

## Resumen

La rápida digitalización de la movilidad y el uso generalizado de dispositivos móviles personales ha dado lugar a una nueva generación de fuentes de *big data* para la caracterización de la movilidad. Este artículo hace un repaso por las oportunidades, retos y estrategias de análisis relacionadas con estas fuentes. Se presentan dos casos de uso que ilustran la aplicación de estos datos al estudio de la movilidad en España, y se apuntan las líneas de investigación que permitirán seguir avanzando en la realización de análisis de demanda de transporte basados en estas fuentes.

*Palabras clave:* movilidad, *big data*, datos de telefonía móvil, demanda de transporte.

## Abstract

The increasing digitalisation of the mobility sector and the pervasive use of personal mobile devices has led to a new generation of big data sources for the characterisation of mobility patterns. This article reviews the opportunities, challenges and analysis strategies associated with these new sources. Two use cases are presented to illustrate the application of these data sources to the study of mobility patterns in Spain. The article also highlights the ongoing research efforts that will continue expanding our capabilities to characterise travel demand patterns with big data.

*Keywords:* mobility, big data, mobile network data, travel demand.

*JEL classification:* R41.

# ESTUDIO DE LA MOVILIDAD CON LA TECNOLOGÍA *BIG DATA*: POSIBILIDADES POR EXPLORAR

Javier BURRIEZA-GALÁN

*Nommon Solutions and Technologies*

## I. INTRODUCCIÓN

La movilidad juega un papel crucial en el desarrollo de nuestras sociedades. La capacidad para trasladarnos entre las actividades que realizamos diariamente es vital para garantizar nuestra calidad de vida y fomentar la interacción social. Al mismo tiempo, la movilidad requiere de una gran cantidad de recursos para su funcionamiento, por lo que su gestión ineficiente puede dar lugar a multitud de impactos adversos. De esta manera, la movilidad no es solo una práctica habilitadora del desarrollo socioeconómico, sino que también es parte fundamental de muchos de los retos de sostenibilidad a los que nos enfrentamos hoy en día.

La gestión eficiente de los sistemas de movilidad requiere de un conocimiento profundo de los patrones de demanda de transporte. Identificar y caracterizar los flujos de movilidad es imprescindible para mejorar la asignación de recursos y diseñar medidas que permitan alinear las necesidades de desplazamiento con los objetivos de sostenibilidad ambiental, social y económica. Desde mediados del siglo XX, la ingeniería de transportes ha dedicado ingentes esfuerzos a desarrollar métodos de medida de la movilidad que permitan obtener dicha información. Un primer conjunto de instrumentos está formado por aquellos dedicados

al conteo de viajeros o vehículos, que permiten cuantificar los flujos en ciertos puntos del sistema de transporte, pero no aportan información acerca de las características de cada viaje (origen y destino, modos elegidos durante el viaje, rutas seguidas en la red, perfil del viajero, etc.). Por ello, los instrumentos de conteo han sido tradicionalmente complementados con un segundo tipo de técnicas: las encuestas de movilidad. Estas encuestas consisten en suministrar un cuestionario a los viajeros para que estos puedan responder a cuestiones acerca de los desplazamientos que han realizado, su valoración de cómo el sistema de transportes satisface sus necesidades de movilidad, o hasta qué punto usarían nuevos servicios e infraestructuras si estos se creasen. La combinación de conteos y encuestas ha permitido avanzar en la capacidad de diagnóstico de la movilidad, pero cuenta con limitaciones importantes. Por un lado, el coste elevado de las encuestas restringe el alcance espacial del ámbito de estudio, el tamaño de muestra alcanzado y las posibilidades de monitorización de la demanda de transporte a lo largo del tiempo. La complejidad creciente de los sistemas de movilidad, con nuevos modos de transporte y más variabilidad en los comportamientos de la población, hace que estas limitaciones sean aún más acuciantes. Por otro lado, las encuestas presentan sesgos difíciles de subsanar en muchos casos,

procedentes tanto de la construcción de la muestra como de la disposición de los entrevistados.

Ante este panorama, la digitalización supone una oportunidad para mejorar la recogida de datos de demanda de transporte. En los últimos años, la proliferación de dispositivos móviles asociados a personas y vehículos, los avances en sensorización conectada y las tecnologías de geolocalización han dado lugar a una serie de fuentes de datos masivos o *big data* que pueden aportar información valiosa acerca de los patrones de actividad y movilidad de la población. El *big data* se caracteriza por su gran volumen y su elevada frecuencia de muestreo, lo que suele implicar una mayor variabilidad intrínseca, al contener un mayor número de situaciones diferentes (Laney, 2001). En su aplicación al estudio de la demanda de transporte, hay que tener en cuenta un elemento común adicional a la mayoría de fuentes *big data*: no han sido creadas con el propósito de analizar la movilidad, sino que son un producto secundario de ciertos servicios (comunicación entre teléfonos, aplicaciones móviles, cobro mediante tarjeta bancaria, etc.). Esto añade complejidad a los métodos de análisis para extraer información de demanda de transporte, ya que hay que abordar cuestiones como la privacidad de los datos personales o la limpieza de datos no relevantes para el estudio. Sin embargo, esta característica es una ventaja fundamental en términos de eficiencia: el dato ya está ahí, esperando su análisis. Esto ha acelerado la aparición y adopción de multitud de soluciones de explotación de fuentes de datos masivas para la caracterización de la movilidad.

Este artículo hace un repaso por el estado actual de estas tecnologías *big data* para el estudio de la movilidad, presentando una serie de casos de aplicación en España y apuntando las líneas de investigación en marcha relacionadas con la explotación de estas nuevas fuentes de datos.

## II. EL USO DEL *BIG DATA* EN LA PLANIFICACIÓN Y GESTIÓN DE LA MOVILIDAD

### 1. Las nuevas fuentes de datos

La diversidad de tecnologías de información y comunicación que han surgido en los últimos años hace que las fuentes de datos asociadas sean de naturaleza muy variada. No obstante, cabe distinguir entre dos grandes grupos: las procedentes de dispositivos móviles y las procedentes de sensores.

Los dispositivos móviles generan registros de su ubicación a lo largo del tiempo. Esto permite reconstruir los patrones de movimiento de las personas, mercancías o vehículos a los que está asociado cada dispositivo. Entre las fuentes procedentes de dispositivos móviles aptas para estudiar la movilidad de personas, destacan: i) los datos de telefonía móvil, generados por la interacción de los teléfonos móviles con la red de antenas del operador de telefonía móvil; ii) los datos de posicionamiento GPS (*global positioning system*) generados por ciertas aplicaciones de los teléfonos móviles inteligentes y otros dispositivos; iii) los datos de sistemas inteligentes de pago de transporte público, generados por la interacción de las tarjetas inteligentes de los

usuarios con las terminales de pago instalados en estaciones y/o vehículos; y iv) los datos de tarjetas bancarias, generados por la interacción de dichas tarjetas con las terminales de pago instalados en puntos de venta de bienes y servicios. Muchas de estas fuentes incluyen datos que van más allá de un mero registro espacio-temporal, lo que permite contextualizar las trayectorias de las personas. Un ejemplo son los datos procedentes del uso de aplicaciones de redes sociales, en las cuales se conoce el contenido del mensaje emitido por el usuario en un instante y lugar determinado, lo que permite analizar qué actividad está realizando o cuál es su valoración acerca de lo que ve (Ayfantopoulou *et al.*, 2020). Además de estas fuentes asociadas a personas, existen otras asociadas a vehículos: por ejemplo, los datos de navegadores GPS de conducción, que permiten registrar la trayectoria y velocidad de los vehículos en sus rutas.

Los sensores generan registros de los objetos que pasan por el radio de acción de cada sensor. Por un lado, existen sensores puntuales, que registran lo que ocurre en una única ubicación sin posibilidad de trazar trayectorias. Algunos de estos ya se vienen empleando desde hace décadas para cuantificar los flujos de vehículos, como es el caso de los aforos de tráfico basados en tubos neumáticos o espiras magnéticas. Por otro lado, existen redes de sensores de intervalo, que permiten trazar el recorrido de personas o vehículos al registrar su presencia en una secuencia de sensores desplegados en el territorio. Las tecnologías detrás de estos sensores son el reconocimiento automático de imágenes (como es el caso

de los sensores de matrícula de vehículos), el wifi y el *bluetooth* (aplicable tanto a receptores embarcados en vehículos como a los asociados a dispositivos móviles personales).

## 2. Oportunidades y retos del *big data* para la movilidad

El uso de nuevas fuentes de datos para caracterizar la movilidad conlleva una serie de oportunidades y retos. Hay tres características básicas que determinan estas oportunidades y retos: son recogidas para otros propósitos, sin necesidad de interacción directa con las personas o vehículos en movimiento, y se registran de manera continua.

Muchas de estas fuentes no se recogen con el propósito de caracterizar la movilidad, sino que aparecen como subproductos de algunos servicios a la ciudadanía, tales como la telefonía móvil, el uso del transporte público o el apoyo a la navegación de conductores. Esto abarata su recogida en comparación a las encuestas de movilidad, ya que el coste está internalizado en la operación del servicio que da lugar a la propia fuente. Esta reducción de costes facilita una actualización más frecuente de la información de demanda de transporte en comparación a las encuestas (Toole *et al.*, 2015), lo que es especialmente relevante en un contexto de cambios constantes en la movilidad. Además, la concurrencia de datos que permiten inferir desplazamientos y otros datos puramente asociados al propósito original de la fuente facilita el desarrollo de modelos explicativos basados en variables latentes, complejas de cuantificar de manera directa (Milne y

Watling, 2019). Un ejemplo es la relación entre movilidad y consumo en el caso de los datos de tarjeta bancaria (Lenormand *et al.*, 2015). Aun con estas oportunidades, la explotación de datos para un fin distinto del que originó su recogida no está exenta de retos. En primer lugar, el acceso a la fuente de datos no es inmediato, ya que el gestor del servicio que origina la fuente debe poner los datos a disposición del interesado sin contravenir sus propios intereses. Si se trata de gestores de servicios privados (operadores de telefonía móvil, aplicaciones móviles, bancos...) este obstáculo se resuelve, bien mediante la creación de equipos de analítica de datos dentro de la empresa que provee el servicio (generando una nueva línea de negocio que vende la información extraíble del dato, secundaria a la del servicio que origina la fuente), bien mediante acuerdos de monetización de los datos para su cesión a terceros especializados en el manejo de grandes volúmenes de datos para caracterizar la movilidad. Si se trata de gestores de servicios públicos externalizados (por ejemplo, operadores de transporte), las administraciones competentes tienen que asegurar que los datos generados por el servicio puedan estar a disposición de los planificadores y gestores de la movilidad si resulta conveniente. En segundo lugar, la mayoría de estas fuentes de datos necesitan la identificación de la persona o vehículo al que corresponde cada registro para cumplir con su función original, por lo que contienen información sensible de carácter personal. Por tanto, es vital garantizar la privacidad mediante técnicas para anonimizar que cumplan todas las garantías legales, sin afectar de manera innecesaria a las posibilidades de análisis que ofrecen

las fuentes. Finalmente, una vez que los datos estén disponibles y cumplan con las garantías de privacidad necesarias, el analista se encontrará con elementos sobrantes (que no interesan para el estudio de la movilidad) y con elementos faltantes (que le gustaría conocer del contexto de cada registro y no se contemplan en la fuente). Por ello, el procesamiento de los datos tiene que tener en cuenta su propósito original. Esto es clave para diseñar metodologías de limpieza y filtrado que aligeren el procesamiento sin mermar la calidad de los datos de partida, así como estrategias para mitigar los posibles sesgos muestrales. Cerrar la brecha entre los datos presentes en la fuente y los datos necesarios para completar el análisis requiere el desarrollo de metodologías de fusión de datos.

Otra característica básica de estas fuentes y que determina algunas de sus ventajas y limitaciones es la recogida pasiva, ya que los datos se registran sin intervenir en las trayectorias de personas o vehículos. Por un lado, esto, junto al abaratamiento que se deriva de la recogida por otros propósitos, permite un salto de orden de magnitud en los tamaños de muestra alcanzados por las encuestas de movilidad (Zhang *et al.*, 2019). A mayor tamaño de muestra, menor incertidumbre asociada a los resultados del análisis, especialmente para flujos de movilidad reducidos o para el estudio de distintas segmentaciones de la demanda. Esto es importante en un contexto de dispersión urbana creciente, donde la heterogeneidad en los patrones de movilidad es cada vez mayor. Además, esto permite conocer mejor los hábitos de desplazamiento de grupos de población muy concretos que

pueden ser de interés para la planificación de la movilidad (por ejemplo, personas de la tercera edad). Por otro lado, la recogida pasiva permite reducir algunos de los sesgos en los que incurrir las encuestas de movilidad: el falseamiento en las respuestas, la omisión de algunas preguntas del cuestionario o la imposibilidad de recoger una muestra no sesgada en ciertas ocasiones (por ejemplo, en las encuestas a conductores en vías de alta capacidad, que han de realizarse en gasolineras para no interrumpir el tráfico, sesgando la muestra hacia cierto tipo de desplazamientos). Junto a estas ventajas, la recogida pasiva conlleva una limitación importante del alcance de estas nuevas fuentes, ya que se trata de instrumentos que solo pueden recoger preferencias reveladas de los individuos: cuáles han sido sus desplazamientos bajo las condiciones existentes en el momento de realizarse. Por el contrario, las encuestas de movilidad pueden combinar esta perspectiva con un análisis de las preferencias declaradas de los individuos: cómo modificarían sus comportamientos bajo otras condiciones (como, por ejemplo, la creación de nuevas infraestructuras o servicios de transporte). De nuevo, la brecha entre el alcance de las fuentes recogidas de manera pasiva y el alcance necesario para el estudio de la movilidad en ciertas situaciones ha de ser cubierta por metodologías de fusión de datos.

Además de tener habitualmente otros propósitos y de no intervenir en los desplazamientos de la población estudiada, la recogida de las fuentes de datos *big data* es continua. Las oportunidades que esto ofrece son inmensas, a expensas de requerir el manejo de grandes volúmenes

de datos. En primer lugar, ofrecen la posibilidad de realizar una actualización constante de la información de demanda de transporte, lo que puede ser determinante en la gestión a corto plazo de la oferta disponible. Por ejemplo, permiten detectar desviaciones en los patrones habituales de tráfico que puedan incrementar la probabilidad de accidentes. En segundo lugar, su recogida continua alimenta bases de datos históricas, muy valiosas por dos motivos: i) hacen posible seleccionar períodos de tiempo del pasado que sean de interés por alguna circunstancia particular (por ejemplo, evaluar la influencia de eventos inesperados en la movilidad); y ii) hacen posible el desarrollo de modelos explicativos y predictivos de la movilidad mediante el uso de técnicas de inteligencia artificial.

### 3. Fusión de datos para un diagnóstico completo

Tal como ocurre con los métodos tradicionales de recogida de datos, ninguna de las nuevas fuentes *big data* para el estudio de la movilidad resuelven por sí solas los retos expuestos en la sección anterior. En un contexto de creciente disponibilidad de fuentes de datos, las metodologías de fusión de datos cobran aún más relevancia. No se trata únicamente de combinar las nuevas fuentes entre sí, sino de elaborar estrategias para un uso conjunto de las fuentes tradicionales y las fuentes *big data* (Chen *et al.*, 2016). Algunos ejemplos de fusión de datos son los siguientes:

- La información de demanda que se puede obtener a partir de las fuentes *big data* puede servir para optimizar el diseño de las campañas de trabajos de

campo asociadas a las fuentes tradicionales (encuestas, aforos, etc.). Por ejemplo, se pueden obtener conclusiones acerca de las fechas más representativas de los flujos de movilidad mediante la exploración de información longitudinal de demanda procedente de datos de telefonía móvil o de datos de sistemas inteligentes de pago de transporte público. Además de dar soporte en la decisión de cuándo ejecutar estos trabajos de campo, también pueden ayudar a repartir los esfuerzos en el territorio. Los datos de telefonía móvil, que tienen una muestra muy amplia, pueden ayudar a identificar aquellos lugares con una mayor variabilidad en sus patrones de movilidad, lo que puede servir para priorizar ciertas zonas a la hora de determinar el tamaño de muestra de la encuesta necesario en cada zona.

- Algunas de las fuentes *big data* más empleadas, como los datos de telefonía móvil, miden el flujo de personas. Sin embargo, en algunos estudios de movilidad el objeto de análisis es el flujo de vehículos. Más allá de aproximaciones sencillas basadas en suponer valores medios de ocupación de los vehículos para hacer una conversión entre viajes-persona y viajes-vehículo, es posible usar los datos de aforos de tráfico para minimizar el error de esta conversión, mediante la comparación de los flujos de personas medidos mediante las fuentes *big data* en ciertos puntos de la red de carreteras con los conteos de vehículos que ofrecen los aforos de tráfico.

- La precisión y alcance espacial de las nuevas fuentes de datos no es homogénea. Por ejemplo, los datos de telefonía móvil permiten determinar la

ubicación del dispositivo con una precisión de centenares de metros (dependiendo del área de cobertura de las antenas), mientras que los datos procedentes de aplicaciones móviles tienen generalmente una mayor precisión, al referirse a la ubicación GPS. Sin embargo, los datos de telefonía móvil presentan muchas ventajas frente a los de ubicación GPS, ya que suelen tener tamaños de muestra muy superiores y una mayor regularidad en los registros que genera cada usuario. En este contexto, la asignación de unas coordenadas concretas a los registros de telefonía móvil puede realizarse con menor error si se conoce la distribución espacial de la presencia de personas dentro del área de cobertura de la antena mediante datos GPS.

– Una de las limitaciones más habituales de las nuevas fuentes de datos de movilidad es que carecen de suficiente información sociodemográfica de los agentes presentes en sus muestras. Esto se puede solventar recurriendo a técnicas de inteligencia artificial que exploten los datos de encuestas de movilidad (en los cuales sí hay presente un gran número de características de los agentes de la muestra, junto a sus patrones de movilidad revelados). Estas técnicas identifican de manera automática las relaciones existentes entre las variables sociodemográficas que no están presentes en la fuente *big data* con otras variables sociodemográficas y de movilidad que están presentes tanto en la encuesta como en la fuente *big data*. Una vez desarrollados, estos modelos de clasificación pueden emplearse para imputar las características faltantes a la base de la fuente *big data*.

### III. CASOS DE USO EN ESPAÑA

España es uno de los países del mundo con mayor experiencia en el uso de nuevas fuentes de datos para la planificación y gestión del transporte. En particular, España ha sido pionera en la aplicación de datos de telefonía móvil para la obtención de matrices origen-destino a gran escala. Entre los numerosos proyectos realizados por agentes públicos y privados en los últimos años, destacan dos estudios promovidos por el Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA) que ilustran las oportunidades que representan estas nuevas fuentes de datos para mejorar el conocimiento sobre la demanda de transporte.

#### 1. Planificación del transporte: el análisis de la movilidad interprovincial de 2018

En el marco de sus competencias de planificación de transporte a nivel estatal, el MITMA realiza estudios de demanda de movilidad que abarcan todo el territorio español. En el pasado, estos estudios se basaban en las llamadas encuestas *Movilia* (con ediciones en los años 2000-2001 y 2006-2007). Dadas las dificultades para realizar encuestas de movilidad con este amplio alcance, el MITMA promovió en 2018 un análisis de la movilidad interprovincial de viajeros en España basado en datos de telefonía móvil.

El objetivo de este estudio fue obtener matrices origen-destino de viajes entre provincias en distintos períodos del año 2017 (tanto estival como no estival, cubriendo varios tipos de día de la semana). Estas matrices ori-

gen-destino están segmentadas por modo de transporte (carretera, ferrocarril, aéreo y marítimo); período del día en el que tiene origen el viaje; y distintas características del viaje y del viajero (tales como el lugar de residencia o el propósito del viaje). Los desplazamientos analizados fueron los de distancias superiores a 50 km, salvo para las provincias con mayor número de viajes interprovinciales (Madrid, Barcelona, Alicante y Vizcaya), para las cuales se analizaron todos los desplazamientos de más de 10 km.

Las matrices origen-destino fueron obtenidas mediante la solución analítica de Nommon Solutions and Technologies, empresa española especializada en el análisis de nuevas fuentes de datos para caracterizar la movilidad de la población, a partir de los registros anonimizados de telefonía móvil procedentes de la red de Orange España (MITMA, 2019). Esta solución integra los datos de telefonía móvil junto a datos de usos del suelo (que permiten mejorar la precisión espacial con la que se ubican las actividades); datos de la oferta y red de servicios de transporte (que permiten interpretar los registros de telefonía generados durante los viajes para identificar el modo de transporte y la ruta seguida por el viajero); datos sociodemográficos (que permiten expandir la muestra) y otros datos de demanda de transporte disponibles que sean de utilidad para ajustar los resultados. La solución preprocesa y limpia los datos de partida, eliminando los errores de la fuente original y filtrando la información que no es necesaria para el análisis de la movilidad. Tras seleccionar una muestra de usuarios con registros suficientes para caracterizar su movilidad, se analizan los datos para iden-

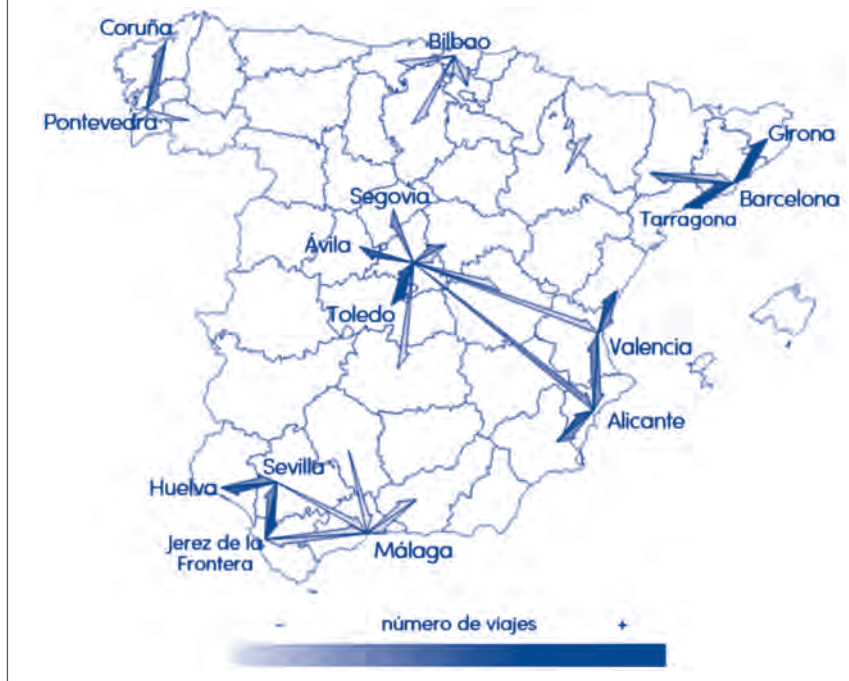
tificar la ubicación y duración de las actividades de los usuarios anónimos. Los desplazamientos entre estas actividades son los viajes, de los cuales es posible extraer más información gracias a los registros generados durante cada viaje. La posibilidad de analizar secuencias de varias semanas para cada usuario permite identificar en qué zona reside, lo que es de utilidad para la expansión de la muestra, según la relación entre población y muestra residente en cada unidad censal. De esta manera, se asigna un factor de expansión a cada cadena de actividades y viajes, lo que permite, finalmente, generar los indicadores de movilidad con la agregación espacial y temporal deseada.

Los resultados han permitido al MITMA caracterizar con precisión los desplazamientos entre provincias y su variabilidad en distintos tipos de día, tanto a lo largo de la semana como a lo largo del año. En particular, se pudieron identificar qué flujos casa-trabajo de larga distancia son los más relevantes en España gracias a las matrices origen-destino de días laborables, tal como se observa en el gráfico 1 (por ejemplo, Girona-Barcelona o Toledo-Madrid). El análisis de estos flujos por modo permite cuantificar la captación del ferrocarril en estas relaciones frente al vehículo privado, lo que es esencial para elaborar estrategias que permitan reducir la huella de carbono de este tipo de desplazamientos. Además, se pudieron identificar las relaciones que más crecen en operaciones salida asociadas a períodos vacacionales, tales como Madrid-Alicante o Vizcaya-Cantabria, como se puede ver en el gráfico 2. Esta información puede servir para ajustar la oferta de transporte

GRÁFICO 1  
RELACIONES PRINCIPALES EN LA MATRIZ ORIGEN-DESTINO DE JUEVES PROMEDIO DE PERÍODO NO ESTIVAL (OCTUBRE DE 2017)



GRÁFICO 2  
RELACIONES PRINCIPALES EN LA MATRIZ ORIGEN-DESTINO DEL VIERNES 14 DE JULIO DE 2017 (OPERACIÓN SALIDA)



público interprovincial de manera más adecuada a la demanda excepcional que se da en períodos pico. Los resultados del estudio han sido publicados como datos abiertos (OTLE, 2020) y están siendo empleados por el MITMA para la construcción del Modelo Nacional de Transportes (MNT), herramienta que permitirá explorar escenarios de políticas públicas de transporte a nivel estatal (MITMA, 2021a).

## 2. La monitorización de la movilidad durante la pandemia de COVID-19

Una de las medidas más extendidas para combatir la pandemia de COVID-19 ha sido la limitación de la movilidad, con el propósito de reducir el contacto social. El MITMA aprovechó su experiencia previa con los datos de telefonía móvil para lanzar un proyecto de monitorización continua de la movilidad a nivel estatal en marzo de 2020, con el

objetivo de comprobar el cumplimiento de las medidas de restricción y vigilar los cambios en los patrones de demanda de transporte derivados de la situación excepcional. Esta información ha sido muy útil para la adaptación de la oferta de transporte público a las necesidades de desplazamiento, así como para el desarrollo de modelos epidemiológicos de difusión de la enfermedad. La capacitación de los técnicos gracias a las experiencias previas con datos de telefonía fue vital para poder promover y gestionar con rapidez un proyecto de estas características.

En este estudio se han obtenido matrices origen-destino y otros indicadores de movilidad (por ejemplo, número de viajes por persona) de manera continua desde marzo de 2020 hasta mayo de 2021, cubriendo todos los días de este período. El MITMA obtuvo esta información con tan solo tres días de decalaje, de manera que podía

disponer en muy poco tiempo de datos actualizados acerca de cómo se comportaba la movilidad en el país. Las matrices origen-destino registraron todos los desplazamientos de más de 500 metros, con una zonificación de tamaño variable (desde agrupaciones de municipios en el ámbito rural hasta nivel de distrito en las ciudades). La información tiene resolución horaria e incluye segmentaciones como el propósito de viaje. Un ejemplo de la información que se puede obtener mediante este estudio es la que se observa en el gráfico 3, donde se muestra la evolución diaria del número de viajes entre las provincias de Madrid y Barcelona desde que irrumpió la pandemia. Se observa cómo los mínimos de movilidad se registraron durante el confinamiento más estricto en marzo y abril de 2020, así como durante el temporal invernal Filomena en enero de 2021. Se observa también el impacto del período estival o la disminución de la movilidad en

GRÁFICO 3  
EVOLUCIÓN DIARIA DEL NÚMERO DE VIAJES ENTRE LAS PROVINCIAS DE MADRID Y BARCELONA (AMBOS SENTIDOS) ENTRE FEBRERO DE 2020 Y FEBRERO DE 2021

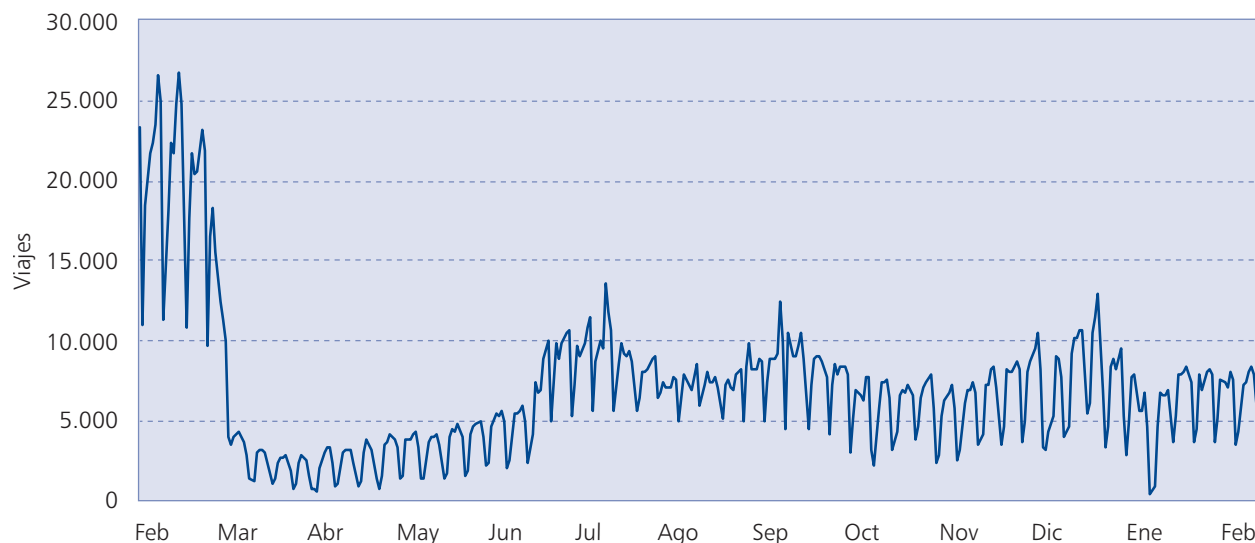
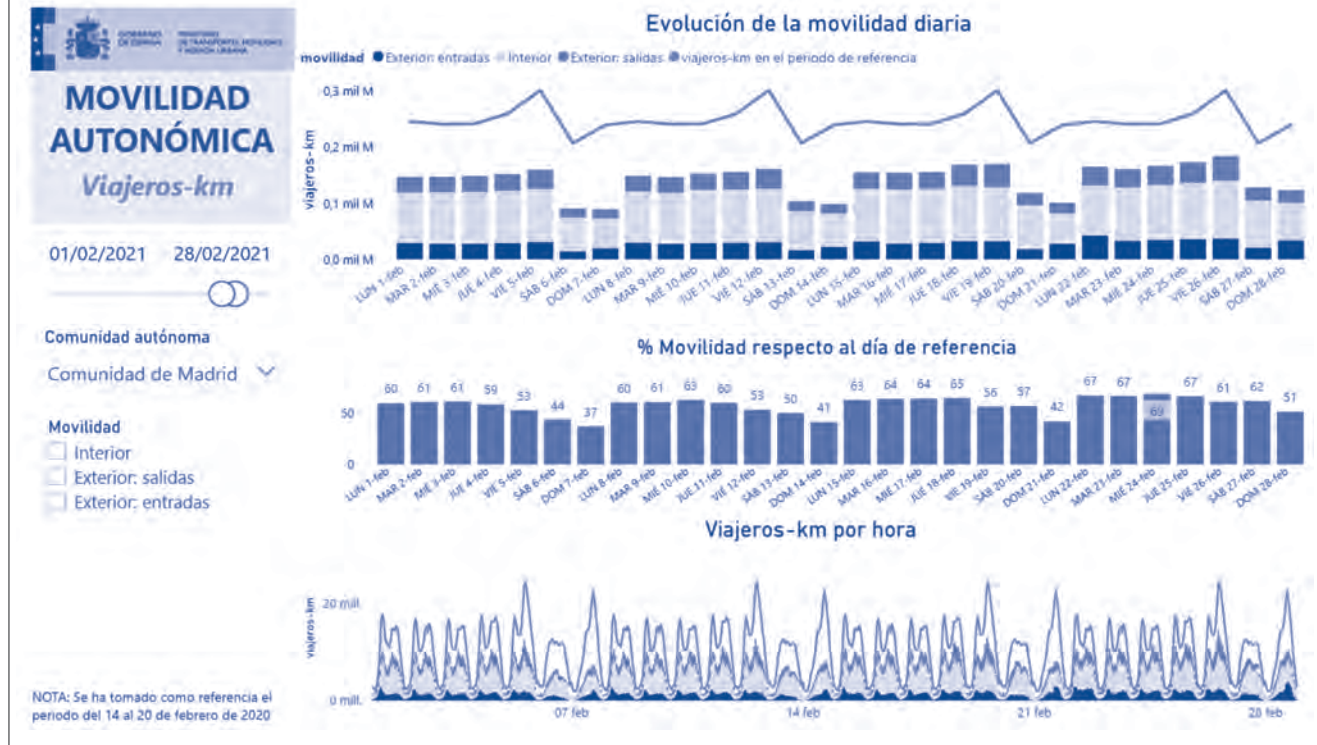


GRÁFICO 4

EJEMPLO DE LA HERRAMIENTA DE VISUALIZACIÓN CREADA POR EL MITMA PARA EXPLOTAR LOS RESULTADOS DEL ANÁLISIS DE MONITORIZACIÓN DE LA MOVILIDAD



octubre y noviembre de 2020 coincidiendo con la segunda ola de la pandemia.

Al igual que en el caso de uso anterior, estas matrices origen-destino y el resto de indicadores fueron obtenidas mediante la solución analítica de Nommon Solutions and Technologies a partir de los registros anonimizados de telefonía móvil procedentes de la red de Orange España (MITMA, 2021b). Dicha solución fue adaptada para incrementar las capacidades de computación y automatizar las validaciones de datos de partida e información de salida, de manera que fuera posible la producción de indicadores de movilidad en un corto período de tiempo desde la recepción de los datos. El MITMA desarrolló un portal de

visualización de la información para poner los resultados a disposición de toda la ciudadanía (gráfico 4), así como un portal de descarga de los datos (MITMA, 2021c).

#### IV. LÍNEAS DE INVESTIGACIÓN EN MARCHA

##### 1. Incorporación de nuevas fuentes de datos

La constante aparición de tecnologías emergentes da lugar a nuevas generaciones de fuentes de datos con gran potencial para aportar distintas perspectivas en el análisis de la movilidad. Algunas de estas tecnologías surgen en el propio ámbito del transporte, como es el caso

del vehículo autónomo y conectado. La progresiva automatización de los servicios de movilidad, no solo del vehículo privado, va a facilitar la recopilación de datos de demanda de transporte (por ejemplo, mediante el análisis detallado de la ocupación del transporte público, dada la necesidad de monitorizar la subida y bajada de viajeros). En otros casos, tecnologías más transversales o que no son propias del sector del transporte son las que aportan el nuevo tipo de dato. Un ejemplo son los datos procedentes de satélites (conocidos como *earth observation*): imágenes de alta resolución que registran con una frecuencia elevada cambios en los usos del suelo. La combinación de esta fuente con los datos procedentes de dispositivos móviles permitirá



el cálculo de nuevos indicadores de actividad socioeconómica y de uso del espacio público.

## 2. Obtención de nuevos indicadores de actividad de la población

La explotación de nuevas fuentes de datos para el análisis de la movilidad se ha centrado habitualmente en la obtención de matrices origen-destino a distintas escalas (por ejemplo, a nivel de ciudad o región mediante datos de telefonía móvil, a nivel de red de carreteras local con sensores wifi o *bluetooth*, etc.). Este foco se debe a que las matrices origen-destino son la manera más habitual de representar la dimensión de la demanda en modelos de transporte, al expresar de una manera sintética los flujos de desplazamientos entre zonas, construyendo una matriz por cada segmentación que se desee. Sin embargo, las nuevas fuentes de datos permiten la construcción de otros indicadores de movilidad y actividad de la población en el territorio que van más allá de las matrices origen-destino. Un ejemplo está constituido por indicadores que miden la existencia de cadenas de viaje triangulares (*tours*), con paradas intermedias para satisfacer algunas actividades (por ejemplo, compras) en los movimientos casa-trabajo. Este tipo de realidades no se reflejan en las matrices origen-destino y, sin embargo, son muy relevantes para entender la competitividad del transporte público frente al vehículo privado. Otro ejemplo está relacionado con indicadores de contacto social, que expresan la concurrencia de personas de distintos segmentos sociodemográficos en ciertos lugares e instantes determinados.

La pandemia de COVID-19 ha incrementado el interés por este tipo de indicadores desde el punto de vista de la epidemiología, pero su aplicación es extensible a otras áreas, como los estudios acerca de la desigualdad y segregación social.

## 3. Del análisis descriptivo al análisis predictivo: aplicación de inteligencia artificial

El estudio realizado por el MITMA durante la pandemia de COVID-19 es un buen ejemplo de las capacidades de las nuevas fuentes de datos para monitorizar la movilidad. No obstante, las oportunidades de la recogida continua de información no se limitan únicamente a una mejor caracterización de la movilidad actual. La construcción de bases de datos de demanda de transporte longitudinales con alto nivel de detalle abre la puerta al desarrollo de modelos predictivos basados en técnicas de inteligencia artificial. Estos modelos se nutren de los patrones cíclicos de demanda asociados al calendario (estacionalidad, diferencias entre días de la semana, etc.), pero también de la incidencia de eventos puntuales, de la meteorología y de otros factores dinámicos en la movilidad. La disponibilidad de estas capacidades predictivas permite un gran número de casos de uso, relevantes tanto para planificadores de transporte como para operadores de servicios de movilidad.

## V. CONCLUSIONES

En los últimos años, el sector de la movilidad ha experimentado un crecimiento en sus capacidades de análisis de la demanda de transporte gracias a la aparición de nuevas fuentes de datos *big*

*data*. Estas fuentes, que habitualmente nacen con un propósito diferente a caracterizar la movilidad, permiten un análisis continuo de los desplazamientos sin intervenir en los mismos. Esto presenta numerosas oportunidades frente a las fuentes de datos tradicionales, tales como un mayor tamaño de muestra, una mayor frecuencia de actualización o menores sesgos en la caracterización de la demanda de transporte. Si bien no están exentas de retos, entre los que se encuentran asegurar la privacidad de los datos o el desarrollo de algoritmos capaces de eliminar los errores de partida, las fuentes de datos *big data* se han convertido ya en un instrumento esencial para elaborar diagnósticos de la movilidad más completos y precisos que en el pasado. Esto permite una mejor toma de decisiones tanto a planificadores como a operadores de transporte, tal como demuestran los dos ejemplos de proyectos elaborados por el Ministerio de Transportes, Movilidad y Agenda Urbana de España basados en datos de telefonía móvil que presenta este artículo.

A pesar de que estas fuentes suponen un gran salto en capacidad analítica, la experiencia sugiere que su aplicación no está necesariamente dirigida a sustituir por completo las fuentes de datos tradicionales, sino a complementar y optimizar su aplicación. La complejidad del fenómeno de la movilidad hace que sea difícil abarcar todas sus dimensiones mediante un único método de investigación. En muchos casos la combinación de fuentes pasivas y encuestas de movilidad es un enfoque óptimo para adquirir una imagen lo más completa posible de la demanda de transporte.

Las nuevas fuentes de datos ya son parte de la caja de herramientas de planificadores y operadores de movilidad, lo que convierte a España en uno de los países más avanzados en la materia. Sin embargo, existen numerosas líneas de investigación en activo que continuarán mejorando las capacidades de análisis gracias a estas fuentes, incorporando nuevas tecnologías y desarrollando nuevos indicadores y métodos de modelado de transporte a partir de la recogida continua de datos.

#### BIBLIOGRAFÍA

- AYFANTOPOULOU, G., BURRIEZA-GALÁN, J., MASEGOSA, A., GRAU, J. M. S., BOUFIDIS, N., MARTÍNEZ, I. M., ... y ROS, O. C. (2020). Cataloging and Assessing City-scale Mobility Data. En *Conference on Sustainable Urban Mobility*, pp. 1139-1148. Springer.
- CHEN, C., MA, J., SUSILO, Y., LIU, Y. y WANG, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research*

*Part C: Emerging Technologies*, 68, pp. 285-299.

- LANEY, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research note*, 6(70), p. 1.
- LENORMAND, M., LOUAIL, T., CANTÚ-ROS, O. G., PICORNELL, M., HERRANZ, R., ARIAS, J. M., BARTHELEMY, M., SAN MIGUEL, M. y RAMASCO, J. J. (2015). Influence of sociodemographic characteristics on human mobility. *Scientific Reports*, 5(1), pp. 1-15.
- MILNE, D. y WATLING, D. (2019). Big data and understanding change in the context of planning transport systems. *Journal of Transport Geography*, 76, pp. 235-244.
- MITMA (2019). *Estudio de la movilidad interprovincial de viajeros aplicando la tecnología big data. Informe metodológico*. [https://cdn.fomento.gob.es/portal-web-drupal/Docs\\_OTLE/MFOM-Estudio\\_Movilidad\\_Interprovincial\\_Informe\\_Metodologico.pdf](https://cdn.fomento.gob.es/portal-web-drupal/Docs_OTLE/MFOM-Estudio_Movilidad_Interprovincial_Informe_Metodologico.pdf)
- MITMA (2021a). *Modelo Nacional de Transportes de España. Diseño general del Modelo Nacional de Transportes*. <https://cdn.mitma.gob.es/portal-web-drupal/hermes/>

[documento\\_diseno\\_general\\_del\\_mnt.pdf](#)

- MITMA (2021b). *Análisis de la movilidad en España con tecnología big data durante el estado de alarma para la gestión de la crisis de la COVID-19. Informe metodológico*. [https://cdn.mitma.gob.es/portal-web-drupal/covid-19/bigdata/mitma\\_-\\_estudio\\_movilidad\\_covid-19\\_informe\\_metodologico\\_v3.pdf](https://cdn.mitma.gob.es/portal-web-drupal/covid-19/bigdata/mitma_-_estudio_movilidad_covid-19_informe_metodologico_v3.pdf)
- MITMA (2021c). *Open Data Movilidad*. [www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad](http://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad)
- OTLE (2020). *Estudio piloto de movilidad interprovincial*. <https://observatoriotransporte.mitma.es/estudio-experimental>
- TOOLE, J. L., COLAK, S., STURT, B., ALEXANDER, L. P., EVSUKOFF, A., y GONZÁLEZ, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58, pp. 162-177.
- ZHANG, R., YE, X., WANG, K., LI, D., y ZHU, J. (2019). Development of commute mode choice model by integrating actively and passively collected travel data. *Sustainability*, 11(10), p. 2730.