo no se han empleado en los procesos de ajuste de resultados, debido a que el alcance del proyecto limita el estudio a ciertos tipos de viajes (interprovinciales, ciertos rangos de distancias, etc.), mientras que los datos de aforo recogen cualquier viaje que circule por la vía en dicho punto.

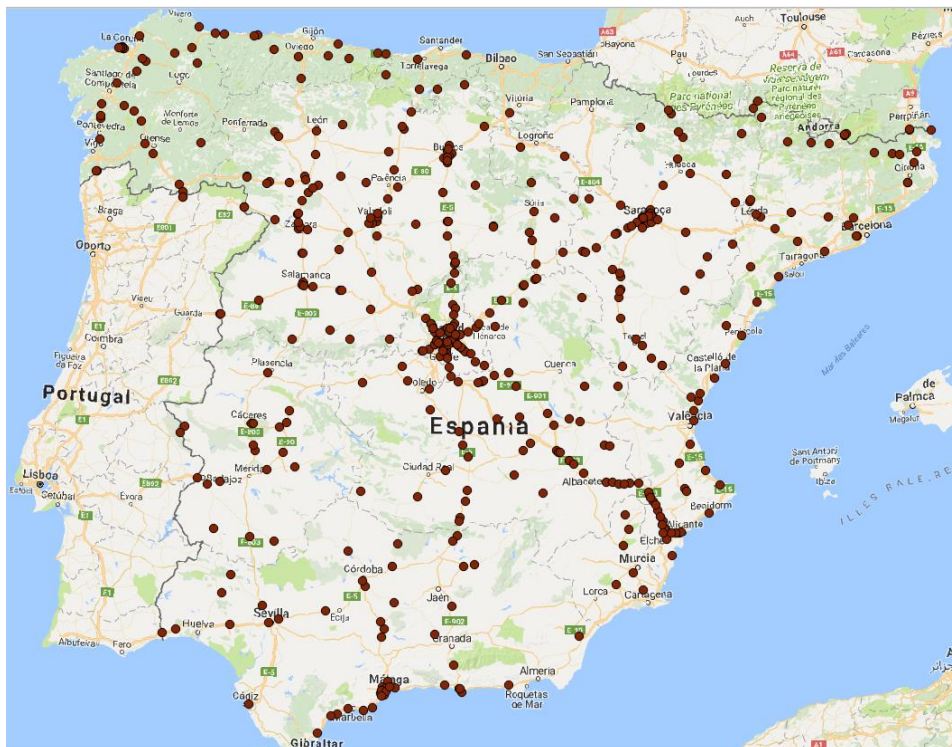


Figura 5. Mapa de aforos permanentes seleccionados para el estudio

### 2.4.3 Datos sobre el modo autobús

Para las concesiones de ámbito nacional del Ministerio de Fomento, se han recopilado tres tipos de datos:

- **Datos de red:** paradas que conforman la concesión y ubicación de las mismas<sup>2</sup>. En la Figura 7 se muestra un ejemplo de la información de red recopilada con sus principales campos.
- **Datos de oferta:** información relativa a cada expedición (número de expedición, número de ruta, sentido...), paradas de origen y destino, fecha de la expedición y hora de salida y llegada.
- **Datos de demanda:** se ha dispuesto, por periodo trimestral, de datos de demanda.

<sup>2</sup> Es importante señalar que la red de autobuses objeto de estudio es la conformada por las paradas para las cuales se tiene información de ubicación.

Para utilizar esta información en los procesos de identificación de modo y ruta, a partir de la información de ubicación de las paradas y de las horas de paso por cada parada se ha generado una secuencia espacio-temporal de posiciones de cada expedición para cada uno de los días de estudio. En la Figura 6 se muestra de manera esquemática la información espacio-temporal generada.



Figura 6. Esquema sobre la información de espacio-temporal generada para el caso del autobús. Para cada expedición se dispone de la parada ( $P_i$ ) y de la hora de paso ( $t_i$ ).

COD_CONTRATO	DESC_CONTRATO	NRUTA	DES_RUTA	ORDEN_PARADA	CODIGO_PARADA	NOMBRE_PARADA	DIRECCION_PARADA	COD_PROV	COD_MUNI	COD_POBLA	NOMBRE_POBLACION	COORDENADA_X	COORDENADA_Y
VAC-010	...	1	...	1	842	PARADA BUS	N-630	10	15	1	ALDEANUEVA DEL CAMINO	-5,92813188	40,25979198
VAC-010	...	1	...	2	880	ESTACION DE AUTOBUSES	AVDA DE ESPAÑA, Nº 4	10	96	1	HERVAS	-5,86195308	40,27241403
VAC-010	...	1	...	3	847	ESTACION DE AUTOBUSES	AVDA DE LAS TERMAS, S/N	10	24	1	BAÑOS	-5,85609912	40,31967096

Figura 7. Ejemplo de información principal sobre red - concesiones de autobuses

## 2.5 Transporte ferroviario

En el caso del modo ferroviario se ha recopilado la siguiente información:

- **Datos de red:** ubicación de las estaciones de tren e información sobre la tramificación de la red, incluyendo la información geométrica de cada tramo e información sobre las relaciones entre estaciones y tramos asociados. En la Figura se muestran los mapas asociados a las estaciones y la tramificación.

- **Datos de oferta:** se ha recopilado información sobre las circulaciones que tienen lugar cada día. La información de circulaciones proporciona los tramos de paso de cada tren a lo largo de su recorrido, entre la estación de origen y la estación de destino. En la Figura 8 se muestra un extracto de la información recopilada sobre circulaciones.
- **Datos de demanda:** se ha recopilado información sobre los flujos entre la estación de origen y destino. Es importante señalar que la información no incluye la demanda asociada a los servicios de cercanías.



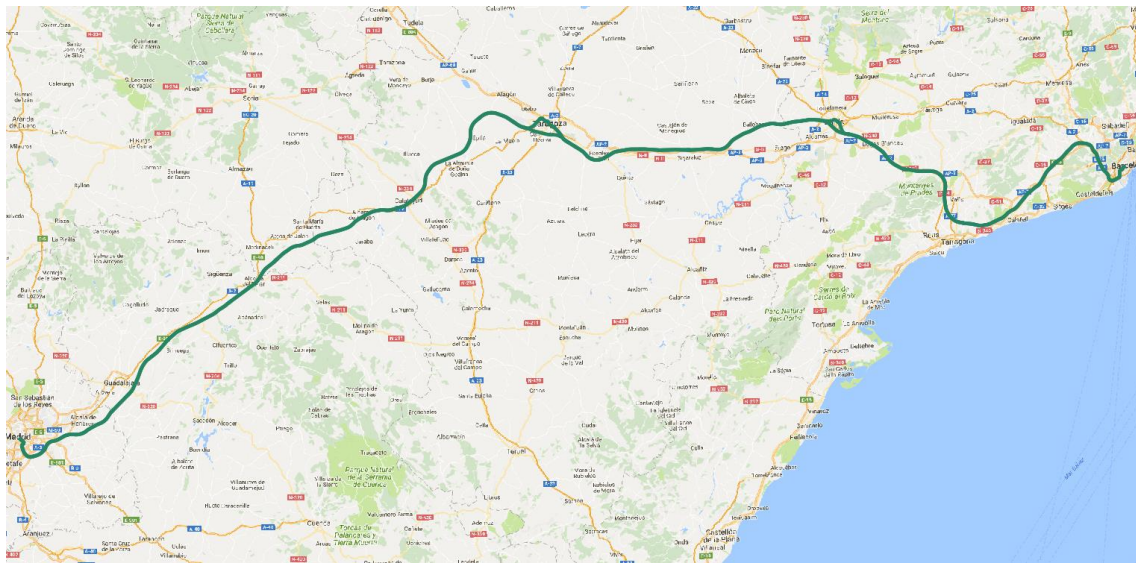
Figura 8. Información de estaciones de tren (izquierda) y tramificación de la red ferroviaria (derecha)

fecha	tren	origen	destino	emp_ferr	CodProducto	cod_tramo	pasos	max_orden01
20170802	26045	13200	13506	RF	C	17220010	1	2375071
20170802	26046	13506	13200	RF	C	17220010	1	2375077
20170802	26046	13506	13200	RF	C	17200060	1	2375082

Figura 8. Ejemplo de la información de circulaciones

Para cada día se han generado aproximadamente unas 45.000 rutas de tren, pudiendo este número disminuir o aumentar en función de la oferta ferroviaria concreta de cada día. En la Figura 9 se muestra el conjunto de rutas alternativas entre las estaciones de Atocha (Madrid) y Sants (Barcelona) para el día 16 de julio de 2017.





*Figura 90. Información de rutas ferroviarias entre las estaciones de Madrid Puerta de Atocha y Sants (oferta para el día 16/07/2017)*

## 2.6 Transporte aéreo

En el caso del modo aéreo se ha recopilado, principalmente, la siguiente información:

- **Datos de red:** información sobre la ubicación de los aeropuertos y distancia ortodrómica entre ellos. En total se han seleccionado 48 aeropuertos de interés. En la Figura 10 se muestra un mapa con la ubicación de los distintos aeropuertos considerados.
- **Datos de demanda:** se ha recopilado información sobre los movimientos entre aeropuertos de origen y destino y el volumen total de pasajeros.

## 2.7 Transporte marítimo

En el caso del modo marítimo se ha recopilado, principalmente, la siguiente información:

- **Datos de red:** información sobre la ubicación de los puertos y distancia marítima entre ellos. En total se han seleccionado 34 puertos de interés. En la Figura 11 se muestra un mapa con la ubicación de los distintos puertos considerados.
- **Datos de demanda:** para cada par de puertos, se dispone del volumen de pasajeros a nivel mensual.

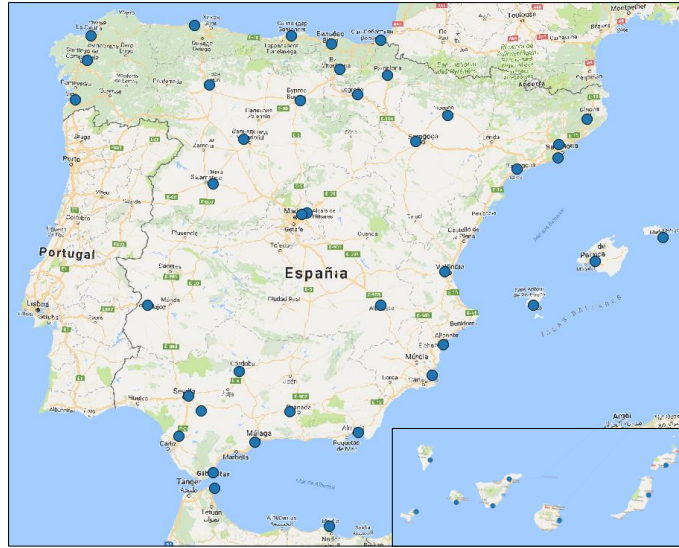


Figura 101. Red de aeropuertos considerada en el estudio

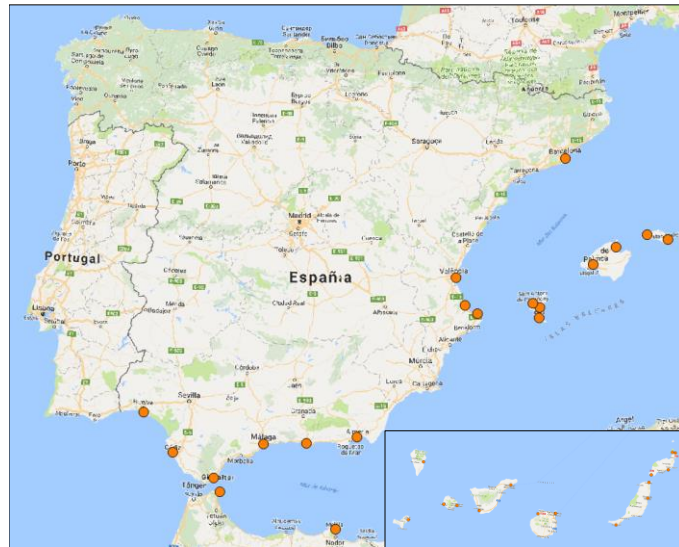


Figura 112. Red de puertos considerada en el estudio

## 3. Metodología

En este epígrafe se describen los diferentes subprocesos que se han llevado a cabo para generar los flujos de movilidad objetivo. Los subprocesos para la extracción de información de movilidad se basan en un conjunto de algoritmos desarrollados por la empresa Kineo, adjudicataria del contrato en unión temporal con Orange España. La secuencia completa de subprocesos es la siguiente:

1. Extracción y anonimización de los registros de telefonía móvil.
2. Pre-procesado y limpieza de los datos.
3. Construcción de la muestra potencial.
4. Extracción de diarios de actividades y viajes.
5. Selección de la muestra efectiva.
6. Elevación de la muestra al total de la población.
7. Validación y ajuste de los resultados.
8. Generación de resultados.

Cada uno de estos subprocesos, así como los algoritmos asociados, se describen en detalle a continuación.

### 3.1 Extracción y anonimización de los registros de telefonía móvil

El primer subproceso consiste en la extracción y anonimización de los CDRs enriquecidos. La anonimización de los CDRs está basada en la utilización de una función hash unidireccional, es decir, una función que permite el cálculo de un identificador anonimizado (similar a un texto aleatorio) a partir del identificador original (habitualmente el IMSI, en el caso de un operador de telefonía) de tal forma que resulta imposible realizar el proceso a la inversa (es decir, una vez aplicada la función no se puede reconstruir el identificador original). Orange ha utilizado lo que se conoce como funciones hash perfectas, que por su diseño evitan las colisiones, es decir, evitan que dos identificadores originales diferentes den como resultado un mismo identificador anonimizado.

Los eventos registrados corresponden tanto a eventos activos (llamadas, mensajes SMS, conexiones a Internet) como a determinados eventos pasivos (por ejemplo, cambios de áreas de cobertura), y por tanto proporcionan una secuencia de las localizaciones en el tiempo.

### 3.2 Pre-procesado y limpieza de los datos

Una vez obtenidos los datos anonimizados, se realiza un pre-procesado de los mismos para facilitar su gestión, ordenando, agrupando y comprimiendo los registros de la forma más eficiente para su posterior manejo. El resultado del pre-procesado permite tener una mayor capacidad de respuesta en fases posteriores del análisis, reduciendo los tiempos de cálculo.

Por otra parte, el proceso de limpieza de datos y depuración de errores es esencial para asegurar la calidad de los datos de partida, evitando que posibles errores de origen desvirtúen los resultados que se obtienen con los algoritmos de extracción de patrones. Las tareas de limpieza de datos y depuración de errores se aplican a los registros anonimizados.

### 3.2.1 Limpieza y depuración de los registros anonimizados

Los registros que muestran la información en un formato no válido son descartados, y se comprueba también la consistencia de los datos registrados en los campos de fecha y hora para eliminar posibles errores.

### 3.2.2 Limpieza y depuración de los datos de red

Los datos de la red de telefonía móvil aportan información sobre las características y localización de las torres de telefonía y la orientación de sus antenas. Como en el caso anterior, los registros que aportan información en un formato no válido, es decir, que no cumplen con las especificaciones definidas por el operador de red, son descartados.

Por otro lado, puede ocurrir que la posición de determinadas torres se haya registrado incorrectamente. La experiencia de las empresas adjudicatarias en proyectos previos demuestra que, en general, los datos de red de los operadores móviles no están nunca exentos de este tipo de errores. En algunos casos, los errores de localización de las torres pueden ser muy evidentes (por ejemplo, torres ubicadas en el mar o en zonas geográficas donde el operador no presta servicio), pero no siempre es así. Por ello, se han desarrollado algoritmos que permiten detectar las torres mal posicionadas a partir del análisis de los eventos registrados en esas torres y de los registros anteriores y posteriores (a modo de ejemplo, un registro en Barcelona precedido y sucedido por una secuencia de registros muy cercanos en el tiempo localizados en el entorno de Madrid, con una diferencia de tiempo incompatible con un viaje en cualquier modo de transporte, puede indicar un error en el posicionamiento de la correspondiente torre de telefonía). Si bien este tipo de errores se presentan en muy limitadas ocasiones, puede tener una influencia importante en los resultados, generando viajes espurios que pueden afectar significativamente a determinadas celdas de la matriz origen-destino, por lo que su depuración resulta fundamental.

## 3.3 Construcción de la muestra potencial

Para construir la muestra potencial (que posteriormente se depurará para dar lugar a la muestra efectiva utilizada en el estudio, tal y como se explica en la sección 3.5), se realiza una selección de los datos con información suficiente para inferir un diario de actividades y viajes de manera fiable, sin perder ninguna actividad relevante en el tiempo o periodo analizado. Dicha selección se lleva a cabo de acuerdo con criterios relacionados con el número de registros generados y el tiempo entre registros. La construcción de la muestra potencial supone por tanto un compromiso entre la cantidad y la calidad de la muestra. En general, la experiencia y las pruebas de validación llevadas a cabo demuestran la conveniencia de obtener una muestra de más calidad, aunque para ello se reduzca ligeramente el tamaño de la muestra, eliminando aquella muestra que podría llevar a no detectar determinados viajes y por tanto falsear los resultados del estudio.

## 3.4 Extracción de diarios de actividades y viajes

El subproceso de generación de los diarios de actividades y viajes es el encargado de transformar los registros de telefonía móvil en información de movilidad, y constituye por ello el núcleo de la solución técnica. El objetivo de este subproceso es generar un diario de actividades y viajes de la muestra potencial y para cada uno de los días de estudio.

La información asociada a cada actividad incluye:

- Localización.
- Tipo de actividad: casa, trabajo, otras actividades frecuentes, otras actividades no frecuentes.
- Hora de inicio.
- Hora de finalización.

La información asociada a cada viaje incluye:

- Origen (igual a la localización de la actividad inmediatamente anterior al viaje).
- Destino (igual a la localización de la actividad inmediatamente posterior al viaje).
- Hora de inicio (igual a la hora de finalización de la actividad anterior)
- Hora de finalización (igual hora de inicio de la actividad siguiente).
- Si se trata de un viaje multietapa, las paradas intermedias (localización, hora de inicio de la estancia, hora de finalización de la estancia).
- Modo de transporte empleado para cada etapa del viaje.

Con el fin de describir con mayor detalle los algoritmos utilizados para inferir esta información, es útil dividir dichos algoritmos en los siguientes grupos: (i) identificación y caracterización de estancias y actividades; (ii) identificación y caracterización de viajes y etapas.

### 3.4.1 Identificación de estancias y actividades

El primer paso consiste en identificar las estancias, es decir, la permanencia durante un cierto tiempo en un lugar determinado. Para ello, se analizan los registros consecutivos en una misma celda de telefonía (o en un conjunto de celdas, dependiendo de la unidad de análisis espacial que se tome como referencia), y se identifican como estancias aquellas localizaciones que perduran durante un tiempo superior a un determinado umbral. A continuación, se identifican aquellas estancias que corresponden a actividades, distinguiéndolas de aquellas que corresponden a estancias intermedias subordinadas a un viaje y que son realizadas entre etapas del mismo. Para ello, los algoritmos combinan distintos criterios basados en:

- Duración de la estancia: por ejemplo, considerando que una estancia por encima de una determinada duración, se considera actividad. Para ello se utiliza un umbral adaptativo en función del tipo de viaje, que tiene en cuenta factores como la localización de la estancia.
- Localización de la estancia: por ejemplo, una estancia de más de 3 horas en un lugar de un núcleo urbano con toda probabilidad es una actividad, mientras que una estancia de 3 horas en un aeropuerto puede corresponder a una conexión entre vuelos y por tanto no constituir una actividad.
- Patrones longitudinales de comportamiento: por ejemplo, el lugar donde se pernocta habitualmente se etiqueta como 'Casa'.
- Información sobre los usos del suelo, que permite refinar la localización de las estancias y actividades y su caracterización: por ejemplo, una celda de telefonía en suelo industrial puede conducir a la localización de la actividad 'trabajo'.



Utilizando estos criterios<sup>3</sup>, se refina la localización de actividades y estancias, asociándolas a un área dentro de la correspondiente celda de telefonía, y se clasifican las actividades en cuatro categorías principales:

1. **Casa.** Su identificación es necesario, porque se utiliza posteriormente en el proceso de elevación muestral. Para su determinación se realiza un análisis longitudinal de los patrones de comportamiento. Los algoritmos descritos en la literatura asignan el lugar de residencia a la localización en la que se permanece durante el período nocturno. Los algoritmos empleados en el proyecto incorporan un conjunto de mejoras con respecto a esta aproximación, permitiendo detectar la muestra que trabaja en horario nocturnos.
2. **Trabajo/estudio.** Se define como la actividad recurrente de una duración significativa (6-8 horas) que se realiza en un lugar diferente al de residencia en función de los usos del suelo.
3. Es importante asimismo puntualizar los siguientes aspectos:
  - i. **Trabajo versus lugar de estudio:** se decidió considerar conjuntamente la información correspondiente a la movilidad obligada recurrente, agrupando a los usuarios que realizan viajes interprovinciales por motivo de trabajo y por motivo de estudio.
  - ii. **Trabajos sin centro de trabajo fijo:** los trabajos en los cuales la actividad principal no se realiza en un centro de trabajo fijo (mensajeros, taxistas, etc.) no son identificados.
4. **Otras actividades frecuentes:** actividades que se realizan de manera recurrente en una misma localización. En este grupo de actividades también pueden estar incluidas actividades de carácter laboral o escolar a tiempo parcial, que no han sido clasificados en el grupo 'trabajo/estudio'.
5. **Otras actividades no frecuentes:** actividades que se realizan sin una frecuencia preestablecida también forman parte del diario de actividades.

### 3.4.2 Caracterización de viajes y etapas. Identificación de modo y ruta

A partir de las estancias y actividades detectadas, se identifican los viajes realizados entre las distintas actividades y las etapas que los componen.

El viaje queda caracterizado al determinar la **hora de inicio y de finalización**, el **modo de transporte** y la **ruta**. Los algoritmos utilizados para determinar estos atributos se basan en la siguiente aproximación:

1. **Análisis de la oferta de transporte:** utilizando la información de la red de transporte (red de carreteras, red ferroviaria, aeropuertos, puertos y estaciones de autobuses) y la oferta de servicios de transporte (conexiones disponibles, horarios, etc.), se identifican todas las alternativas posibles para desplazarse desde el origen al destino del viaje, las etapas que las componen, y su duración.
2. **Determinación de modo y ruta:** los registros espacio-temporales generados durante el viaje se superponen con los itinerarios correspondientes a las distintas alternativas (trazado de las carreteras y vías de ferrocarril, localización de puertos y aeropuerto) mediante técnicas de *map matching*, seleccionando el modo de transporte (y, en el caso de los viajes por carretera y tren, la ruta) compatible

<sup>3</sup> Si bien resulta muy difícil capturar perfectamente toda la casuística existente (por ejemplo, las reuniones celebradas en un aeropuerto pueden motivar viajes y, sin embargo, en determinadas condiciones, esa actividad podría ser erróneamente identificada como una escala entre dos etapas), la experiencia de Kineo Mobility Analytics en anteriores proyectos relacionados con estudios de demanda en corredores de larga distancia demuestra que la solución propuesta captura correctamente los viajes multietapa en la gran mayoría de los casos, y que casos especiales como el anteriormente citado del aeropuerto no suelen afectar de manera significativa a los resultados.

con dichos registros. Pueden presentarse casos en los que aparezca alguna ambigüedad (por ejemplo, si el trazado de una carretera transcurre muy cerca al del ferrocarril de principio a fin, o si dos carreteras alternativas discurren muy próximas), pero en general estos casos son menos usuales en viajes de media y larga distancia y más habituales en los de corta distancia. En esos casos, las técnicas de *map matching* se complementan con criterios basados en velocidades y tiempos de viaje estimados. Si la ambigüedad persiste y siguen existiendo dos o más modos/rutas compatibles con los registros observados, la ruta, se etiquetan temporalmente como ‘indeterminado’ si existe ambigüedad únicamente entre un número concreto de opciones (por ejemplo, si existe ambigüedad si el desplazamiento se ha realizado en tren o en carretera pero se tiene certeza de que no se ha realizado en avión), o ‘desconocido’ si el desplazamiento puede haberse realizado en cualquiera de las opciones de transporte disponibles. Para asignar el modo de transporte y, si procede, la ruta, se reparten esos desplazamientos de acuerdo a las distribuciones empíricas de probabilidad obtenidas del resto de la muestra para ese par origen-destino<sup>4</sup>. Al final de este proceso, cada viaje tendrá asignado un modo entre los siguientes: carretera, ferroviario, aéreo o marítimo. Para el caso de los desplazamientos por carretera se añade información de ruta.

3. **Determinación de la hora de inicio y fin de cada etapa.** Combinando la información sobre la duración de cada etapa para el modo y ruta identificados y la distribución espacio-temporal de los registros generados durante el viaje, se determina la hora de inicio y fin de cada etapa del viaje.
4. **Clasificación de vehículos para los viajes por carretera.** A partir de los patrones de movilidad de los usuarios a lo largo de varias semanas (por ejemplo, número de kilómetros recorridos, estancia en centros logísticos, etc.) y la movilidad específica del día de análisis, se divide a los usuarios que viajan por carretera en dos grupos, usuarios con movilidad característica de transportistas (vehículos pesados de mercancías) y resto de usuarios (vehículos ligeros y autobús), con el fin de excluir los viajes de vehículos pesados de mercancías de la matriz de viajes interprovinciales de pasajeros. A continuación, para los pares origen-destino entre provincias de distintas comunidades autónomas, se emplea la información de oferta y demanda de las concesiones de autobuses del Ministerio de Fomento para realizar la segmentación entre vehículo privado y autobús. Es necesario reflejar que los algoritmos empleados para la identificación del modo autobús tienen un carácter experimental, se han probado por primera vez en el marco de este proyecto, y por ello para el análisis o comparación entre flujos de diferentes O/D es aconsejable realizarla sobre el total del modo “carretera”.

### 3.5 Selección de la muestra efectiva

Una vez generados los diarios de actividades y viajes, se genera la muestra efectiva, que busca evitar cualquier error debido al hecho de que existen más móviles que población. La solución incluye un algoritmo para depurar los datos, en función del nivel de similitud y coincidencias de las actividades y viajes detectados durante el periodo de estudio.

---

<sup>4</sup> Este proceso de reparto empírico para determinar el modo y/o ruta de los viajes a partir de las distribuciones de probabilidad observadas en la muestra se considera la mejor aproximación cuando únicamente se dispone de información extraída de los datos de telefonía móvil. Este cálculo se ha realizado para validar que la información obtenida únicamente con datos de telefonía móvil proporciona resultados ya de por sí satisfactorios. Para la obtención de los resultados finales, al disponer de información de demanda de viajes de modos colectivos, se utilizará dicha información para resolver la incertidumbre de los casos ‘indeterminados’ y ‘desconocidos’ obtenidos mediante telefonía móvil.

## 3.6 Elevación de la muestra al total de la población

Finalmente, una vez seleccionada la muestra efectiva, para obtener los indicadores requeridos es necesario elevar dicha muestra al total de la población objeto de estudio. Cabe distinguir dos casos: la elevación de la muestra de residentes en España y la elevación de la muestra de no residentes.

### 3.6.1 Elevación de la muestra de residentes en España

Para los residentes en España, se han utilizado factores de elevación basados en la residencia inferida a partir de los datos analizados, tomando como marco muestral el Padrón Municipal de Habitantes. La aplicación de factores basados en la residencia debe elaborarse siempre sobre las distintas zonas del territorio, para corregir la heterogeneidad de la cuota de mercado de cualquier operador de telefonía. De esta forma cada provincia tiene un factor de expansión y a nivel nacional este factor medio para residentes ha sido de 8,29.

### 3.6.2 Elevación de la muestra de no residentes

Para los no residentes en España, se utiliza como marco muestral la encuesta FRONTUR realizada por el INE, que recoge datos relativos a la entrada en España de visitantes no residentes en España. La muestra efectiva de usuarios extranjeros se clasifica por nacionalidad y tipología de extranjero (turista o excursionista). Para clasificar a los visitantes como turistas o excursionistas se lleva a cabo un análisis longitudinal de sus patrones de comportamiento a lo largo de toda su estancia en el territorio nacional, para determinar si el viajero pernocta en España (turista) o no (excursionista). Los valores de elevación obtenidos para los no residentes han sido en su mayoría inferiores a 20, salvo para los casos de EE.UU., Resto de América y Resto del Mundo que lo superan. Resto de América supera el valor de 100.

## 3.7 Validación de resultados

Los resultados de demanda de transporte obtenidos a partir de datos de telefonía móvil se han analizado y contrastado con las otras fuentes de datos de demanda de transporte anteriormente reflejadas. En algunos casos, es posible evaluar la correspondencia de los resultados con los proporcionados por otras fuentes: un ejemplo es el contraste del número de etapas identificadas en modo ferrocarril con origen en una determinada estación con el número total de pasajeros que inician una etapa de su viaje en dicha estación (salvo en algunos casos particulares donde no se dispone de información para ciertos servicios, como los servicios de Cercanías). En otros casos, sólo es posible evaluar la coherencia de los mismos con otras fuentes de datos: un ejemplo es el contraste con datos de demanda portuaria, donde solo se dispone de un volumen de ciertos tipos de viajes obtenido a partir de los datos facilitados por los operadores, con una agregación mensual.

Para evaluar la validez de los resultados obtenidos mediante telefonía móvil, se han llevado a cabo análisis de correlación, comparando los resultados obtenidos para cada uno de los modos de transporte con los datos de estadísticos de demanda. Se ha evaluado si la estructura de los viajes identificada con telefonía móvil es similar a la reportada por los operadores (comparando el coeficiente de correlación) y la capacidad de información de telefonía móvil para estimar el volumen total de viajes (analizando la pendiente de la recta de ajuste). En los casos en los cuales se han encontrado discrepancias significativas entre los resultados



obtenidos a partir de los datos de telefonía móvil y la información de referencia cierta, se ha procedido a reanalizar los algoritmos utilizados hasta verificar que las diferencias observadas correspondan únicamente a las limitaciones intrínsecas de los datos de telefonía móvil.

### 3.8 Ajuste de resultados

Los resultados obtenidos a partir de telefonía móvil se han ajustado con la siguiente metodología:

Según se ha explicado en la sección 3.4.2, en función del grado de certeza con la que se determina el modo de transporte a partir de los datos de telefonía móvil, los viajes identificados se clasifican en tres grupos:

- **Determinados:** son aquellos casos (normalmente la mayoría), para los que se determina con certeza el modo de transporte a partir de técnicas de *map matching*.
- **Indeterminados:** son aquellos casos en los que se conoce con certeza el subconjunto de opciones de transporte que son compatibles con la información de telefonía móvil generada por el dispositivo móvil a lo largo del viaje. Por ejemplo, es posible determinar que un viaje Madrid-Valencia no se ha realizado en avión, pero no se puede determinar con certeza si el viaje se ha realizado en ferrocarril o por carretera.
- **Desconocidos:** son aquellos casos en los que cualquiera de las opciones de transporte es posible.

Para definir el método de ajuste de resultados, se han tenido en cuenta los siguientes criterios:

1. Tomar como punto de partida la información obtenida de los datos de telefonía móvil, que constituye la base del estudio, con el fin de utilizar la información procedente de la telefonía móvil que no es posible obtener por otras fuentes, en particular los orígenes y destinos puerta-a-puerta y la demanda de viajes en vehículo privado. Con objeto de no distorsionar artificialmente los resultados obtenidos mediante telefonía móvil, se ha empleado un método de ajuste que proporciona valores compatibles con dichos resultados.
2. Utilizar los datos de proporcionados por los operadores como referencia dentro del intervalo de valores posibles proporcionado por la telefonía móvil.
3. Emplear una aproximación desagregada, incorporando los viajes 'indeterminados' y 'desconocidos' a cada modo de transporte, en lugar de aplicar un factor de ajuste agregado, con el objeto de conservar las características de relaciones origen-destino puerta-a-puerta de los viajes muestrales y de mantener la consistencia entre los distintos resultados generados (matriz de viajes, matriz de etapas).

Teniendo en cuenta estos criterios, el método de ajuste seleccionado es el siguiente:

- Para cada par origen-destino, el número mínimo de viajes en un modo concreto de transporte lo fija el valor de viajes 'determinados' identificados en dicho modo a partir de los datos de telefonía móvil. El máximo viene determinado por el valor más alto compatible con la información obtenida también mediante telefonía móvil. Así, una vez aplicado el ajuste, el volumen de viajes en cada modo de transporte para un par origen-destino está siempre contenido entre:
  - el número de viajes 'determinados' identificados para ese modo;
  - la suma de los viajes 'determinados' identificados para ese modo, los viajes 'indeterminados' para los que dicho modo es una alternativa posible, y los viajes con modo 'desconocido'.

- Si el volumen de viajes ‘determinados’ es superior a la información de demanda proporcionada por los operadores, se mantiene el valor de viajes determinados obtenido a partir de los datos de telefonía. De esta forma, el método es robusto frente a posibles deficiencias o falta de información en los datos de demanda procedentes de fuentes alternativas a la telefonía móvil, permitiendo realizar estimaciones que superan las limitaciones de las otras fuentes de datos. Por ejemplo, los datos de billeteo de Renfe Operadora disponibles no incluyen la demanda en Cercanías, por lo que para los pares origen-destino en los que existen dichos servicios, se toma como dato de demanda en ferrocarril el proporcionado por la telefonía móvil, que además de los servicios de larga distancia incorpora también los viajes realizados en Cercanías. Otro ejemplo relevante son los datos de transporte marítimo, en los que los datos de telefonía permiten capturar viajeros de cruceros o embarcaciones privadas o de recreo cuyos viajes no están incluidos en los datos de demanda disponibles.
- Si el volumen de viajes determinados para un determinado modo es inferior al reportado por los operadores, se tiene en cuenta el volumen de viajes ‘indeterminados’ en modo y los casos ‘desconocidos’, reevaluando su incorporación al grupo de ‘determinados’. Los viajes ‘indeterminados’ y ‘desconocidos’ se ordenan según su grado de similitud con los distintos modos, de forma que los viajes con mejores métricas de similitud para cada modo son los primeros en pasar al grupo de ‘determinados’ de dicho modo. En primer lugar, se ajustan los casos ‘indeterminados’ para cada uno de los modos. Posteriormente, se ajustan los casos ‘desconocidos’. El orden de prioridad del ajuste se basa en la calidad y fiabilidad de la información de demanda disponible para realizar el ajuste. Teniendo en cuenta este criterio, el orden de prioridad empleado es el siguiente: ferrocarril, avión y autobús (para el modo marítimo no se realizar ningún ajuste por los motivos expuestos anteriormente). Para realizar el ajuste del modo ferroviario en los pares origen-destino que incluyen servicios de Cercanías, para los cuales no se dispone de datos de demanda, el reparto de los viajes ‘indeterminados’ y ‘desconocidos’ se efectúa de forma probabilística, basándose únicamente en la información obtenida a partir de datos de telefonía móvil, según el reparto de viajes ‘determinados’ observado, siguiendo el método descrito en la sección 3.4.2.

Con el proceso de ajuste de resultados se consigue una mejora en la estimación de la demanda en transporte público, así como refinar el volumen total de viajes en vehículo privado. En la Figura 13 se muestra, a modo de ejemplo, la comparativa de los resultados para los modos ferrocarril y aéreo antes y después del ajuste.

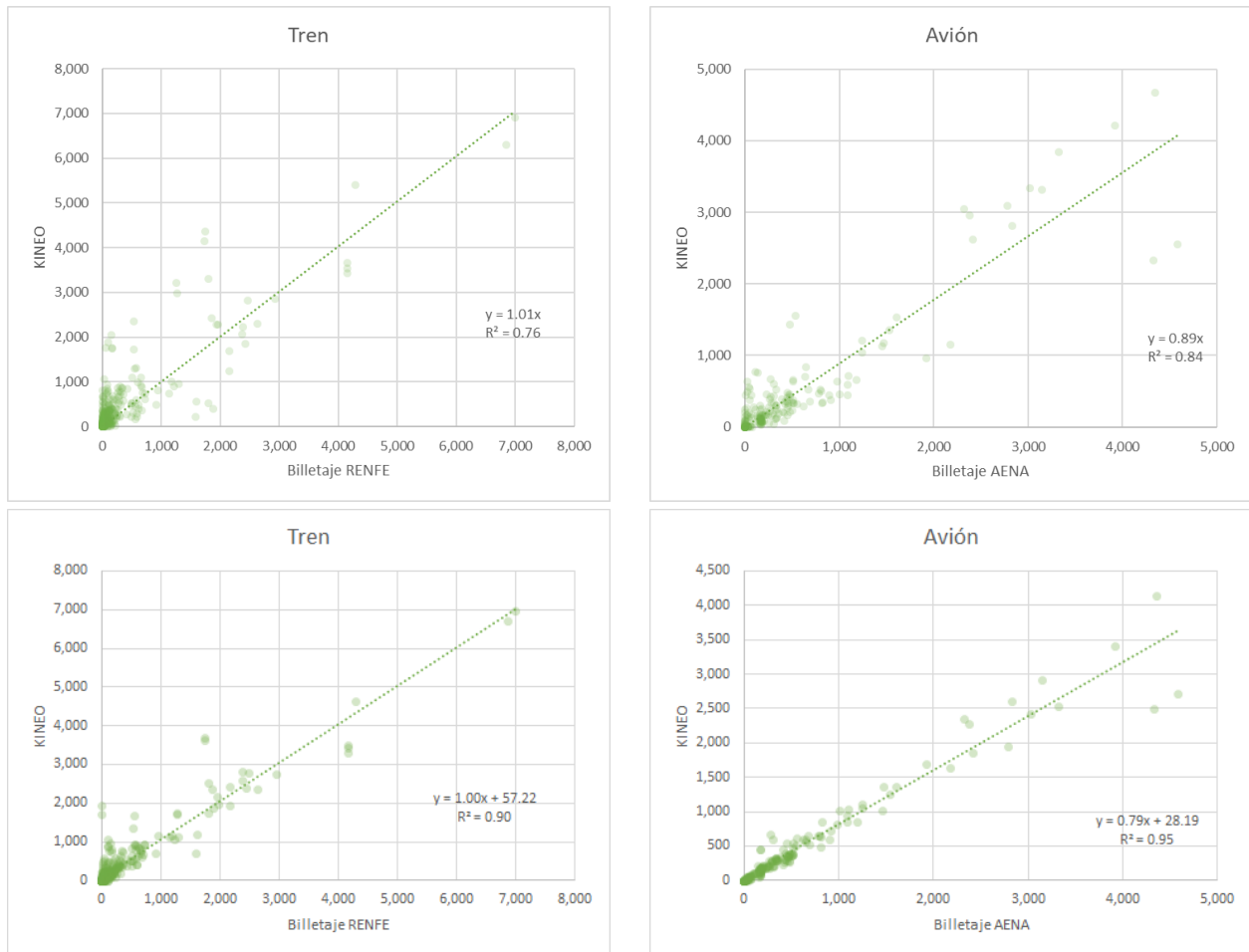


Figura 123. Comparativa de los resultados para los modos ferrocarril y aéreo antes y después del proceso de ajuste para un día tipo. En la fila superior se muestran los resultados antes del ajuste. En la fila inferior se muestran los resultados obtenidos tras el ajuste. En el modo ferroviario se han eliminado los pares O/D con servicios de cercanías

### 3.9 Generación de resultados

Una vez que los resultados han sido validados y ajustados, se generan las matrices de flujos.

Para asegurar el cumplimiento del Reglamento General de Protección de Datos (GDPR), se ha verificado que el tamaño poblacional asociado a los distintos grupos de viajeros, según los distintos criterios de segmentación empleados, es suficientemente grande como para evitar cualquier riesgo razonable de reidentificación de algún individuo mediante un hipotético proceso de fusión con otras fuentes de datos. Esta condición se cumple en todos los casos, por lo que no ha sido necesario agregar los resultados a un nivel superior.

## 4. Resultados del trabajo

El Ministerio de Fomento no recibió como resultado del trabajo los datos anonimizados del posicionamiento de los terminales, sino el producto de aplicar la tecnología big data a dichos datos expresados en matrices de flujos de viajes y etapas.

En este sentido, el Ministerio recibió como resultado:

- **Matrices de viajes/etapas:** un fichero Excel por periodo, con el nombre “periodo.xlsx”, donde ‘periodo’ puede tomar los valores de ‘julio’ u ‘octubre’. Cada uno de estos ficheros contiene la siguiente información:
  - **IND:** Portada e índice del fichero, con hipervínculos que permiten explorar los resultados de manera cómoda y eficiente.
  - **C\_X:** conjunto de pestañas con la descripción de los campos de las distintas tablas de resultados. ‘X’ representa el tipo de resultado generado y puede tomar los valores VI (viajes) o ET (etapas).
  - **X\_PAA:** conjunto de pestañas con las distintas tablas de resultados. ‘X’ representa el tipo de resultado generado y puede tomar los valores VI (viajes) o ET (etapas). ‘P’ representa el periodo de análisis y puede tomar los valores ‘J’ (julio/agosto) u ‘O’ (octubre). ‘AA’ es un código numérico (por ejemplo 02) que representa el identificador del grupo de días.

Tabla 2. Especificación de la matriz de viajes/etapas

Matrices de viajes/etapas			
Columna	Nombre	Valores	Descripción
1	Origen	59 zonas	Identificador de la zona de origen del viaje: código postal provincia <sup>5</sup>
2	Destino	59 zonas	Identificador de la zona de destino del viaje: código postal provincia <sup>6</sup>
3	Año	2017	Año de estudio
4	Mes	07-08-10	Meses de estudio
5	Día	01-31: días específicos Lunes-domingo: días promedios	Nombre del día de la semana para los días promedio y número de día para los días sueltos
6	Periodo	P1: 00:00 - 06:00 P2: 06:00 - 10:00 P3: 10:00 - 17:00 P4: 17:00 - 00:00	Franja horaria correspondiente al inicio del viaje
7	Modo	autobús/avión/barco/carretera/privado/tren	Modo de transporte principal del viaje / utilizado en etapa
8	Distancia	D1: viajes entre 10 y 50 km. D2: viajes de más de 50 km.	Rango de distancia al que pertenecen los viajes
9	Residencia	Nacionalidad según los valores de la <i>Tabla 1</i>	País o grupo de países de residencia de los viajeros
10	Actividad_Origen	Casa/trabajo-estudio/ frecuente/ esporádica	Tipo de actividad realizada en el lugar de origen del viaje
11	Actividad_Destino	Casa/trabajo-estudio/ frecuente/ esporádica	Tipo de actividad realizada en el lugar de destino del viaje
12	Viajeros	Número decimal (3 dígitos)	Número de viajeros
13	Viajeros-km	Número decimal (3 dígitos)	Número de viajeros multiplicado por los kilómetros del viaje

<sup>5</sup> Para las Islas ver epígrafe de zonificación

<sup>6</sup> Para las islas ver epígrafe de zonificación

