

# CÓMO APLICAR BIG DATA EN LA PLANIFICACIÓN DEL TRANSPORTE:

## EL USO DE DATOS DE GPS EN EL ANÁLISIS DE LA MOVILIDAD URBANA

División de Transporte

NOTA TÉCNICA N°  
**IDB-TN-1944**

### **Autores:**

Javier Gutiérrez Puebla  
Carolina Benítez  
Juan Carlos García Palomares  
Gustavo Romanillos Arroyo  
Elías Rubinstein  
Juan Manuel Leña  
Karisa Maia Ribeiro  
Lynn Scholl  
Borja Moya Gómez  
Ana Condeço Melhorado

### **Editora:**

Carolina Benítez

Julio 2020

# CÓMO APLICAR BIG DATA EN LA PLANIFICACIÓN DEL TRANSPORTE:

## EL USO DE DATOS DE GPS EN EL ANÁLISIS DE LA MOVILIDAD URBANA

### **Autores:**

Javier Gutiérrez Puebla  
Carolina Benítez  
Juan Carlos García Palomares  
Gustavo Romanillos Arroyo  
Elías Rubinstein  
Juan Manuel Leñaño  
Karisa Maia Ribeiro  
Lynn Scholl  
Borja Moya Gómez  
Ana Condeço Melhorado

### **Editora:**

Carolina Benítez

Julio 2020

**Catalogación en la fuente proporcionada por la  
Biblioteca Felipe Herrera del  
Banco Interamericano de Desarrollo**

Cómo aplicar big data en la planificación del transporte: el uso de datos de GPS en el análisis de la movilidad urbana / Javier Gutiérrez Puebla, Carolina Benítez, Juan Carlos García Palomares, Gustavo Romanillos Arroyo, Elías Rubinstein, Juan Manuel Leaño, Karisa Maia Ribeiro, Lynn Scholl, Borja Moya Gómez, Ana Condeço Melhorado.

p. cm. — [Nota técnica del BID ; 1944]

Incluye referencias bibliográficas.

1. Urban transportation-Data processing. 2. Big data. 3. Urban Transportation-Planning. 4. Global Positioning System. I. Gutiérrez Puebla, Javier. II. Benítez, Carolina. III. García Palomares, Juan Carlos. IV. Romanillos Arroyo, Gustavo. V. Rubinstein, Elías. VI. Leaño, Juan Manuel. VII. Ribeiro, Karisa Maia. VIII. Scholl, Lynn. IX. Moya Gómez, Borja. X. Condeço-Melhorado, Ana. XI. Banco Interamericano de Desarrollo. División de Transporte. XII. Serie.

IDB-TN-1944

Diseño y Diagramación: **FILEZ**

Edición de estilo: **Ximena Abeledo**

Contacto: **Carolina Benítez (cbenitez@iadb.org)**

<http://www.iadb.org>

Copyright © 2020 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



## Resumen

La producción de grandes volúmenes de datos masivos, Big Data, abre interesantes posibilidades para entender los flujos de movilidad de nuestras ciudades de la región, dada su complementariedad con los métodos tradicionales de recolección de datos. La presente nota técnica tiene como objetivo presentar las oportunidades que el uso de Big Data a partir de registros GPS ofrecen a la hora de entender y planificar la movilidad urbana. Aportará información sobre las fuentes de datos de las que se puede obtener información GPS, las características específicas de este tipo de datos, las metodologías para derivar de los datos información útil en la planificación del transporte, principales ventajas y limitaciones.

Incluye, asimismo, estudios de caso que permitirán ilustrar la multiplicidad de posibles aplicaciones y encontrar en un mismo documento una agrupación y sistematización sobre los conocimientos generados por técnicos e investigadores en los últimos años.

**Clasificaciones JEL:** O2, O3, O18, O32, O35, Z18

**Palabras clave:** Transporte urbano, procesamiento de datos, Big Data planeamiento del transporte urbano, sistemas de geoposicionamiento.





# Cómo aplicar Big Data en la planificación del transporte:

## El uso de datos de GPS en el análisis de la movilidad urbana

# Contenidos

<b>RESUMEN EJECUTIVO</b>	04
<b>1. INTRODUCCIÓN: BIG DATA, REGISTROS GPS Y MOVILIDAD INTELIGENTE</b>	05
1.1 El contexto de la cuarta revolución industrial y las Smart Cities	06
1.2 La producción de datos masivos: Big Data geolocalizado	07
1.3 Big Data, GPS y movilidad	11
1.4 Objetivos y estructura de la nota técnica	14
<b>2. FUENTES DE DATOS Y CAMPOS DE APLICACIÓN</b>	17
2.1 De las fuentes de datos tradicionales a las nuevas fuentes de datos	18
2.2 Datos obtenidos a partir de localizadores GPS	19
2.3 Nuevas fuentes de datos basadas en registros GPS	21
2.4 Campos de aplicación	34
<b>3. CARACTERÍSTICAS DE LOS REGISTROS GPS</b>	36
3.1 Registros GPS: Tipos de ficheros	37
3.2 Información contenida en los registros GPS	39
3.3 Resolución espacial de los datos: exactitud posicional	41
3.4 Resolución temporal de los datos	42
3.5 Limitaciones de los estudios basados en datos GPS	43
<b>4. ANÁLISIS DE MOVILIDAD A PARTIR DE DATOS DE LOCALIZACIÓN PUNTUALES</b>	46
4.1 Metodología a seguir para el cálculo de Matrices OD de viajes a partir de datos de localización puntual	48
4.2 Análisis de patrones de movilidad de los turistas utilizando localizaciones puntuales de fotografías geolocalizadas	54
<b>5. ANÁLISIS DE MOVILIDAD A PARTIR DE TRACKS (RECORRIDOS)</b>	58
5.1 Obtención de datos de recorridos GPS: fuentes y usos potenciales	59
5.2 Metodología para el tratamiento y enriquecimiento de los datos	64
5.3 Análisis de patrones de movilidad de personas a partir de tracks GPS	68
5.4 Monitorización y análisis de la movilidad en la ciudad y el territorio a partir de tracks GPS	72
5.5 Modelización de la movilidad: simulaciones para escenarios futuros en cuanto a la implementación de políticas e infraestructuras	77

<b>6. EXPERIENCIAS INTERNACIONALES .....</b>	<b>82</b>
6.1 Visualización y análisis de rutas a partir de datos de tracks de bicicletas públicas .....	85
6.2 Estimación de la demanda de servicios de taxi .....	93
6.3 Evaluación de los impactos de la congestión metropolitana en la accesibilidad usando datos de TomTom Speed Profiles .....	98
6.4 Uso de datos de tarjeta inteligente y GPS: el caso del Transantiago .....	104
6.5 Análisis de los accidentes de tráfico a partir de la integración de datos oficiales y de Waze .....	108
<b>7. CONSIDERACIONES FINALES .....</b>	<b>114</b>
<b>BIBLIOGRAFÍA .....</b>	<b>119</b>
<b>ANEXO: DESCARGA DE DATOS DE LA API DE TWITTER .....</b>	<b>123</b>

## Resumen Ejecutivo

*Big Data* es un término que se ha popularizado en los últimos años y hace referencia a la producción de grandes cantidades de datos masivos obtenidos a través de múltiples redes de sensores, dispositivos y de las actividades de los usuarios en Internet. Estos datos nos aportan información sobre multitud de procesos, que requieren herramientas específicas para su tratamiento. Cabe citar algunos ejemplos tales como, el crecimiento de las ciudades (sensores remotos), la contaminación del aire (sensores localizados en las ciudades) hasta los patrones de movilidad de la población (teléfonos móviles) o su consumo (tarjetas bancarias).

Una de las fuentes de datos masivos más utilizadas en la planificación del transporte es la constituida por los registros de sistemas de posicionamiento global, conocido por sus siglas en inglés, GPS. Los registros GPS ofrecen información muy útil y variada. La misma, se genera de forma constante, con una alta resolución espacial y temporal, lo que permite monitorizar y analizar diferentes procesos de la movilidad urbana. Los datos de dispositivos GPS permiten analizar temáticas tales como; las velocidades de circulación, la distribución de los viajes, la carga de la red, los problemas de agrupamientos de autobuses, los accidentes de tráfico, etc. El flujo continuo de datos hace posible estudiar estas características de la movilidad, tanto en situaciones habituales como en días atípicos o el impacto de eventos especiales. A diferencia de otras metodologías de obtención de datos, en los que existen serias dificultades para identificar el modo de transporte, en el caso de los registros GPS la identificación del modo está asegurada cuando se trata de datos provenientes de dispositivos instalados en los vehículos: autobuses, vehículo privado, bicicletas, patinetes eléctricos, etc.

En noviembre de 2019, se ha publicado un documento sobre *“Cómo aplicar big data en la planificación del transporte urbano. El uso de datos de telefonía móvil”*<sup>1</sup>. En la misma línea, la presente nota técnica surge con el objetivo de presentar las oportunidades que el uso de *Big Data* a través de registros GPS brindan a la planificación de la movilidad urbana y en un intento de contribuir también

al objetivo de mejorar la capacitación de los cuadros técnicos de las ciudades de la región para aprovechar las ventajas de las nuevas fuentes de datos en la planificación del transporte. A lo largo del documento, se podrá visualizar una agrupación y sistematización sobre los conocimientos generados por técnicos e investigadores en los últimos años en distintos países del mundo, orientando a aquellos planificadores de transporte y técnicos en datos que quieren adentrarse en este campo. Asimismo, se describen las características específicas de este tipo de datos y se presentan las metodologías para obtener datos con información útil en transporte, mostrando sus principales ventajas y también sus limitaciones, basados en estudios de casos que permitirán ilustrar la multiplicidad de posibles aplicaciones.

En la región de América Latina y el Caribe, existen numerosas ciudades que hacen uso de datos de GPS en el estudio de la movilidad urbana y la planificación del transporte. Un ejemplo claro es el uso de tarjetas de transporte o de aplicaciones tales como Waze. La ventaja para la región es que las metodologías para el tratamiento de datos de registros de GPS, aplicados al sector, se encuentran bastante estudiadas. Existe además una penetración muy alta de muchas de estas fuentes en la región y, a diferencia de otro tipo de datos (como los de telefonía móvil), el acceso a muchas de estas fuentes es relativamente fácil. Por lo tanto, aunque se evidencian distintas aplicaciones exitosas en varias ciudades, su adopción generalizada en el resto de las ciudades presenta una gran oportunidad.

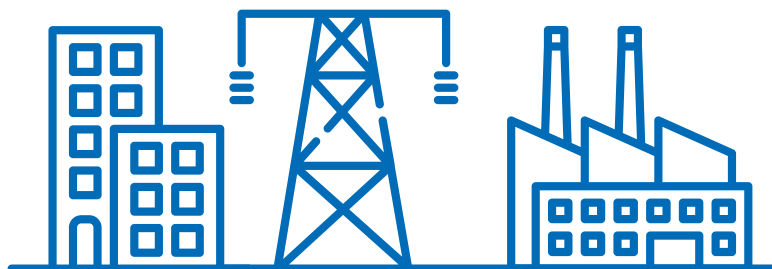
La movilidad del futuro requerirá y exigirá cuadros técnicos multidisciplinarios y especializados en el manejo de numerosas fuentes de datos, incluidas las que aquí se presentan. Para ello, en el corto y mediano plazo, los cuadros técnicos expertos en movilidad necesitarán entender las características de los nuevos datos y herramientas de análisis a fin de poder trabajar en coordinación con los científicos de datos y obtener el mayor beneficio posible de las aportaciones de la ciencia de datos a la planificación del transporte.

<sup>1</sup><https://publications.iadb.org/es/como-aplicar-big-data-en-la-planificacion-del-transporte-urbano-el-uso-de-datos-de-telefoniamovil>



# 1

**Introducción:  
Big Data, registros GPS  
y movilidad inteligente**



## 1.1 El contexto de la cuarta revolución industrial y las *Smart Cities*

No cabe duda de que estamos presenciando una era de profundos cambios económicos y sociales impulsados por el avance de la tecnología. La revolución tecnológica está transformando nuestra forma de vivir, trabajar y relacionarnos. Vivimos en una sociedad hiperconectada. Por primera vez en la historia, miles de millones de personas están continuamente conectadas entre sí, a través de sus dispositivos móviles. Este evento, deja una huella digital, generando una ingente cantidad de datos.

Esta revolución tecnológica, conocida como la cuarta revolución industrial, provoca una confluencia continua de nuevas tecnologías, tales como la inteligencia artificial, el internet de las cosas, la robotización, el vehículo autónomo o la nanotecnología. Muchas de estas innovaciones tecnológicas están solo en sus comienzos, pero están llegando a un punto de inflexión a medida que se desarrollan y amplifican mutuamente en una fusión de tecnologías a través de los mundos físico, digital y biológico (Schwab, 2018).

El soporte de esta revolución es la tecnología de la información y las telecomunicaciones. Los nuevos desarrollos tecnológicos se basan en la creciente capacidad para producir, intercambiar y procesar datos. Miles de millones de sensores y dispositivos están conectados entre sí (internet de las cosas) produciendo una gran cantidad de datos (Big Data) que son almacenados y procesados en la nube mediante nuevas herramientas de analítica de datos e inteligencia artificial. Para la transferencia de datos masivos se necesitan redes de telecomunicaciones con mayor capacidad y rapidez, lo que será posible gracias al despliegue del 5G. El 5G es un requisito para el desarrollo del internet de las cosas y para la implantación de tecnologías que exigen un flujo rápido y fiable de datos, como es la del vehículo autónomo, que, a su vez, necesita de soluciones basadas en la inteligencia artificial para tomar decisiones ante cualquier incidencia. Al final, unas y otras tecnologías se potencian dando lugar a nuevas tecnologías. La cuarta revolución industrial está caracterizada no tanto por el desarrollo de tecnologías específicas sino por la convergencia de tecnologías.

Otros rasgos distintivos de esta revolución son la gran rapidez con la que se están produciendo los cambios y su carácter transversal, afectando a casi todas las facetas de la economía y la sociedad. En este contexto se sitúa el concepto de *Smart Cities*, ciudades inteligentes en las que la tecnología se pone al servicio

de la prestación de servicios, la calidad de vida, la participación ciudadana, el medio ambiente, la economía y la movilidad urbana. Las tecnologías de la información y las comunicaciones actúan como el nodo de enlace entre el mundo virtual, donde se halla la información, y el mundo físico, donde se proyecta dicha información espacial y temporalmente (Figura 1). Se puede decir que los datos son impulsos neuronales de un sistema nervioso digital que recubre poco a poco el mundo físico con sentidos y músculos, evolucionando las ciudades a un entorno cada vez más inteligente (Tascón y Coullaut, 2016). El Big Data enriquece el conocimiento acerca de las ciudades y ofrece nuevas oportunidades para la interacción social y la planificación (Batty, 2013).



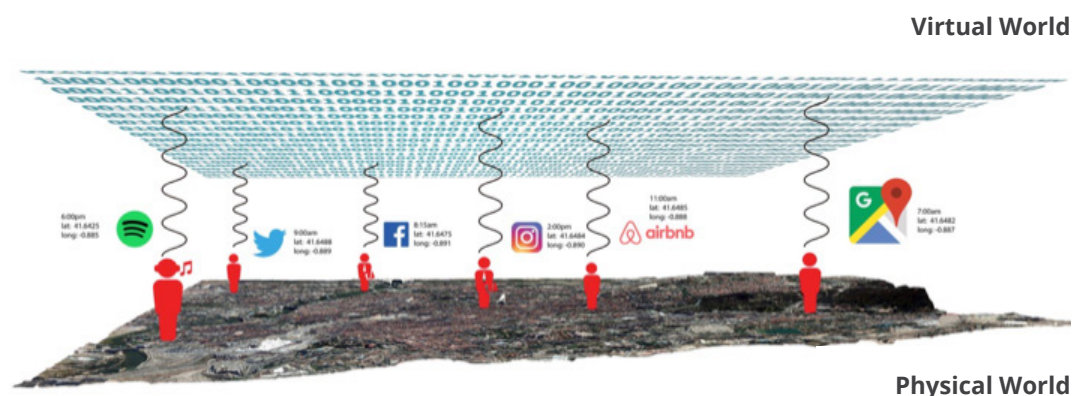
## 1.2 La producción de datos masivos: Big Data geolocalizado

El término *Big Data* se ha popularizado enormemente en los últimos años y hace referencia a la producción de grandes cantidades de datos por medio de múltiples redes de sensores y dispositivos. Gran parte de estos datos se generan de forma automática por distintos tipos de sensores. Así, por ejemplo, generamos cantidades ingentes de datos geolocalizados cuando nos desplazamos (registrados a través del GPS de nuestros teléfonos inteligentes), cuando utilizamos nuestro teléfono móvil, cuando pagamos con tarjetas de crédito o presentamos una tarjeta de fidelización, o cuando hacemos uso de nuestra tarjeta inteligente de transporte (Gutiérrez Puebla et al., 2016). Todos esos datos se capturan y almacenan de forma automática.

También se genera una enorme cantidad de datos a partir del uso de internet. Este tipo de datos se conoce como contenidos generados por los usuarios. En el mundo de la web 3.0, los usuarios de internet ya no adoptan una actitud pasiva, de meros receptores de información, sino que se convierten en productores de grandes cantidades de datos, particularmente a través de las redes sociales (Gutiérrez Puebla et al., 2016), en las que se almacenan datos cuando las utilizamos, ya se trate de redes generalistas, como Twitter, o especializadas, como las redes sociales de fotografías compartidas (Flickr, Instagram) (Figura 1).

Muchos de estos datos, tanto los capturados por sensores y dispositivos como los generados por los usuarios en internet, son de gran utilidad, al reflejar los patrones de movilidad de los individuos, ya sea en todos los modos (por ejemplo, datos de telefonía móvil) o en alguno de ellos, (por ejemplo, la tarjeta de transporte inteligente o las tarjetas para uso de bicicletas públicas).





**Figura 1:** La huella digital de los humanos y su localización en el espacio y en el tiempo. La secuencia temporal de la geolocalización de un individuo permite seguir sus desplazamientos en la ciudad. Fuente: BID, 2019.

Las nuevas fuentes de datos masivos (*Big Data*) presentan rasgos distintivos frente a las fuentes de datos tradicionales. Para caracterizar el *Big Data*, generalmente, se recurre al esquema de las 3 vs (volumen, velocidad y variedad); a veces se añaden algunas características adicionales hasta alcanzar las 5 vs (veracidad y valor) o incluso las 7 vs (variabilidad y visualización). Aquí se enumeran las principales características del *Big Data*, conocidas como las 5 vs (Gutiérrez Puebla, 2018):

**VOLUMEN:** el *Big Data* se caracteriza por su volumen, su carácter masivo. Según la compañía tecnológica Oracle, los datos producidos en todo el mundo crecen a un ritmo del 40% anual. Las unidades de medida para estas ingentes cantidades de datos son los terabytes (un billón de bytes), petabytes (mil billones de bytes), exabytes (un millón de billones de bytes), etc. Twitter, por ejemplo, almacena medio millón de tweets al minuto, lo que al cabo de un año supone un enorme volumen de datos. A diferencia de lo que ocurría en el pasado, los datos históricos se conservan (no se borran) porque sirven de apoyo para analizar procesos en el pasado y realizar predicciones futuras.

**VELOCIDAD:** los datos son generados, almacenados y procesados a gran velocidad. Se generan de forma continua, de manera que es posible seguir procesos del mundo real en streaming y hacer análisis de datos en tiempo casi real. La API (*Application Programming Interface*) de Twitter, por ejemplo, permite descargar tweets geolocalizados en tiempo real, con lo que se pueden monitorizar concentraciones de población por eventos especiales. Frente a las estadísticas oficiales, que ofrecen informaciones de un proceso en un momento determinado (una foto fija), como ocurre por ejemplo con el censo, la velocidad asociada al *Big Data* aporta datos con alta resolución temporal, lo que permite seguir procesos de forma continua.

**VARIEDAD:** la variedad hace referencia a la diversidad de tipos, formatos y fuentes de datos, desde datos estructurados (que pueden ser presentados en forma de tabla y manejados con sistemas gestores de bases de datos SQL) a datos semiestructurados (como los ficheros HTML o JSON) y datos no estructurados (como archivos de texto, correos electrónicos, imágenes o vídeos, que pueden ser tratados con gestores NoSQL). Datos de distintas fuentes y características pueden ser combinados entre sí y con información de



fuentes tradicionales para producir una información más completa.

**VERACIDAD:** la veracidad hace referencia a la fiabilidad de los datos. El hecho de que los datos sean masivos, (en muchos casos se puede trabajar con poblaciones en lugar de con muestras, que pueden presentar problemas de representatividad y sesgo asociado al análisis de la misma) y de que deriven de comportamientos de los usuarios de los servicios de transporte, son atributos de la veracidad de los datos, aunque no implica que sean completamente fiables<sup>2</sup>.

**VALOR:** en la actual sociedad tecnológica la disponibilidad de datos cobra un valor creciente, hasta el punto que se afirma, con frecuencia, que los datos son el petróleo de la cuarta revolución industrial. Pero el dato en sí mismo no tiene valor. Lo que le da valor es convertirlo en información para generar conocimiento útil para la acción y para la toma de decisiones. En este sentido, la analítica de datos y la inteligencia artificial aportan gran valor a los datos, tanto para descubrir patrones y estudiar procesos como para realizar predicciones. La mayor parte de los datos masivos están localizados en el tiempo y en el espacio, es decir, contienen las coordenadas temporales del momento en que se generó el dato (año, mes, día, hora, minuto y segundo) y también las coordenadas espaciales del lugar (localización por latitud y longitud u otros sistemas de geolocalización). En general, la resolución temporal de los datos es bastante alta (por ejemplo, un registro cada hora como promedio para los usuarios de telefonía móvil) en comparación con las fuentes de datos tradicionales. También es, generalmente, alta la resolución espacial de la mayor parte de los registros provenientes de nuevas fuentes de datos. Cuando se trata de sensores o dispositivos fijos (por ejemplo, espiras magnéticas<sup>3</sup> o cámaras de video) su geolocalización es muy precisa. Cuando los dispositivos son móviles (portados por personas o instalados en vehículos), generalmente, la geolocalización se hace por GPS, lo que conlleva un error posicional de unos 2 o 3 metros o incluso de solo centímetros si se utiliza GPS diferencial<sup>4</sup>. En el caso de los datos de telefonía móvil, la localización espacial es menos precisa, ya que se hace a partir de las antenas y sus áreas de influencia, lo que supone un error de algunos cientos de metros en el interior de las ciudades y de algunos miles de metros en las áreas rurales, en las que la densidad de antenas es menor (BID, 2019).

Los datos masivos que almacenan información de individuos (telefonía móvil, redes sociales, tarjetas de transporte, etc.) contienen además un campo en el que se almacena el identificador de usuario. De esta forma, se puede seguir la localización espacial de un mismo usuario a lo largo del tiempo y por lo tanto analizar sus patrones de movilidad (Figura 2). Cuanto mayor sea la resolución espacial y temporal, más precisa será la información obtenida. Los datos de telefonía móvil tienen una alta resolución espacial y permiten

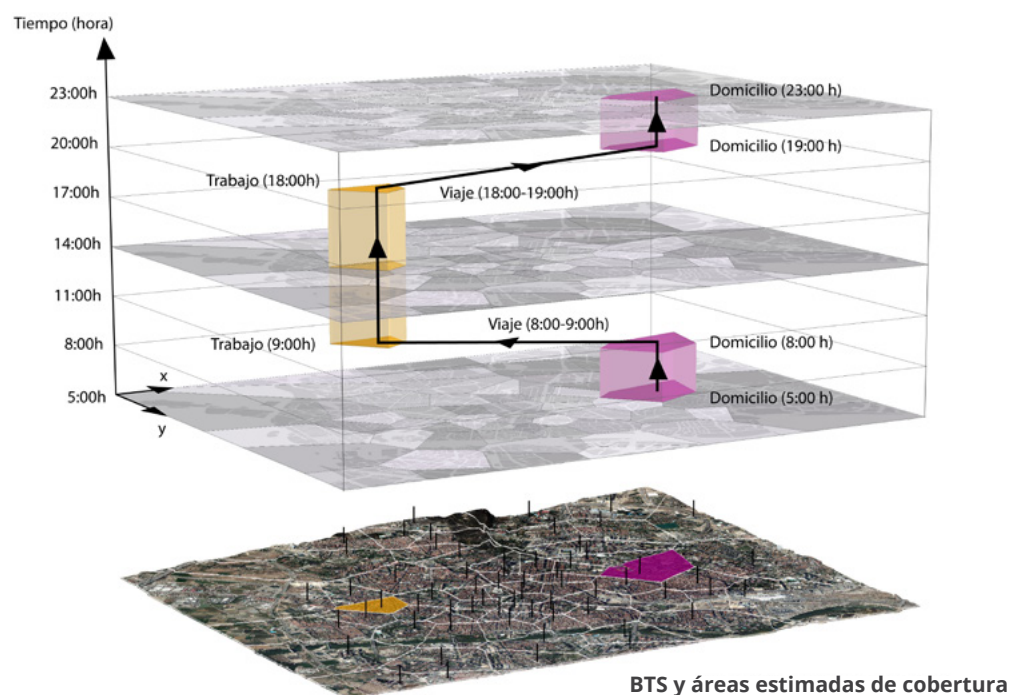
<sup>2</sup> En algunas fuentes de datos se disponen de muestras (generalmente de gran tamaño), como ocurre con los datos de telefonía o de redes sociales, en otras, en cambio; se dispone de la totalidad de los datos, no de muestras (uso de la tarjeta inteligente de transporte en las ciudades en las que su uso es obligatorio en los transportes urbanos).

<sup>3</sup> Existen dos tipos de sensores para la detección y monitorización de vehículos: (i) las tecnologías intrusivas, que están instaladas en o a lo largo del pavimento (las espiras magnéticas, son sensores que pertenecen a esta categoría); (ii) las tecnologías no intrusivas que se encuentran por encima o a los lados de la carretera causando mínimo efecto sobre el flujo de tráfico.

<sup>4</sup> Se denomina GPS diferencial a un sistema que brinda a los receptores de GPS correcciones de los datos que se recibe de los satélites GPS, con el objetivo de brindar una mayor precisión y exactitud de la posición que se pretende calcular.

captar con detalle la relación de viajes realizados por una misma persona a lo largo de un día, pero su granularidad<sup>5</sup> espacial es media-alta, de forma que no se conoce con exactitud la localización de los orígenes y los destinos de los viajes. Los datos geolocalizados de redes sociales (como Twitter) tienen una alta granularidad espacial (ya que la geolocalización se hace por GPS), pero su resolución temporal suele ser bastante más baja que en el caso de los datos de telefonía móvil, por lo que, en general, solo reflejan una parte de los viajes en el estudio de la movilidad diaria. Sin embargo, aportan información de gran interés en el análisis de la movilidad interurbana. Las aplicaciones instaladas en *smartphones* (como Google Maps) o los localizadores instalados en vehículos (como en autobuses urbanos o en bicicletas de servicios de movilidad compartida) aportan información detallada sobre los viajes por alta resolución espacial (dispositivos GPS) y temporal. Se trata de lo que se conoce como *tracks*, es decir, secuencias de puntos con alta granularidad espacio-temporal.

#### Trayectorias en prisma espacio-temporal



**Figura 2:** Trayectorias espacio-temporales. El eje vertical representa el tiempo (horas del día). El plano horizontal representa el espacio. Las líneas verticales indican estancias (permanencias en un mismo lugar asociadas a la realización de actividades). Las líneas diagonales se corresponden con viajes (desplazamientos entre lugares conllevando un cierto consumo de tiempo). (Fuente: BID: Guía metodológica para el uso de Big Data a partir de datos de telefonía móvil para entender los patrones de movilidad en ciudades de América Latina y El Caribe)

<sup>5</sup> Se refiere al grado de especificidad a la que se define un nivel de detalle de los datos en una base.

El uso de este tipo de datos tiene un gran potencial en el campo de la movilidad, con multitud de aplicaciones, pero en la actualidad estas fuentes de datos están poco aprovechadas. El *Big Data* aporta información de gran valor, complementaria a la de las fuentes usadas tradicionalmente en los estudios de transporte y movilidad (BID, 2019).



### 1.3 Big Data, GPS y movilidad

La movilidad es uno de los grandes retos de las ciudades del siglo XXI. Existe una creciente demanda de movilidad que se traduce en el aumento del número y la longitud de los viajes, así como una mayor diversidad de motivos para realizarlos, y un uso intensivo de medios motorizados. Es necesario avanzar en modelos de planificación y gestión sostenibles, que garanticen el funcionamiento del sistema económico, promuevan la equidad social y sean respetuosos con el medio ambiente. Se busca un uso eficiente y equilibrado del transporte, que no consuma materia y energía de manera excesiva y que facilite la reducción de la contaminación mediante un mayor uso de los modos activos (desplazamientos peatonales o en bicicleta), y una planificación más inteligente del transporte público (Banister, 2008).

La planificación y gestión de la movilidad sostenible, requiere datos que permitan un conocimiento profundo de las pautas de movilidad de la población, pues de ellos depende la elaboración de diagnósticos correctos y la utilización de las herramientas de modelización y la estimación de los impactos de las actuaciones a promover (Miralles-Guasch, 2012). Sin embargo, los gestores y planificadores de las ciudades, habitualmente, se enfrentan a una situación de escasez de datos. La principal fuente de datos para conocer las pautas de movilidad en las ciudades son las Encuestas Domiciliarias de Movilidad (EDM). Aportan datos de gran interés, pues proporcionan información detallada y específica de la movilidad, pero a la vez tienen costes elevados, lo que reduce la temporalidad (frecuentemente se hacen cada diez años) y el tamaño de las muestras utilizadas; con lo que los datos suelen estar desactualizados y no reúnen el suficiente nivel de detalle espacial para ciertos tipos de análisis. Es necesario disponer de otras fuentes de datos, que proporcionen información complementaria y con una escala espacial y temporal de mayor detalle. Todo ello, a ser posible, con costes económicos menores (Osorio Arjona y García-Palomares, 2017).

En este contexto, la producción de datos masivos (*Big Data*) abre interesantes posibilidades en el campo de la gestión de la movilidad y la planificación del transporte urbano. Estos datos permiten disponer de información siempre actualizada, monitorizar procesos y analizar los fenómenos de forma continua. El *Big Data* puede contribuir a actualizar la información de las encuestas de movilidad o simplemente a aportar información en las ciudades que no dispongan de encuestas, particularmente en lo que se refiere a las matrices de viajes, lo que resulta crítico en la planificación del transporte y la gestión de la movilidad urbana.

Una gran parte de los datos que se generan en la actualidad en las ciudades, están geolocalizados mediante algún sistema global de navegación por satélite

(*Global Navigation Satellite System, GNSS*). Los GNSS son constelaciones de satélites coordinados para posicionar dispositivos en cualquier parte del globo terrestre. El sistema GPS (*Global Positioning System*) es el GNSS de los Estados Unidos de América, el más extendido a nivel global y, claramente, el más relevante en el contexto de esta nota técnica. Otro sistema es el GLONASS de la Federación Rusa, y con distintos niveles de desarrollo, también se deben citar el BEIDOU de China, el IRNSS de India, el QZSS de Japón y el GALILEO, de la Unión Europea. Junto con GPS, el más relevante en el contexto de esta nota técnica es GALILEO, dado el lanzamiento del Centro de Información para América Latina GALILEO, por parte de la Unión Europea, que tiene como objetivo desarrollar la implementación y el uso de GALILEO en América Latina, en colaboración con el sistema GPS. De aquí en adelante, dado el papel tan menor que juegan los restantes GNSS, nos referiremos esencialmente al sistema GPS.

La producción masiva de los datos geolocalizados por GPS abre un nuevo panorama en el estudio de la movilidad urbana en las ciudades de América Latina. Los dispositivos GPS instalados en los teléfonos móviles inteligentes y en los vehículos registran, con gran precisión, la localización de los individuos y los vehículos en el espacio y en el tiempo, suministrando información de gran calidad sobre las características de los viajes. Con respecto a lo primero, muchas aplicaciones de nuestros teléfonos móviles registran nuestra localización y generan registros sobre nuestros viajes, como es el caso de Google Maps o Waze. Los dispositivos de numerosos vehículos también generan registros GPS, como ocurre con los sistemas de bicicletas públicas y de gestión de flotas de transporte público (autobuses) o de transporte de carga (dispositivos de localización instalados en vehículos y contenedores). Estos datos permiten conocer la localización de los vehículos y contenedores en tiempo real y analizar los viajes realizados. Además, algunas encuestas de movilidad ya no solo se basan en los cuestionarios tradicionales, sino que además incluyen el almacenamiento de registros GPS por parte de grupos de encuestados que aceptan ser rastreados o "*trackeados*" durante un determinado periodo de tiempo, aportando datos mucho más precisos.

Con la creciente incorporación de la tecnología, a los servicios de transporte y la puesta en marcha de nuevos servicios de movilidad compartida, la cantidad de datos de registros GPS generados se está incrementando de forma exponencial. Los nuevos servicios de movilidad compartida (bicicletas, motos eléctricas, coches eléctricos, patinetes eléctricos, etc.) se están extendiendo a gran velocidad por numerosas ciudades (Figura 3). De hecho, varios fabricantes de automóviles afirman que en el futuro el foco principal de su negocio irá pivotando desde la producción hasta la provisión de servicios de movilidad. El usuario consulta una aplicación en su teléfono inteligente para conocer cuáles son los vehículos de un servicio más próximos a él y cuál sería el costo y el tiempo de viaje hasta un destino seleccionado. Obviamente, para que esta funcionalidad sea operativa, es necesario que tanto el usuario como los vehículos estén geolocalizados, el primero por el GPS de su teléfono móvil y los segundos mediante localizadores GPS instalados en los vehículos. Así, todos los vehículos que prestan servicios de movilidad compartida tienen en común el estar dotados de dispositivos GPS que generan una enorme cantidad de datos sobre los viajes realizados, de gran utilidad para la gestión y la planificación del transporte. El desarrollo de nuevas plataformas de agregadores de servicios supone un paso adelante a la hora de tratar conjuntamente datos de distintos proveedores de servicios de movilidad. La tendencia y próximos desafíos

es la de alcanzar soluciones de *Mobility as a Service (MaaS)* en las que una única plataforma aporta información sobre todos los servicios de movilidad compartida de la ciudad, incluyendo el transporte público y opciones basadas en la intermodalidad, permitiendo realizar el pago de todos los servicios, independientemente de qué compañía o compañías se utilicen (Jittrapirom et al., 2017).



**Figura 3:** Nuevos servicios de movilidad compartida en las ciudades: servicios de patinetes y motos eléctricos en Madrid. Fuente: Fotografías de los autores

Por otro lado, cada vez más se va extendiendo el vehículo conectado y en un futuro próximo el uso del vehículo autónomo, se generalizará en nuestras ciudades. Los vehículos conectados están generando actualmente, una gran cantidad de datos que son enviados a los propietarios de los mismos a través de apps y a los fabricantes y/o empresas de servicios de gestión de flotas de vehículos a través de internet. Estos datos son obtenidos mediante localizadores GPS instalados en los vehículos conectados. Los fabricantes almacenan de esta forma enormes volúmenes de datos sobre los vehículos que producen o las flotas que gestionan, como número de kilómetros recorridos, uso de combustible, consumos, rutas recorridas, tiempos de viaje, lugar del último aparcamiento, etc. Estos datos son utilizados para proveer de información y recomendaciones a sus propietarios, por ejemplo, el momento de hacer la próxima revisión de los vehículos. Por su parte, las empresas de servicios de gestión de flotas aportan recomendaciones a sus clientes en cuestiones como la seguridad vial, la eficiencia, el cumplimiento de la normativa y la productividad. En el caso del vehículo autónomo, se requiere un gran flujo de datos en tiempo real para poder circular. Antes de comenzar un viaje el vehículo requiere conocer su localización (origen del viaje), la localización del lugar de destino e información actualizada en tiempo real sobre la red vial para calcular la ruta óptima e incluso ir modificándola, dinámicamente, en función de las circunstancias del tráfico. Por lo tanto, no solo requiere de sensores para hacer un reconocimiento de su entorno, sino también de dispositivos GPS para localizar el vehículo en todo momento a lo largo de la ruta.

Por lo expresado anteriormente, los datos masivos geolocalizados mediante dispositivos GPS resultan críticos en la gestión de la movilidad y la planificación del transporte en la actualidad y en el futuro lo serán cada vez más. El término de Smart Mobility, estrechamente conectado al de Smart City, hace referencia a este nuevo paradigma de la movilidad sostenible que gravita en torno al Big Data y las tecnologías de la información y las telecomunicaciones (Benevolo et al., 2016).





## 1.4 Objetivos y estructura de la nota técnica

Esta nota técnica tiene como objetivo presentar las oportunidades del uso de Big Data a partir de registros GPS, a fin de entender y planificar la movilidad urbana y pretende servir de material de consulta a los técnicos especialistas en datos del sector transporte, a la hora de realizar estudios de movilidad basados en datos de registros GPS. También, aportará información sobre las fuentes de datos de las que se puede obtener información GPS, las características específicas de este tipo de datos, las metodologías para derivar de los datos información útil en la planificación del transporte y las principales ventajas y limitaciones. Incluye, asimismo, estudios de caso que permiten ilustrar la multiplicidad de posibles aplicaciones. El aporte de este estudio permite encontrar en un mismo documento una agrupación y sistematización sobre los conocimientos generados por técnicos e investigadores en los últimos años en distintos países del mundo, orientando a aquellos planificadores de transporte y técnicos en datos que quieren adentrarse en este campo.

Muchos de los lectores potenciales de este documento no harán una lectura secuencial del mismo, de principio a fin, sino que previsiblemente tendrán interés en consultar un capítulo u otro de forma separada, en función de sus intereses específicos. Por ello, se ha optado por un formato en el que los diferentes capítulos y subcapítulos pretenden ser autoexplicativos en la medida de lo posible, lo que obliga a algunas reiteraciones, inevitables para facilitar una lectura no necesariamente secuencial.

### LA NOTA TÉCNICA SE ESTRUCTURA EN SIETE CAPÍTULOOS:

Este primer capítulo realiza una introducción a la revolución del *Big Data* y la aparición de nuevas fuentes de datos, con especial atención a los datos que contienen información geolocalizada, el llamado *Big Data* geolocalizado. Se presenta el *Big Data* con relación a las dos principales fuentes de creación de datos: los sensores y dispositivos, muchos relacionados con el desarrollo de las Smart Cities, y los datos provenientes de internet. La mayor atención se centra en los datos generados mediante los sistemas globales de navegación por satélite (*Global Navigation Satellite System, GNSS*), en especial, el sistema GPS (*Global Positioning System*), siendo el más difundido a nivel global y el más relevante en el contexto de esta publicación, por lo que se toma de referencia.

El segundo capítulo presenta las fuentes de las que podemos recoger datos de GPS y su utilidad con relación a la planificación del transporte y la movilidad urbana. En primer lugar, se hace una presentación de la evolución que se ha producido en las fuentes de datos, desde las más tradicionales, como las EDM, a la irrupción de las nuevas fuentes de datos en el marco de la revolución de las tecnologías de la información y las comunicaciones, sus ventajas, limitaciones y, sobre todo, sus complementariedades. Posteriormente se presentan y describen las nuevas fuentes basadas en registros GPS, como los datos provenientes de redes sociales con información geolocalizada (como Twitter, Flickr o Instagram), de otras redes sociales relacionadas propiamente con la movilidad (como Waze o Strava), o de servicios web (como Google Maps). En esta descripción de las fuentes se incluyen además los datos de *tracks* de autobuses urbanos, taxis o vehículos de nuevos servicios de movilidad compartida, los generados por el uso de tarjetas inteligentes de transporte o

de redes de transporte privado (como TomTom). Finalmente, se presentan los campos de aplicación de todas estas fuentes, en un análisis que considera la presencia de personas y sus distribuciones horarias en la ciudad, el uso de las redes de transporte público, la obtención de aforos de viajes y matrices origen destino, el estudio de la accesibilidad, el análisis de eventos y su impacto en la movilidad urbana y la percepción de los servicios del transporte.

El tercer capítulo describe las características de los registros GPS. Se diferencia entre datos puntuales, que muestran simplemente localizaciones (por ejemplo, accidentes de tráfico) y datos lineales (*tracks*), que muestran secuencias de puntos a partir de las cuales se pueden reconstruir líneas representativas de los viajes. Además, se presentan los diferentes tipos de ficheros usados para almacenar este tipo de datos, desde las tablas de texto y ficheros .csv (*comma-separated-values*), a los propios de tecnologías GPS como los ficheros .gpx (*GPS Exchange Format*). Posteriormente se describe la información contenida en los registros GPS con énfasis en dos de los aspectos más importantes de la misma: la resolución espacial y la resolución temporal de los datos. Finalmente, se dedica un apartado final a las limitaciones que tienen los datos de GPS y que deben considerarse en los estudios o análisis de planificación del transporte.

Los capítulos cuarto y quinto muestran metodologías para obtener información de movilidad a partir de datos de GPS, ya sea puntuales (capítulo cuatro) o de *tracks* (capítulo cinco). En el primer caso se toma como referente los datos puntuales procedentes de redes sociales como Twitter y Flickr, y se muestran los pasos a seguir para analizar pautas de movilidad a partir de las localizaciones cambiantes de los usuarios de estas redes sociales. En primer lugar, se utiliza Twitter y las localizaciones puntuales de aquellos usuarios que publican sus mensajes de manera geolocalizada. Se muestra la importancia de los procesos de limpieza y pre-proceso de datos, hasta disponer de una base que contenga únicamente datos con los que se obtenga información fiable sobre la movilidad de los usuarios. Para ello, es necesario seleccionar una muestra de usuarios válidos para el análisis y eliminar los registros que puedan generar ruido o interferencias en los resultados finales. Se muestran además las limitaciones de los datos, en especial en relación con la resolución temporal, y las operaciones para determinar espacios de actividades frecuentes y a partir de ellos matrices de viajes. Además, se introducen algunas formas de validación de los resultados obtenidos. Junto a la obtención de matrices de viajes usando datos de Twitter, se muestra la metodología a utilizar para conocer flujos de movilidad de turistas en espacios urbanos a partir de datos geolocalizados de redes de fotografías compartidas como Flickr.



El capítulo cinco, muestra las oportunidades del uso de *tracks*, diferenciando desde el inicio entre los *tracks* provenientes de compañías o entidades y los datos capturados por iniciativas específicas con el propósito particular de analizar la movilidad. Entre los primeros, están los de compañías como Endomondo, Garmin, Strava, Wikiloc, a los que se puede acceder de diferentes formas y pueden ser usados en distintos estudios de movilidad. Entre estos datos, están los capturados por los sistemas de bicicletas públicas o por los nuevos servicios de movilidad compartida. En el segundo caso, los datos de *tracks* proceden de distintas iniciativas, lanzadas con el objetivo específico de recopilar rutas de individuos e información asociada a dichas rutas, con la finalidad de realizar estudios e investigaciones concretas. Estas iniciativas llaman a la participación de voluntarios que dan su consentimiento para que sus datos sean utilizados con un objetivo establecido. Estos datos recopilados suelen corresponder con *tracks* individuales para cada ruta realizada y los participantes suelen aportar información relativa a sí mismos y a la ruta que realizan, como la edad, el género o el motivo del viaje. El volumen de datos, es mucho menor que en los *tracks* provenientes de compañías o entidades públicas.

Una vez presentadas las posibles procedencias de los *tracks*, se realiza una aproximación a la metodología para el tratamiento de los datos, que incluye la propia incorporación de los mismos a un sistema de información geográfica, su geolocalización, los procesos de enriquecimiento y de map matching (ajuste cartográfico) para ajustar las rutas a la geometría de una red de transporte, de carreteras o calles. Con los datos ya depurados, se presenta su utilidad a la hora de analizar patrones de movilidad, el impacto de eventos urbanos o la realización de simulaciones de escenarios futuros en la implementación de políticas o actuaciones en las infraestructuras.

En el capítulo seis se incluyen algunos casos de estudio para ilustrar las posibilidades de estas fuentes de datos. Los casos de estudio muestran diversas experiencias en los contextos de Europa y América, a fin de mostrar aplicaciones a distintos modos de transporte. Además, se presentan experiencias que usan tanto *tracks* (recorridos) como datos de puntos de interés (datos puntuales), para abordar diferentes temáticas. Así, los casos de estudio son sobre: 1) el análisis de los *tracks* de un sistema de bicicletas público para conocer los patrones de movilidad de sus usuarios, permitiendo obtener un diagnóstico muy útil a la hora de planificar y realizar inversiones en infraestructura, 2) la utilización de registros GPS de las flotas de taxis para estudiar la demanda en este modo de transporte, 3) el uso de datos de TomTom para el análisis dinámico de la congestión y la accesibilidad en las ciudades, 4) la combinación de datos de registros GPS de los autobuses, la tarjeta de transporte y otras fuentes para analizar la movilidad de los usuarios del autobús, y 5) la integración de datos de fuentes oficiales y de Waze para el estudio de los accidentes de tráfico.

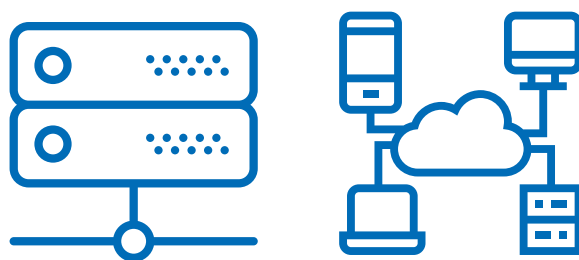
En las consideraciones finales, capítulo siete, se recogen las oportunidades que ofrecen los datos basados en registros GPS y sus principales características, a fin de servir de utilidad para que planificadores y técnicos en datos del sector de transporte puedan aplicarlos al diagnóstico de la movilidad urbana.





# 2

## Fuentes de datos y campos de aplicación



## 2.1 De las fuentes de datos tradicionales a las nuevas fuentes de datos

Las encuestas domiciliarias de movilidad (EDM) han sido durante mucho tiempo la principal fuente de información para el estudio de la movilidad. Estas encuestas recogen información sobre los viajes y también sobre las características sociodemográficas y económicas de la población. Al tratarse de una fuente específica para el estudio de la movilidad y para alimentar los modelos de transporte, las encuestas de movilidad aportan una información de gran valor para técnicos y planificadores. Sin embargo, las EDM también presentan algunos inconvenientes: se requieren enormes muestras de hogares encuestados, con un considerable trabajo de campo, con largos plazos desde que comienza la operación hasta que se dispone de la información y elevados costos para su realización. Por ello la periodicidad de estas encuestas es relativamente baja (normalmente se realizan cada 10 años), con lo que la información que aportan queda rápidamente desactualizada. Otras limitaciones de esta fuente de datos son, por ejemplo, los viajes no reportados por los entrevistados, la imprecisión en el tiempo de salida y llegada del viaje y la falta de información precisa sobre localizaciones o rutas (BID, 2019).

Existen otros tipos de encuestas de panel de movilidad que recogen información de una muestra de población, generalmente pequeña, durante un periodo determinado de tiempo (por ejemplo, una semana) y en distintos momentos temporales. De esta forma es posible recopilar información sobre cambios en patrones de movilidad, normalmente como consecuencia de algún tipo de actuación en materia de transporte. Adicionalmente, se dispone de conteos y aforos en los distintos medios de transporte. En otras ocasiones se utilizan técnicas cualitativas, como diarios de viaje, observaciones, entrevistas, grupos de discusión, etc., basándose en la interacción directa con la población objeto de estudio. Estas técnicas se han usado con cierta frecuencia para analizar la movilidad de determinados grupos de población. Asimismo, se puede obtener información de otras fuentes, no propiamente orientadas a los estudios de movilidad, como los censos de población, para obtener datos sobre los lugares de trabajo y vivienda y del medio de desplazamiento o el tiempo utilizado (Osorio-Arjona y García-Palomares, 2017).

El avance tecnológico abre nuevas oportunidades en los estudios de movilidad, ofreciendo nuevas fuentes de datos que aportan información complementaria a la de las fuentes tradicionales. Entre las ventajas que ofrece el *Big Data* en general para el estudio de la movilidad urbana destacan las siguientes (BID, 2019):

- **GRAN ALCANCE ESPACIAL:** Las redes de sensores y dispositivos que capturan datos masivos normalmente están desplegadas por todo el territorio. Esto permite analizar la movilidad de los usuarios tanto dentro como fuera del área de estudio definida, pudiendo caracterizar con bastante detalle el origen y destino final de los viajes.
- **GRANDES VOLÚMENES DE DATOS:** Estas redes de sensores y dispositivos permiten trabajar con grandes volúmenes de datos, ya sea de muestras de gran tamaño o incluso de datos de todos los usuarios de un sistema de transporte.
- **RESOLUCIÓN TEMPORAL ALTA:** Los datos masivos generalmente se recogen de manera continua (datos continuos), lo que permite tener información constantemente actualizada y monitorizar la movilidad en tiempo real o casi real.
- **RESOLUCIÓN ESPACIAL ALTA-MEDIA/ALTA:** El nivel de detalle en la geolocalización de los datos masivos varía, pero en todo caso es bastante alta, llegando a ser muy precisa cuando trabajamos con datos de GPS.
- **COSTO Y TIEMPO:** Los costos que supone el uso del Big Data para el análisis de la movilidad son mucho menores que los empleados en la realización de las encuestas de movilidad. También es mucho menor el periodo necesario para la elaboración de los análisis y la obtención de la información.

Los registros de telefonía móvil son, sin duda, un ejemplo muy conocido entre las fuentes de datos masivos. A ellos se dedicó una publicación anterior que explica cómo aplicar big data a la planificación de transporte urbano (BID, 2019). Esta nueva publicación se centrará en datos geolocalizados a través de GPS, con una gran diversidad de fuentes de datos, que ofrecerán información complementaria relativa a la movilidad en distintos modos de transporte.



## 2.2 Datos obtenidos a partir de localizadores GPS

Los registros GPS contienen al menos datos sobre el usuario (identificador de usuario) y las coordenadas geográficas (latitud, longitud) y temporales (año, mes, día, minuto y segundo) del lugar y el momento en el que se hace un registro. Por lo tanto, se trata de datos puntuales. Las localizaciones de personas o vehículos a lo largo del tiempo se almacenan como secuencias de puntos, lo que hace posible inferir información sobre los viajes.

La resolución espacial de los datos es más o menos homogénea, con un error de unos 3 o 4 metros, dependiendo de las condiciones en que se hizo la medición por medio del GPS, pero la resolución temporal es muy variable:

- **PUNTOS DE INTERÉS (DATOS DE LOCALIZACIÓN PUNTUALES):** En algunos casos se dispone de datos de localizaciones registrados por acciones del usuario de forma esporádica. Este tipo de datos pueden ser considerados

como datos de localización simples o puntos de interés. Es lo que ocurre por ejemplo con Waze, cuando los usuarios registran la información de un accidente de tráfico, con Twitter, cuando el usuario envía un tweet geolocalizado. Es posible inferir lugares de residencia y de trabajo, lugares de actividades poco frecuentes o esporádicas, lugares de ocio fuera de su ciudad, etc., con lo que se pueden reconstruir ciertos tipos de viajes.

- **TRACKS O RECORRIDOS:** En otros casos se dispone de datos de viajes registrados de forma automática a intervalos regulares de tiempo, generalmente de unos segundos, que se conocen como recorridos o *tracks* en inglés. Se trata de secuencias de puntos que describen las rutas con mayor o menor precisión en función de la resolución temporal del dispositivo (intervalo de tiempo entre localizaciones). Es lo que ocurre, por ejemplo, con las rutas registradas por aplicaciones como Google Maps o con los dispositivos tanto de los vehículos particulares conectados como de los vehículos de los servicios de movilidad compartida. Los *tracks* aportan una información muy rica para el análisis de los patrones de movilidad. Algunas encuestas de movilidad (por ejemplo, la última de Santiago de Chile) incluyen el “trackeo” de entrevistados que acceden a que sus viajes puedan ser registrados en una app durante un periodo de tiempo, ofreciendo así una información complementaria a la captada por la propia encuesta (por ejemplo, hora exacta del comienzo y el fin del desplazamiento, velocidad, tiempo de viaje, ruta seguida, etc.).

Los datos anteriores son registros individuales y están debidamente anonimizados, de forma que el identificador de usuario es simplemente una clave aleatoria que no permite identificar a la persona que realiza las acciones.

Algunas empresas no ofrecen datos individuales, desagregados, sino datos agregados y tratados espacialmente. Es el caso, por ejemplo, de los datos de tiempos de viaje que se pueden obtener de la API de Google Maps o de las matrices de viajes que ofrece la API de TomTom. En estos casos la privacidad de los usuarios está todavía más asegurada, pero el técnico o investigador no puede sacar el máximo provecho de los datos, ya que tiene que limitarse a las funcionalidades que le ofrece la API.

Una cuestión clave al trabajar con datos provenientes de registros GPS es la de su accesibilidad. Algunos de estos datos son producidos por las administraciones públicas, de manera que, debidamente anonimizados pueden ser puestos a disposición de la sociedad. Así, por ejemplo, la Empresa Municipal de Transportes de Madrid cuelga en su página web datos sobre los *tracks* de los autobuses y del sistema de bicicletas públicas de alquiler (Bicimad), que pueden ser descargados gratuitamente. Pero la mayor parte de los datos GPS son generados por empresas, con diferentes políticas de acceso a los datos. Algunas directamente no permiten el acceso a los datos. Otras, como Twitter, los venden desagregados y anonimizados. Y otras ofrecen productos derivados de ellos, como los citados casos de Google Maps y TomTom. Tanto Twitter como Google Maps permiten descargar muestras gratuitas, que permiten que el técnico o investigador conozca sus características e incluso pueda realizar estudios con ellas.



## 2.3 Nuevas fuentes de datos basadas en registros GPS

### REDES SOCIALES

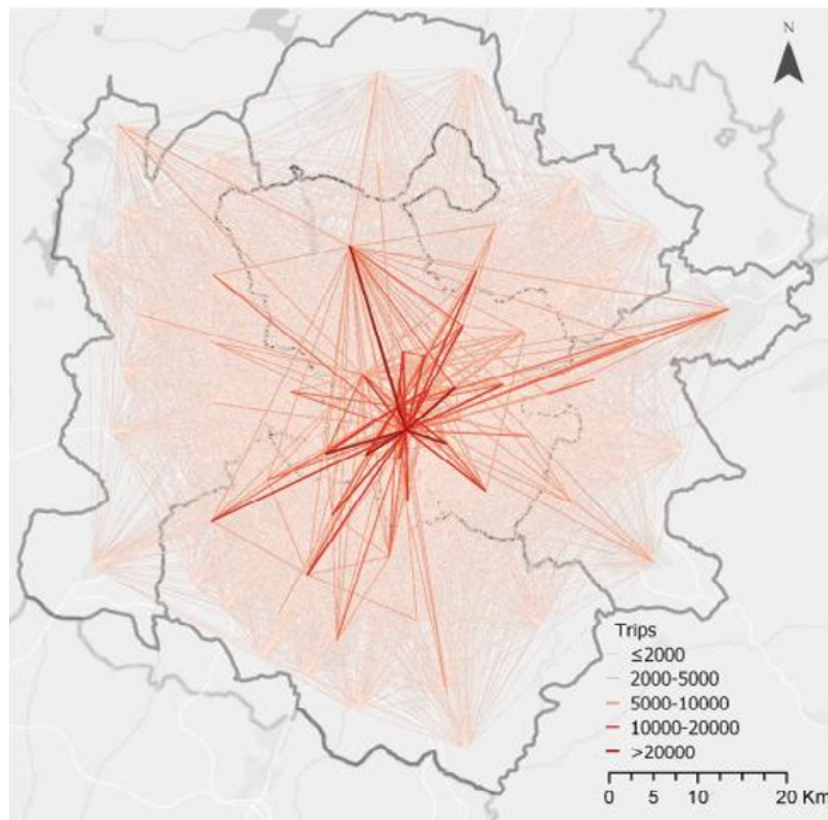
Una de las herramientas más utilizadas en Internet son las redes sociales, que permiten crear, almacenar, compartir e intercambiar información con otros usuarios, generando diariamente una gran cantidad de datos. Muchas empresas usan las redes sociales como recursos para obtener información, estudiar campos de actuación o publicitarse. Los datos generados por las redes sociales, que frecuentemente están geolocalizados, tienen un carácter dinámico debido a su constante actualización, lo que permite conocer las pautas de distribución de los usuarios y su movilidad. Además, las redes sociales son valiosas para informar acerca del tráfico a distintas horas del día, compartir horarios o eventos, dar a conocer empresas y modos de transporte, o planificar proyectos que afecten a distintas áreas urbanas (Osorio-Arjona y García-Palomares, 2017).

Una de las redes sociales más difundidas es Twitter. Este servicio cuenta con unos 500 millones de usuarios en todo el mundo y genera alrededor de medio millón de tweets al minuto. Una pequeña parte de ellos son tweets geolocalizados, es decir, mensajes para los que se conoce la localización del emisor por sus coordenadas geográficas (Gutiérrez Puebla et al., 2016). Numerosos estudios realizados con datos masivos de redes sociales han utilizado Twitter no solo por tratarse de una plataforma de cobertura mundial, sino también por el hecho de que sus datos están disponibles en la red de forma gratuita a medida que se producen, es decir, en tiempo real. Se pueden descargar tweets geolocalizados (o no), registrando el hashtag, indicador de usuario, momento en el que se envía, idioma, tipo de dispositivo, texto del mensaje, etc.

Cada tweet deja una “huella” digital del lugar y el momento en que fue enviado. Si se procesan los datos según identificador de usuario, se puede tener una aproximación de los lugares que visita cada usuario a lo largo del tiempo. Así es posible utilizar esta fuente para analizar los patrones de movilidad de la población (Wu et al., 2014) y generar matrices origen-destino, tanto a escala urbana (Osorio-Arjona y García-Palomares, 2019) como a escala nacional o internacional.

La fiabilidad de este tipo de datos ha sido validada en el trabajo de Lenormand et al. (2014), en donde se contrastan los datos de Twitter con información de teléfonos móviles y datos oficiales de censos. El estudio concluyó que las tres fuentes ofrecen en general resultados comparables en términos de pautas de movilidad de los individuos. Por su parte, Osorio-Arjona y García-Palomares (2019) validaron sus resultados con los obtenidos a partir de una encuesta domiciliaria de movilidad, concluyendo que tales resultados son comparables. El hecho de que los datos de Twitter ofrecen muestras de menor tamaño que las de telefonía móvil, obliga a trabajar con unidades espaciales de mayor extensión. En resumen, a pesar del sesgo inherente a esta fuente de datos, los análisis realizados demuestran que las matrices de viajes que aportan son coherentes y ofrecen una aproximación fiable a los patrones generales

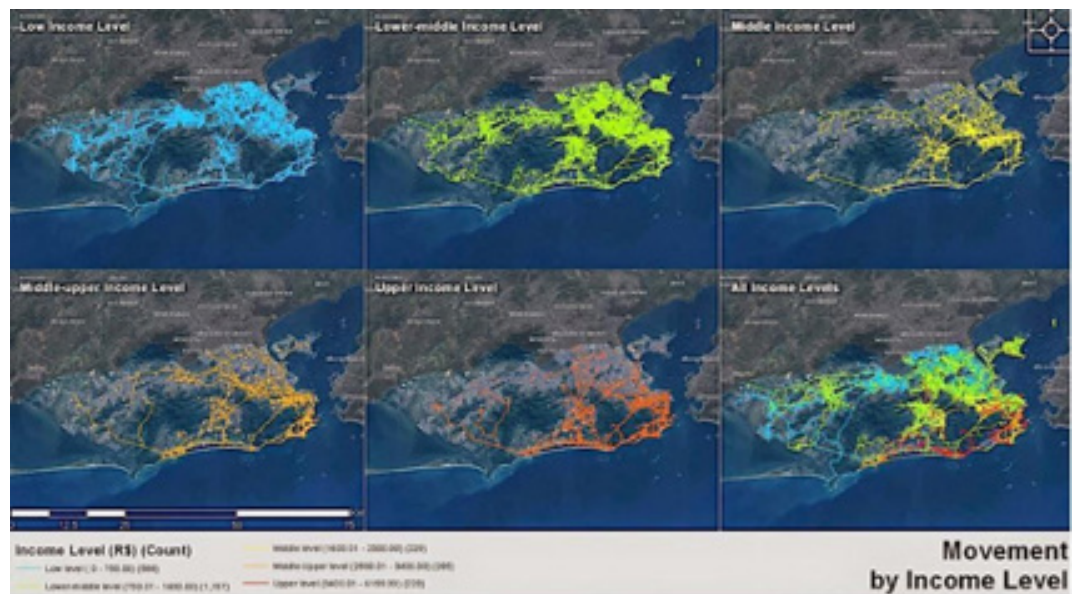
de movilidad de ciudades y regiones, a un costo muy reducido, lo que tiene especial interés para aquellas que no disponen de datos fiables de movilidad.



**Figura 4:** Representación de los flujos de una matriz origen-destino obtenida con datos de Twitter en el área metropolitana de Madrid. Fuente: Osorio-Arjona y García-Palomares, 2019.

Asimismo, se han utilizado tweets geolocalizados para analizar el grado de mezcla social en el uso del espacio, rastreando el movimiento de los grupos sociales en ciudades como Río de Janeiro (Netto et al., 2015), Louisville (Shelton et al., 2015) o Concepción (Salas-Olmedo y Rojas Quezada, 2017). Los estudios sobre segregación social basados en la información suministrada por las fuentes oficiales (censos y registros oficiales), que ofrecen datos relativos al lugar de residencia, se limitan al análisis de la segregación residencial. En cambio, los estudios dinámicos basados en nuevas fuentes de datos, como Twitter, aportan indicadores de multiculturalidad y mezcla social referidos al uso del espacio a lo largo del día (segregación dinámica). Gracias a la geolocalización es posible cruzar los datos de las redes sociales con información sociodemográfica procedente de las estadísticas oficiales, de forma que se puede estimar el nivel de ingresos de cada usuario a partir de la sección censal en la que habita, lo que se puede inferir a partir del lugar desde donde envían más tweets en horario nocturno (Figura 5).

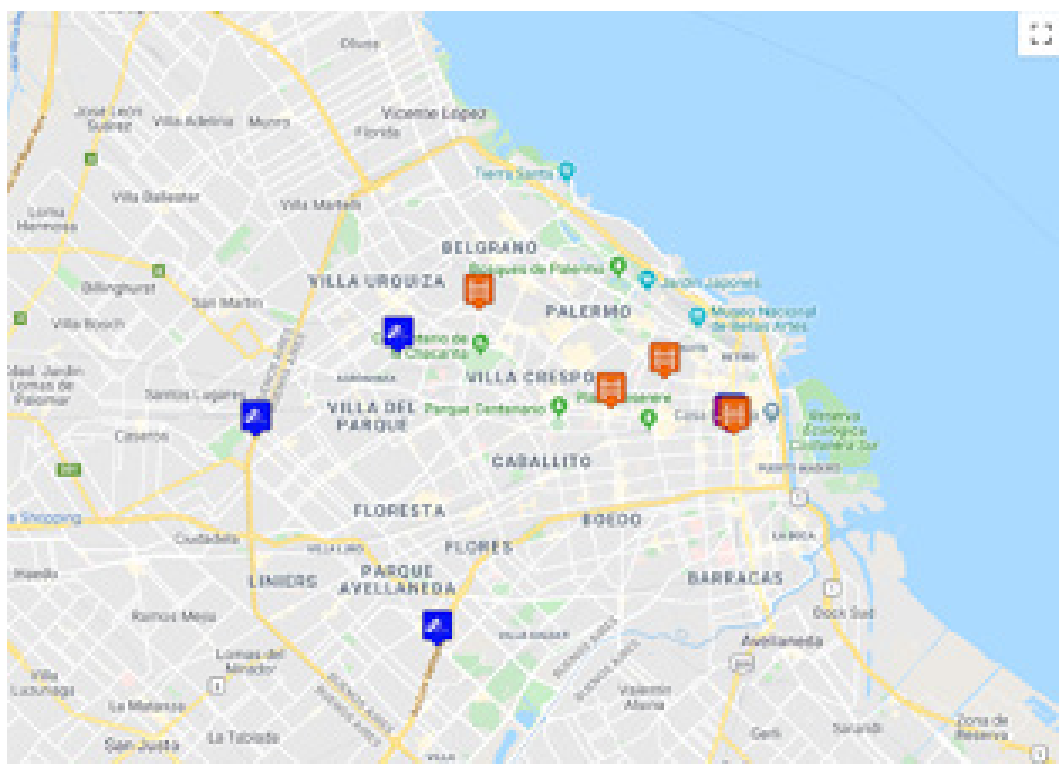




**Figura 5:** Una visión dinámica de la segregación social en Río de Janeiro. Los colores indican patrones de localización a lo largo del día según grupos sociales: azul (bajos ingresos), verde (bajos-medios), amarillo (medios), naranja (medios-altos) y rojo (altos). El mapa de abajo a la derecha muestra una combinación de los cinco mapas primeros y permite diferenciar entre espacios segregados (solo un color) y espacios con mayor mezcla social (varios colores). Fuente: Netto et al., 2015.

Por otro lado, el análisis de los contenidos de los tweets puede aportar información en tiempo real sobre las condiciones del tráfico y accidentes de circulación. Los tweets deben ser filtrados para seleccionar los que tienen información sobre el tráfico y localizados en un mapa para informar de forma rápida a los usuarios. Actualmente, hay un prototipo, (de acceso público en <http://intranet.isistan.unicen.edu.ar/>) que monitoriza la Ciudad Autónoma de Buenos Aires (CABA), mostrando sobre un mapa la localización de los incidentes y los tweets que los reportan. Esta información sirve de ayuda a los conductores para elegir una ruta que evite las retenciones derivadas de esos incidentes (Vallejos et al., 2017).





**Figura 6:** Mapa de incidentes de tráfico en tiempo real en la Ciudad Autónoma de Buenos Aires (CABA) a partir de tweets. Fuente: <http://intranet.isistan.unicen.edu.ar/>

Otras redes sociales de gran interés son las de fotografías compartidas, como Instagram, Flickr y Panoramio, que permiten analizar los registros de fotografías a partir de su geolocalización y sus coordenadas temporales. Estas fuentes permiten conocer los espacios más fotografiados por los turistas y, en consecuencia, los más visitados de las ciudades, siendo posible también realizar el seguimiento de la movilidad de los turistas en las ciudades a partir de las fotografías que toman y comparten en las redes (García-Palomares et al., 2015).

## WAZE

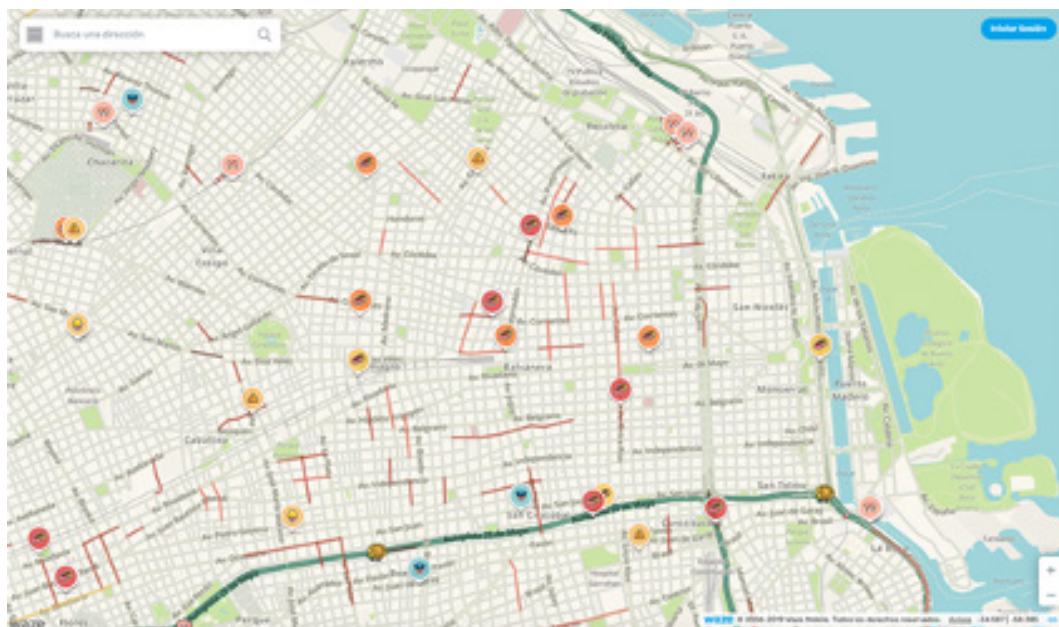
A través de una App para móviles inteligentes y apoyada en registros GPS, Waze proporciona la mejor ruta en coche, con información en tiempo real sobre la situación de la red vial. La aplicación surgió en Israel en 2008, donde a finales de ese año tenía algo más de 20.000 usuarios. Hoy Waze tiene presencia en más de 185 países y supera los 115 millones de usuarios (<https://www.waze.com/>). La plataforma ha sido adquirida por Google e integrada también dentro de su aplicación Google Maps.

Waze genera gran cantidad de información sobre el tráfico, que puede ser utilizada por los planificadores urbanos en los estudios de movilidad. Desde el momento en que los usuarios abren la aplicación para solicitar información sobre una ruta, la aplicación recoge y ofrece en tiempo real datos sobre la velocidad a la que se circula, los puntos donde se detiene el tránsito, los tiempos estimados de trayecto, etc. De esta manera, la red vial en la que se basa el cálculo de las rutas se alimenta de forma similar a la de las redes



de Google Maps o TomTom. Además, Waze se alimenta de información proporcionada por sus usuarios, llamados “wazers”, que pueden crear nueva información sobre el estado del tráfico, localizando posibles eventos que aparecen en la ruta o convirtiéndose en editores de la cartografía. El usuario puede interactuar con la aplicación proporcionando información cuando llega a un atasco, un accidente, un peligro, una obra, carretera cortada o a radares de policía. Esa información puede ser nueva (creando el evento) o actualizada (validando dicho evento o reportando la finalización del mismo). La aplicación almacena las coordenadas de los eventos creados por los usuarios, la hora a la que fue registrado y la duración del mismo, el sentido de circulación del vehículo, el tipo de calle y el grado de confianza en ese evento y usuario.

La Figura 7 reproduce el mapa del centro de Buenos Aires según Waze el jueves 17 de octubre a las 08:00 hora local. En ella podemos ver las velocidades en los tramos de la red y los eventos reportados por los usuarios en ese momento.



**Figura 7:** Situación del tráfico en el centro de Buenos Aires reportada por Waze (17/10/2019, 08:00H). Fuente: <https://tn.com.ar/waze>

La información de Waze es accesible a través de su API y de otros procedimientos de descarga. Por un lado, podemos acceder a la información de eventos del tráfico. Existe una amplia tipología de eventos, normalmente de carácter puntual, que pueden ser utilizados para analizar los niveles de accidentalidad viaria. Varios trabajos lo han realizado ya en ciudades iberoamericanas, como, por ejemplo, Belo Horizonte – Brazil (Dos Santos et al., 2016) o Ciudad de México (Pérez et al., 2018). Por otro lado, es posible utilizar la información de la red de carreteras, que incluye datos relativos a las características de cada tramo (tipo de vía, número de carriles, etc.) y sus velocidades. Esta información puede ser usada, por ejemplo, en estudios de velocidades de tráfico y análisis de los impactos de la congestión.

## TARJETAS INTELIGENTES DE TRANSPORTE (TRANSPORT SMART CARDS)

Una fuente de gran interés para el estudio de la movilidad en transporte público es la enorme cantidad de datos generada por las tarjetas inteligentes de transporte (transport smart cards). Se trata de las conocidas tarjetas de transporte público, que progresivamente han ido desplazando a otros medios de pago (billetes magnéticos o tickets), para imponerse como un sistema seguro de pago y validación, que libera al conductor de la tarea del pago de los viajeros y que suministra ingentes cantidades de datos a las empresas y agencias de transporte público (Pelletier et al., 2011). Las tarjetas inteligentes proporcionan una información exhaustiva (no muestral) sobre los viajes en transporte público en las ciudades donde su uso es obligatorio (por ejemplo, Buenos Aires) y casi exhaustiva en otras ciudades (dependiendo de su tasa de penetración). Por ello, la cantidad de datos a procesar es realmente elevada. Así, por ejemplo, la tarjeta de transporte de Santiago de Chile, que es de uso obligado en los autobuses y abrumadoramente mayoritaria en metro (con una tasa de penetración global del 97% de los viajes), genera alrededor de 40 millones de registros a la semana (Gschwender et al., 2016).

Existen básicamente dos tipos de tarjetas: sin contacto y de contacto. Las tarjetas sin contacto registran automáticamente la entrada y salida del viajero de una estación o de un vehículo de transporte público a través de ondas de alta frecuencia; en cambio, las tarjetas de contacto exigen que el usuario pase la tarjeta por un lector, cosa que normalmente sólo hace en la entrada. En el caso de las estaciones de metro o ferrocarril, la geolocalización va asociada a cada una de ellas, así como también en las paradas de autobús cuando el registro del viaje se realiza en máquinas instaladas en las propias paradas. Si la validación del viaje se hace en lectores instalados en los autobuses, la localización del autobús en ese momento puede quedar registrada si el lector de la tarjeta está conectado a un dispositivo GPS instalado en el vehículo. Cada vez que se hace una validación se registran normalmente datos como hora del día y día de la semana, identificador de la tarjeta, identificador de la línea y de la dirección, parada o estación de acceso (y, en su caso, de salida), identificador del autobús, etc. Por lo tanto, mediante estas tarjetas se recogen datos relativos a todos los viajes de los poseedores de las mismas en transporte público. Dependiendo del tipo de tarjeta, en muchos casos también se dispone de las características sociodemográficas del usuario (registradas al adquirir o renovar la tarjeta).

La explotación de estos datos en estudios de movilidad entraña cierta dificultad, no sólo por el gran volumen de datos que se genera, sino también por las lagunas de información existentes en la generación de los datos. Los retos son mayores en el caso de las tarjetas de contacto, ya que, al registrar solamente datos sobre la entrada del viajero, para conocer la salida se deben desarrollar algoritmos que tengan en cuenta el comportamiento del usuario a lo largo del día y de la semana, considerando los otros lugares por los que accede al transporte público. En el caso de los viajes de ida y vuelta, la parada o estación de salida del viaje de ida se puede inferir a partir de la parada o estación de entrada en el viaje de vuelta. Sin embargo, se dan situaciones más complejas en el caso de cadenas de viajes que incluyen viajes no basados en casa. Por otro lado, es importante desarrollar procedimientos para diferenciar correctamente entre viajes y etapas, de forma que un trasbordo debe ser reconocido como tal y no como un viaje adicional, en función del tiempo

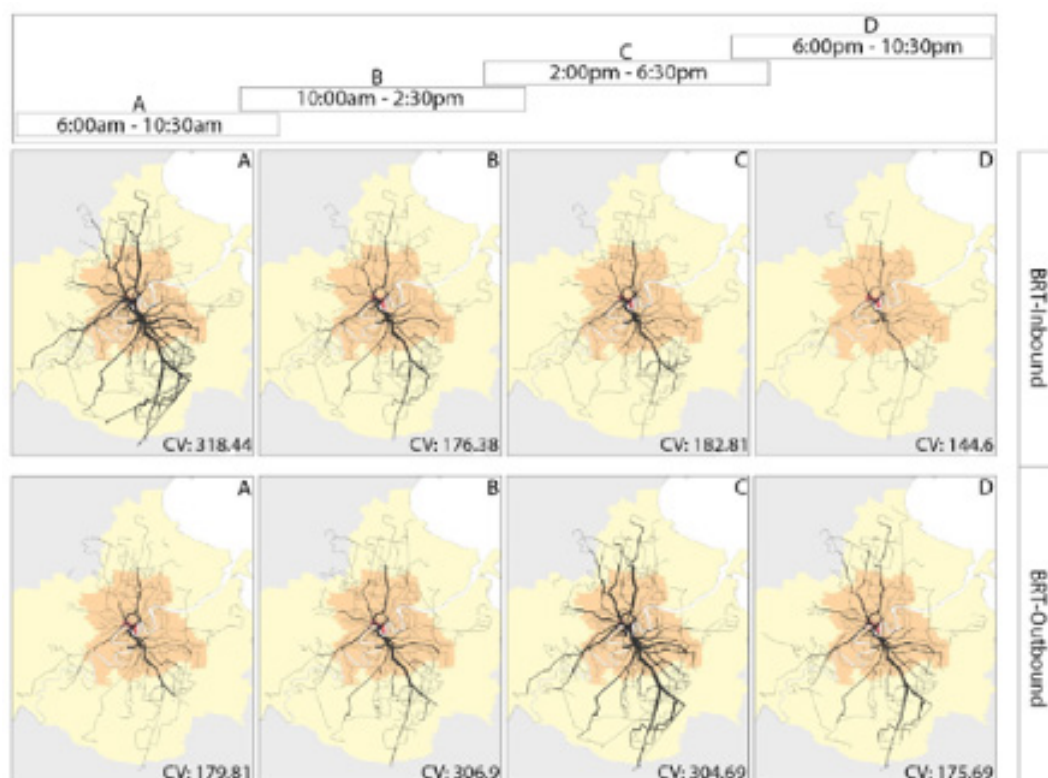
transcurrido entre validaciones. Para identificar las estaciones de bajada en viajes de varias etapas resulta de gran utilidad el cruce de distintas bases de datos, como las transacciones de las tarjetas, los registros GPS que muestran la posición de los autobuses en cada momento del día y la cartografía digital de líneas, paradas y estaciones (Munizaga y Palma, 2012).

Por otro lado, con las tarjetas de transporte quedan registradas las paradas de subida y bajada del autobús, pero no los puntos intermedios, que, sin embargo, pueden ser conocidos a partir del identificador de la línea en que presta servicio el autobús. Para esta tarea resulta de gran interés disponer de ficheros GTFS (General Transit Feed Specification), que contienen la cartografía de las líneas y las paradas. Cruzando ambas fuentes es posible visualizar en un mapa las trayectorias de los distintos viajes y analizar la intensidad de viajes en cada tramo de la línea o la red (Tao et al., 2014) (Figura 8). En el caso de tarjetas para las que no existe información socioeconómica del usuario, es posible enriquecer la información de los viajes cruzándola con la obtenida a través de encuestas de movilidad. A la inversa, esta fuente de datos puede utilizarse para corregir o completar los datos obtenidos mediante encuestas de movilidad (Trepanier et al., 2019).

A partir de los datos de las tarjetas inteligentes de transporte, es posible calcular la longitud de los viajes y generar matrices origen-destino (Munizaga et al., 2010). Además, la información espacio-temporal se puede cruzar con los datos socioeconómicos de los viajeros, lo que proporciona información muy valiosa para la planificación (Pelletier et al., 2011). Así mismo, cruzando los datos del uso de las tarjetas con la cartografía de las líneas, es posible analizar la carga en cada tramo de la red y estudiar de forma dinámica los flujos en transporte público (Tao et al., 2014).







**Figura 8:** Matrices OD y flujos en transporte público a partir de las tarjetas inteligentes de transporte, con predominio de viajes centrípetos en la mañana y centrífugos en la tarde. Fuente: Tao et al., 2014

## TRACKS

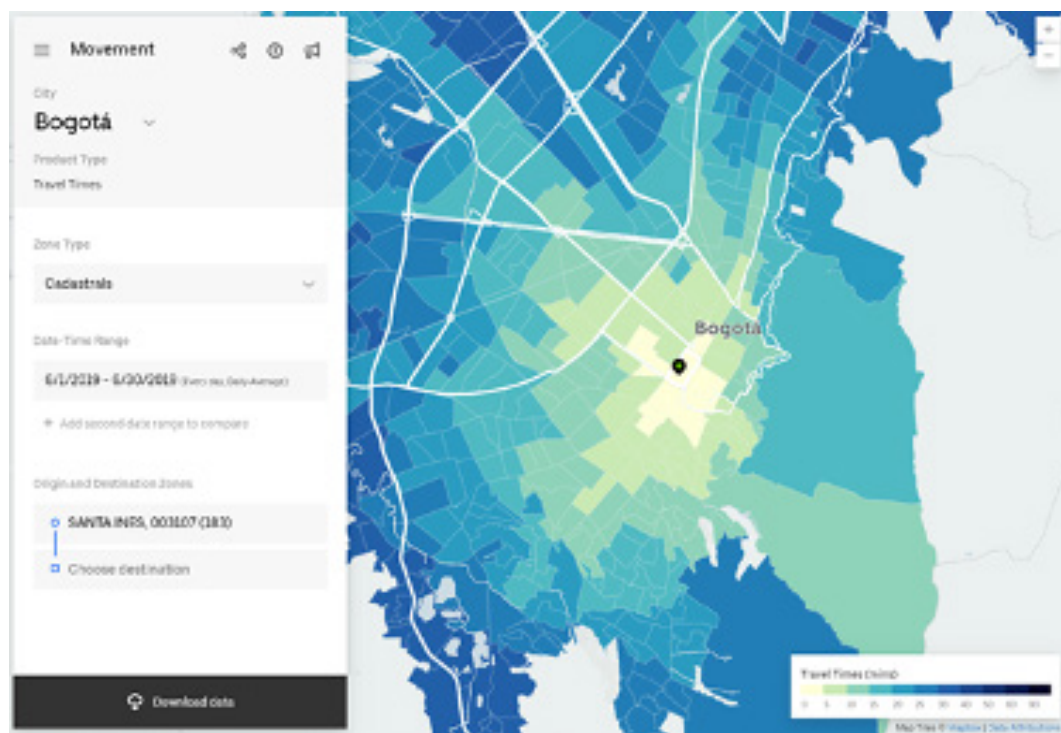
Un gran número de los vehículos que circulan por las ciudades están equipados con dispositivos GPS: autobuses urbanos, taxis, vehículos de nuevos servicios de movilidad compartida (bicicletas, motos, patinetes eléctricos, coches, etc.), vehículos particulares conectados, etc. Todos estos vehículos producen millones de datos de recorridos (*tracks*), de gran valor para gestores y planificadores.

Los autobuses urbanos de numerosas ciudades también están equipados con dispositivos GPS, que permiten la gestión de flotas en tiempo real y alimentan los sistemas de información al usuario, ofreciéndole información constante sobre los tiempos de espera. Así, por ejemplo, todos los autobuses urbanos de Santiago de Chile están dotados de dispositivos GPS que registran la localización de los vehículos cada 30 segundos, generando entre 80 y 100 millones de observaciones a la semana (Gschwender et al., 2016). A diferencia de los ficheros GTFS (General Transit Feed Specification), que incluyen datos sobre horarios, frecuencias y tiempos de viaje teóricos, los ficheros generados por los dispositivos GPS aportan información sobre velocidades y tiempos de viaje reales.

Los datos que producen los dispositivos GPS permiten monitorizar el avance de los autobuses en sus rutas (Cortés et al., 2011), comprobar los retrasos sobre los horarios previstos y estudiar el fenómeno del agrupamiento de autobuses (*bus bunching*), que produce un círculo vicioso en los retrasos de los autobuses (Byon et al., 2011). Detectar los problemas en tiempo real permite

actuar también en tiempo real para mitigar los problemas derivados del agrupamiento de autobuses. Así mismo, los datos históricos permiten analizar la variabilidad de los tiempos de viaje y los factores que inciden en ella, como el uso del suelo, el número de paradas, el número de semáforos, la longitud de la línea, la lluvia, etc. (Mazloumi et al., 2009). El análisis de datos históricos sobre tiempos de llegada de los autobuses y eventos de agrupamiento con base en ficheros de registros GPS permite hacer predicciones, ya sea mediante modelos explicativos o recurriendo a la inteligencia artificial, y tomar medidas para mejorar la puntualidad y evitar futuros episodios de agrupamiento de autobuses.

Las flotas de taxis de algunas empresas también cuentan con localizadores GPS para optimizar el servicio a los clientes, de forma que ante la solicitud de un servicio pueda acudir alguno de los vehículos más cercanos al cliente. Los *tracks* almacenados permiten que los gestores de la empresa puedan seguir los recorridos de los taxistas y analizar su rendimiento: número de viajes, kilómetros recorridos en cada viaje, viajes en vacío, áreas de máxima demanda según horas del día, etc. Así mismo, las empresas de alquiler de vehículos con conductor (como Uber o Cabify) cuentan con este tipo de dispositivos para analizar las rutas de sus empleados. En general, estos datos son de difícil acceso, ya que pertenecen a empresas privadas. Una excepción es Uber, que ha creado un servicio web que permite visualizar tiempos de viaje y velocidades entre zonas de transporte en varias ciudades del mundo, algunas de ellas latinoamericanas, y descargar, de forma gratuita, los datos de tiempos de viaje en ficheros csv (ver Figura 9 y enlace en: <https://movement.uber.com/?lang=es-ES>).



**Figura9:** Página web del portal de datos de Uber. Fuente: <https://movement.uber.com/?lang=es-ES>

Por otro lado, los localizadores GPS son esenciales para la operación de los servicios de movilidad compartida. Estos dispositivos previenen los robos de vehículos, cuya localización puede ser seguida en tiempo real desde las centrales de operación. Pero, además, la geolocalización es imprescindible para la operativa de los servicios de movilidad que no cuentan con bases propias (los llamados *free-floating o dockless*): una app debe geolocalizar al usuario y a los vehículos disponibles en su entorno, para ofrecerle la posibilidad de alquilar alguno de los más próximos, ya se trate de los de una empresa en particular (apps específicas) o los de varias empresas (apps de agregadores de servicios y servicios de movilidad *MaaS*). En cuanto al acceso de gestores y planificadores a los datos generados por los GPS de estos vehículos, en gran parte están en poder de empresas, pero otros muchos están en manos de las autoridades que gestionan algunos de estos servicios, como los de bicicletas compartidas en ciudades como Buenos Aires o Santiago de Chile.

### GOOGLE MAPS

Esta popular plataforma de Google permite el cálculo de caminos mínimos en la web o mediante su app, mostrando la ruta a seguir entre un origen y uno o varios destinos, así como el tiempo de viaje. En muchas ciudades es posible elegir el modo de transporte, diferenciando entre auto, transporte público, bicicleta y a pie. Los tiempos de viaje, tanto en vehículo privado como en transporte público, varían según las horas del día en función de la congestión y de las frecuencias, respectivamente. Los cálculos de los tiempos de viaje en transporte privado se basan en datos históricos de millones y millones de *tracks* almacenados por los servidores de Google (que pueden ser visualizados por cada usuario utilizando la opción de “cronología” en Google Maps), aplicando algoritmos para inferir el modo de transporte. La estimación de los tiempos de viaje en tiempo real se basa en algoritmos que combinan datos históricos con datos del tráfico en tiempo real, obtenidos de los propios teléfonos móviles. Las muestras en que se basan los cálculos de Google son gigantescas, ya que la inmensa mayoría de teléfonos inteligentes tienen instalado Google Maps y activada la geolocalización.

El servicio que ofrece la web y app de Google Maps es útil para el usuario habitual, que sólo requiere el cálculo de una o varias rutas, pero no tanto para el técnico o investigador, que necesita calcular millones de rutas. Para este fin está disponible la API de Google Maps<sup>6</sup>, que permite calcular matrices origen-destino de tiempos de viaje en los distintos modos de transporte. La consulta es gratuita o no en función del número de orígenes y destinos. En el caso del transporte privado se puede elegir la hora y el día de la semana. Esos datos reflejan la congestión en las ciudades y son de gran utilidad para técnicos e investigadores. Así, por ejemplo, García-Albertos et al. (2019) analizaron los cambios en la accesibilidad en Madrid según franjas horarias en función de las condiciones de congestión, con datos obtenidos de la API de Google Maps.

Además, en esta API es posible obtener matrices de tiempos de viaje en transporte público gracias a los ficheros GTFS suministrados por las autoridades de transporte de las ciudades. Se trata, por lo tanto, de tiempos de viaje “teóricos”, en función de horarios y frecuencias, no de los tiempos de viaje reales, que en el caso de los autobuses pueden ser analizados, como

<sup>6</sup> <https://developers.google.com/maps/documentation/?hl=es>

hemos visto, a partir de los *tracks* que almacenan las autoridades de transporte de numerosas ciudades. Google Maps ha hecho alianzas con autoridades de transporte locales en distintas ciudades. De esta manera, la app cuenta con información en tiempo real del transporte público y permite a sus usuarios conocer retrasos, tiempos de viaje y transbordos en la red de transporte público. Los ficheros GTFS pueden ser cargados en un SIG para analizar el nivel de cobertura de la red de transporte público, las velocidades medias y las superposiciones de las líneas (Hadas, 2013), así como para calcular matrices de tiempos de viaje según franjas horarias, que pueden ser utilizadas en estudios dinámicos de accesibilidad (Stepniak et al., 2019).

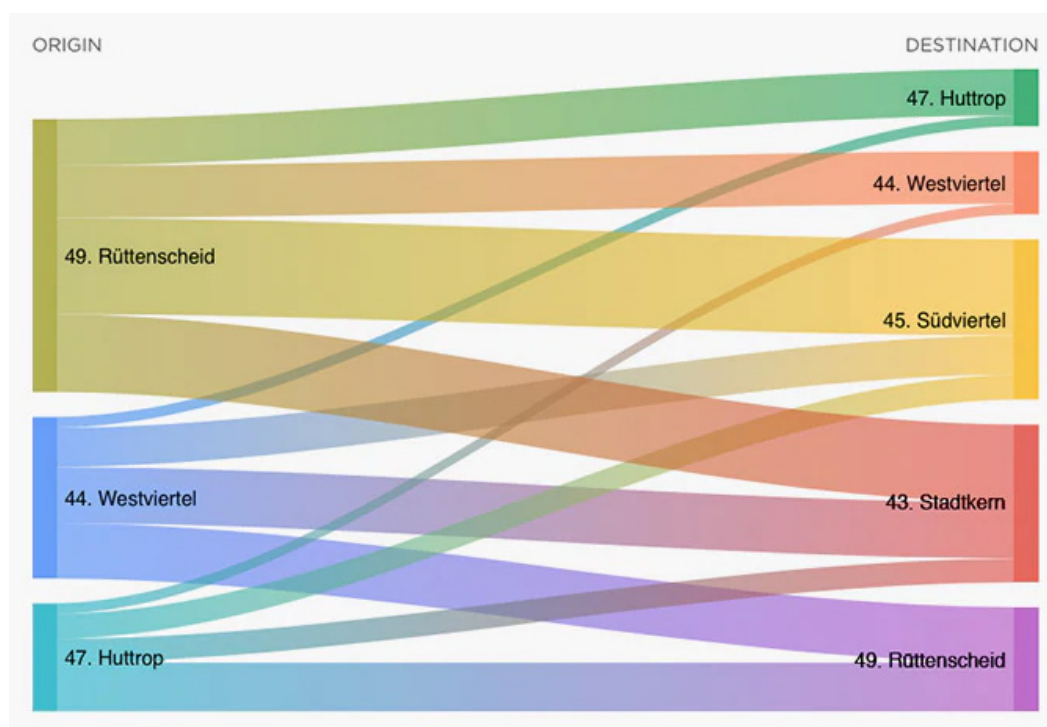


**Figura 10:** Evolución de los tiempos de viaje (en segundos) entre el distrito de negocios de Madrid (AZCA) y el aeropuerto obtenidos a partir de la API de Google Maps. La línea roja indica el tiempo de viaje medio en el conjunto del día. Fuente: García-Albertos et al., 2019

## TOMTOM

La información sobre redes de transporte privado ha mejorado decisivamente en los últimos años gracias a las compañías de navegadores para automóviles, particularmente Navteq y Teleatlas (ésta última adquirida por TomTom). Estas detalladas redes viales contienen información sobre las características del viario (tiempos de recorrido de cada tramo, direcciones prohibidas, prohibiciones de giros, etc.) para hacer posible el cálculo de rutas óptimas en los navegadores. Pero, además, pueden ser instaladas en un Sistema de Información Geográfica para realizar análisis de redes. Para este fin, es especialmente útil el producto TomTom Speed Profiles, una red digital de transporte privado que se puede utilizar en softwares como ArcGIS e incluye las velocidades de los vehículos en cada uno de los tramos de la red vial (excepto el viario local) cada cinco minutos. Estos datos permiten realizar análisis dinámicos de accesibilidad considerando el efecto de la congestión. Se trata de datos históricos obtenidos por distintos dispositivos, incluyendo los propios navegadores y los GPS de teléfonos móviles. Esta nueva fuente de datos ha sido poco explotada hasta la fecha, sin embargo, existen algunos estudios sobre congestión en ciudades europeas que utilizan TomTom Speed Profiles (por ejemplo, Moya-Gómez y García-Palomares, 2017).

Además, esta compañía ha desarrollado la API TomTom Move<sup>7</sup>, que incluye dos productos de interés para técnicos e investigadores del transporte. *Traffic Stats* proporciona información sobre la situación del tráfico en la red de carreteras de una ciudad durante todo el día. Los usuarios pueden crear una consulta personalizada para un área o ruta específica, para días específicos o para los períodos de tiempo en el rango de fechas que sean de su interés. Así, por ejemplo, se puede medir el tiempo de viaje antes y después de un cambio en la red de carreteras para evaluar su impacto o detectar el impacto de eventos e incidentes en la congestión. Por su parte, el producto O/D Analysis permite obtener matrices de viajes origen-destino (OD) en vehículo privado entre zonas de transporte, pudiendo diferenciar también entre días de la semana, horas del día y zona de transporte de paso. Tanto los datos de velocidades y tiempos de viaje, como las matrices OD en vehículo privado, se pueden utilizar para alimentar modelos de transporte.



**Figura 11:** Matriz de viajes OD presentada como gráfico de flujos. Fuente: <https://move.tomtom.com>

## STRAVA

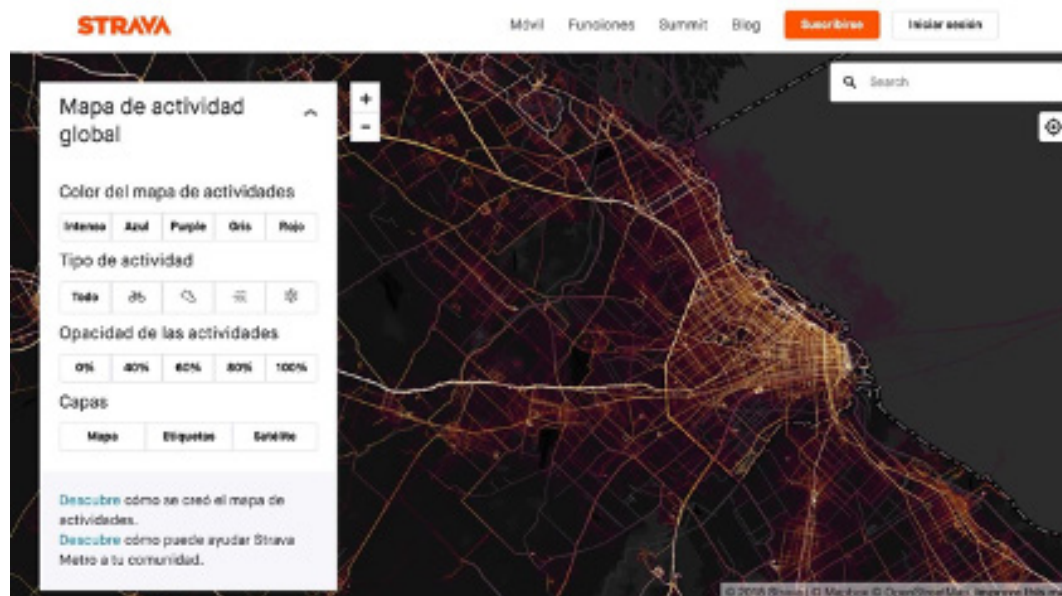
El volumen de *tracks* GPS registrados por compañías propietarias de aplicaciones esencialmente ligadas a actividades deportivas y de ocio ha experimentado un crecimiento exponencial en los últimos años. Su primer impulso vino con la llegada de los teléfonos inteligentes, que permitían la instalación de aplicaciones. La más reciente popularización de los denominados "*wearable devices*", como pulseras de actividad y relojes inteligentes, ha aumentado aún más su uso. Entre las compañías líderes podemos destacar Under Armour,

<sup>7</sup> <https://move.tomtom.com/>



que ofrece actualmente un conjunto de aplicaciones que suman más de 200 millones de usuarios, Runkeeper, que actualmente cuenta con más de 50 millones de usuarios, u otras como Strava, que recopiló 84 millones de rutas ciclistas en 2018, un 24 % más que el año anterior. Estas compañías registran importantes colecciones de datos que ofrecen, para su uso por terceros, de manera anonimizada o agregada.

Particularmente interesante es el caso de los datos registrados por la compañía Strava, a través del conjunto de aplicaciones que comercializa. La compañía lanzó en 2012 su plataforma Strava Metro, orientada a la explotación comercial de los datos registrados para la realización de estudios de movilidad, habiendo contribuido a la realización de más de 300 estudios de movilidad y planeamiento hasta la fecha.



**Figura 12:** Mapa de actividades registradas por Strava en Buenos Aires. Fuente: [www.strava.com/heatmap](http://www.strava.com/heatmap)

Entre la información que se puede obtener con base en sus datos, la compañía destaca la localización de los espacios y calles más frecuentados por corredores y de las arterias urbanas más importantes en cuanto a movilidad ciclista, así como la estimación del uso de infraestructuras ciclistas. Por otro lado, Strava Metro permite conocer el volumen de viajes ciclistas correspondientes a deporte, ocio o trabajo. Este último se incrementó en un 42% en el año 2019, algo relevante a considerar para los escépticos que aún creen que la muestra de este tipo de apps corresponde tan sólo con usuarios ligados al deporte o al ocio.



## 2.4 Campos de aplicación

Este apartado pretende realizar una síntesis de los campos de aplicación que se han ido adelantando en la descripción de cada una de las fuentes, como los análisis de presencia, la generación de matrices origen-destino de viajes o los estudios de percepción de los usuarios en el uso de las redes de transporte:

- **ANÁLISIS DE PRESENCIA DE LA POBLACIÓN Y USO DE LOS MODOS DE TRANSPORTE:** Las fuentes que tienen una gran frecuencia de uso permiten conocer la distribución de la población en los diferentes momentos temporales, ya sea a lo largo del día, la semana o durante la celebración de determinados eventos. La cartografía de presencia de población según momentos temporales es de gran utilidad en los estudios de transporte, en tanto que aportan datos de demanda. Para conocer esa distribución de la población según franjas horarias es posible usar tanto datos de localizaciones puntuales como *tracks*. Una fuente muy usada en los estudios de presencia es Twitter (ver apartado 2.3). En cuanto a los estudios de distribución espacial y temporal según modos de transporte, se pueden utilizar con este fin los datos del uso de tarjetas inteligentes de transporte (Gschwender et al., 2016) y los registros GPS tomados por los servicios de taxi o por los sistemas de bicicletas públicas.
- **OBTENCIÓN DE DATOS DE AFOROS DE VIAJES:** Los datos de *tracks* obtenidos por los navegadores de los coches o de diferentes aplicaciones móviles han sido usados para obtener el número de viajes que pasa por un determinado tramo de la red o por el conjunto de la red. Como hemos visto, productos como las API de TomTom Traffic Stats, Google Transit o Waze (apartado 2.3) proporcionan información sobre la situación del tráfico en la red de carreteras de una ciudad durante todo el día. También con datos de *tracks* del uso de bicicletas públicas es posible obtener los aforos de viajes de los tramos de la red, como los casos de Strava o aplicaciones como la realizada en Madrid a partir de datos del sistema de bicicletas públicas o de *tracks* de ciclistas voluntarios (ver Romanillos et al., 2018). *Calculo de matrices Origen-Destino de viajes:* La gran frecuencia de información que proporcionan las nuevas fuentes de datos es muy útil para obtener matrices OD de viajes. Nuevamente podemos obtener matrices de viajes tanto de datos puntuales como de datos de *tracks*. En esta nota técnica veremos la metodología de trabajo para ambos casos (capítulos 4 y 5, respectivamente). Twitter es una de las fuentes más usadas en el caso de datos puntuales. Otras fuentes de datos para la obtención de matrices OD son las tarjetas inteligentes de transporte público (ver Tao et al., 2014; Dewulf et al., 2015) o los sistemas de navegación de vehículos, como los datos de O/D Analysis de TomTom.
- **ESTUDIOS DE ACCESIBILIDAD:** Los datos obtenidos a partir de GPS también permiten realizar estudios de accesibilidad, con el objetivo de medir la facilidad para alcanzar los destinos de interés en un determinado modo de transporte. En este caso, la información de los *tracks* permite conocer tiempos y velocidades de viaje en los diferentes tramos de la red, y usar esos datos para el cálculo de la accesibilidad según franjas horarias. Así, los datos de TomTom se han utilizado para conocer los cambios temporales de las distribuciones de accesibilidad diaria en las grandes áreas metropolitanas europeas (ver Moya y García-Palomares, 2017), mientras que los *tracks* de

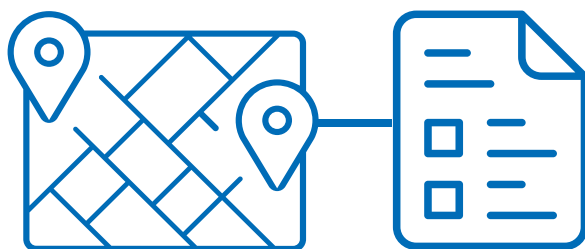
bicicletas públicas se han empleado en el análisis de la complementariedad y la competencia entre la bicicleta y el resto de los modos de transporte (Romanillos y Gutiérrez, 2019). Además, como hemos visto, la API de Google permite también obtener matrices de tiempos de viaje y utilizarlas en estudios de accesibilidad (como en García-Albertos et al., 2019).

- **IMPACTOS DE EVENTOS EN LA MOVILIDAD URBANA:** La celebración de determinados eventos en las ciudades produce un gran impacto en la movilidad. El empleo de fuentes que proporcionen datos geolocalizados a partir de GPS, a ser posible en tiempo real, es fundamental para la gestión de los mismos. Datos de redes sociales como Twitter se han empleado para la detección de eventos urbanos, por ejemplo, en los Países Bajos (Steur, 2014). Las tarjetas inteligentes de transporte también se han utilizado para la predicción de comportamientos del uso del transporte público durante la celebración de eventos especiales, como en el caso de la ciudad de Singapur (Pereira et al., 2015). Algunos trabajos han usado datos de GPS para analizar la vulnerabilidad a diferentes sucesos. En Wuhan (China) se ha construido una base de datos a partir de información obtenida de sistemas GPS del servicio de taxis para, mediante un matriz OD de viajes en una situación de normalidad, analizar patrones de movilidad cuando se producen cortes en alguno de los tres puentes principales que conectan los tres distritos de la ciudad separados por el río Yangtsé (Yue et al., 2012). Además, podemos usar los datos de Waze para estudiar todo tipo de eventos y la distribución de los accidentes y los diferentes eventos que se producen en la red de carreteras.
- **INFORMACIÓN PARA CONOCER LA PERCEPCIÓN DEL SERVICIO DE TRANSPORTE:** Finalmente, el uso de datos puntuales de redes sociales localizadas a partir de GPS permite explotar los comentarios que los usuarios realizan sobre la valoración de los sistemas de transporte. Se trata de, mediante un análisis del lenguaje, clasificar los sentimientos que transmiten y tratar de identificar por qué una línea particular de transporte público tiene un sentimiento positivo o negativo. Luong y Houston (2015) usaron este tipo de análisis para analizar el servicio de tren en la ciudad de Los Ángeles.



# 3

## Características de los Registros GPS



En este apartado se describen las características de los registros GPS, diferenciando entre datos puntuales, que muestran simplemente localizaciones (por ejemplo, accidentes de tráfico) y datos lineales (*tracks*), que muestran secuencias de puntos a partir de las cuales se pueden reconstruir líneas representativas de los viajes. Veremos los tipos de ficheros usados para almacenar estos datos y sus principales características.



### 3.1 Registros GPS: Tipos de ficheros

Cuando se trabaja con datos de registros GPS se manejan distintos tipos de ficheros que conviene conocer. En estos ficheros suelen destacar los formatos abiertos, que facilitan el intercambio de información entre los diferentes softwares. Los registros de datos puntuales, procedentes, por ejemplo, de redes sociales o de diferentes aplicaciones que usan el GPS de los dispositivos móviles suelen descargarse en tablas de texto tipo *csv*. En otros casos, los datos se descargan ya en formatos diseñados para la representación propia de datos geográficos, como son los ficheros GeoJSON. Cuando los datos provienen directamente de dispositivos GPS, los ficheros más usados son los *gpx*, que son ficheros de intercambio con los que trabajan los GPS. Otros ficheros de intercambio entre softwares son los *kml* y *kmz*. Todos estos datos son llevados, tratados y analizados en un Sistema de Información Geográfica (SIG). A continuación, se describen brevemente las características de cada uno de estos tipos de ficheros.

#### TABLAS DE TEXTO O .CSV

Los archivos *csv* son un tipo de documento de texto utilizado para representar datos en forma de tabla. Estos ficheros representan las columnas de la tabla separando los datos de cada columna por comas o puntos y comas, y las filas por saltos de línea. Esta estructura de datos permite almacenar grandes cantidades de datos de una forma muy sencilla, ligera y en un formato abierto, lo que facilita el intercambio y su descarga desde los diferentes espacios web.

#### FICHEROS TIPO GEOJSON

El formato GeoJSON es utilizado habitualmente en las aplicaciones cartográficas en internet. Es un formato de gran utilidad, pues permite el intercambio de datos geográficos de una manera rápida, ligera y sencilla, utilizando un formato abierto. Los ficheros GeoJSON son ficheros con estructura de SIG, ya que están diseñados para representar, por un lado, elementos geográficos sencillos y,

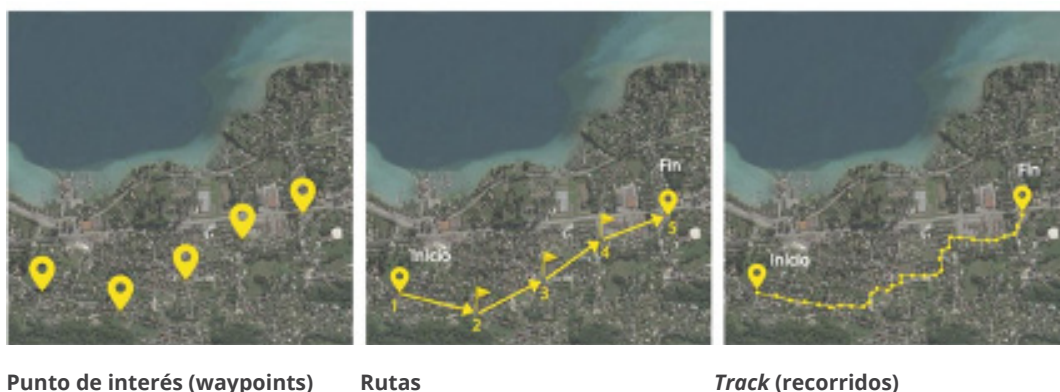


por otro, sus atributos no espaciales. Habitualmente, los ficheros GeoJSON almacenan datos de tipo puntual, pero pueden trabajar también con datos en forma de líneas y de polígonos. Al igual que los datos GPS, los ficheros GeoJSON usan como sistema de coordenadas de referencia el *World Geodetic System 1984* (WGS84) y trabajan con unidades en grados decimales para georreferenciar los datos. Puede verse más información en <https://geojson.org>.

## GPX

Los ficheros *gpx* son los formatos de intercambio de la tecnología GPS. Están pensados para transferir los datos obtenidos por GPS a otro tipo de aplicaciones. Los ficheros *gpx* permiten el intercambio de datos tomados de forma puntual (*waypoints*) y de forma lineal a través de rutas (*routes*) o recorridos (*tracks*) (Figura 13).

- Los puntos de interés (*waypoints*) o localizaciones puntuales son elementos puntuales definidos por una posición geográfica (latitud y longitud: x y) y, en la mayoría de los casos, por la altitud (z). Los GPS suelen almacenar estas localizaciones en ficheros con extensión *.wpt* y después son exportados a *gpx*.
- Una ruta es un camino formado por localizaciones puntuales o puntos de interés que han sido introducidos en un orden predefinido (con un origen y un destino final) y que permite que el GPS nos dirija entre ellos a través de una cartografía de referencia, mediante instrucciones visuales o sonoras, como sigue recto, gira a la derecha, pasa el objeto X, etc. Las rutas se almacenan en ficheros con extensión *.rte*, y, al igual que los puntos de interés, son exportadas después en *gpx*.
- Finalmente, un *track* (recorrido) es la unión ordenada de una serie de puntos para los que se han tomado coordenadas geográficas y que definen un recorrido realizado por el GPS o diseñado sobre un mapa. Los *tracks* se almacenan en ficheros *trk*, que son exportados posteriormente a *gpx*.



**Figura 13:** Ejemplo de datos recogidos con GPS. Fuente: elaboración propia

### OTROS FICHEROS DE INTERCAMBIO: *KML*, *KMZ*, *LOC* ETC.

Los datos de GPS, tanto en formato de localizaciones puntuales, rutas o *tracks*, pueden intercambiarse utilizando otro tipo de ficheros como son, por ejemplo, los *kml* y *kmz* o los *loc*.

Los ficheros *kml* están basados en el lenguaje XML y se usan para representar datos geográficos en entornos web. Los ficheros *kml* son los utilizados, por ejemplo, en el software Google Earth para especificar características de los lugares. Para ello, contienen información de un título, una descripción básica del lugar, sus coordenadas de longitud y latitud y, opcionalmente, la altitud. Dado el gran uso de las aplicaciones con Google Earth, los ficheros *kml* han adquirido un estatus de estándar abierto como formato de intercambio de información geográfica. Para este intercambio, los ficheros *kml* a menudo suelen distribuirse comprimidos como ficheros *kmz*, que, además, pueden incluir otro tipo de archivos (como imágenes, textos, etc.) y otros recursos asociados a la cartografía de los datos.

Algunos dispositivos GPS trabajan con otros formatos de intercambio, como el tipo *loc*. Estos son ficheros similares a los *gpx*, de manera que se trata de archivos que permiten guardar las localizaciones, ya sean puntuales, de rutas o *tracks*.

### FICHEROS GIS: SHP, GEODATABASE U OTROS

Los ficheros anteriores son importados desde un SIG para su tratamiento, análisis y visualización. Se utilizan diferentes tipos de ficheros en función del software. Entre los más habituales están los ficheros *shp* (shapefile), formato desarrollado por la compañía ESRI para su software ArcVIEW y que, por lo sencillo de su estructura, se ha convertido en un estándar dentro de los SIG, siendo muy utilizado por otros software comerciales o libres. La compañía ESRI ha desarrollado otro tipo de ficheros también muy usados, como las coberturas de *ArcInfo* o las *geodatabase* de ArcGIS. Otros softwares comerciales, como MapInfo y Grass, y otros de código libre, como qGIS, trabajan, también, con otros formatos, en general muy similares entre sí.



## 3.2 Información contenida en los registros GPS

Los registros GPS permiten conocer información sobre localizaciones geográficas de puntos de interés, rutas o *tracks*, a los que se asocia otra información muy útil a la hora de usar estos datos en estudios de movilidad. Los registros más completos son los referentes a *tracks* (recorridos). Los datos asociados a un *track* incluyen:

- **PROPIEDADES GENERALES DEL FICHERO:** Puede incluir información como el nombre del *track*, un nombre corto o identificador, descripción del recorrido, identificador del usuario que ha generado el recorrido u otra información asociada a ese usuario. En algunos casos contiene también información sobre el dispositivo GPS que grabó el *track*.

- **INFORMACIÓN GENERAL SOBRE LA ESTRUCTURA DEL TRACK:** Incluye información sobre el número de puntos que definen el *track*, el intervalo de grabación de puntos (resolución temporal), campos de datos adicionales incluidos en el *track*, etc.
- **COORDENADAS GEOGRÁFICAS DE LOS PUNTOS QUE DEFINEN EL TRACK:** Están tomadas en grados decimales y usando el sistema de proyección WGS84 e incluyen la longitud (x) y latitud (y) de cada uno de los puntos que definen el recorrido realizado.
- **VALORES DE ALTITUD DE LOS PUNTOS QUE DEFINEN EL TRACK:** Al igual que las coordenadas de latitud y longitud, los dispositivos GPS almacenan habitualmente los valores de altitud (z) de cada uno de los puntos que define el track.
- **MOMENTOS TEMPORALES:** Se almacena la fecha, hora, minuto y segundo en el que ha sido tomado cada uno de los puntos que define el track.

A partir de estos datos, los propios dispositivos GPS o el SIG permiten obtener información derivada que resulta de gran utilidad en los estudios de movilidad:

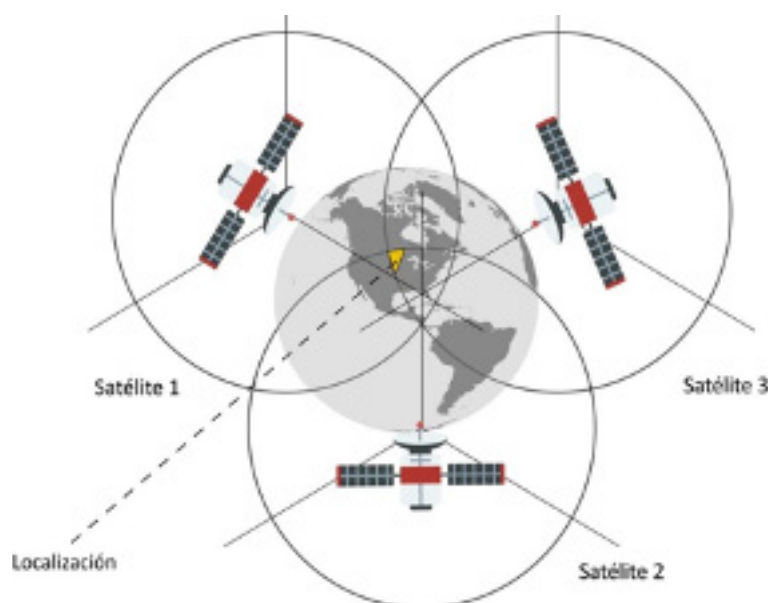
- **MOMENTOS TEMPORALES:** A partir de la fecha y hora se pueden precisar la hora de salida y llegada del recorrido, su duración, los tiempos en movimiento y en parado, etc.
- **DISTANCIA:** La mayoría de los dispositivos GPS generan diferentes campos de distancias, como la distancia proyectada (distancia recorrida horizontalmente), la distancia con altitudes (distancia real recorrida considerando la orografía del terreno), la distancia lineal (línea recta entre el primer y el último punto del track), etc.
- **VELOCIDADES:** Los propios dispositivos GPS generan información de velocidades durante el recorrido realizado, incluyendo valores de velocidad máxima, velocidad media, ritmo, velocidad vertical, etc.
- **INFORMACIÓN SOBRE ALTITUD Y PENDIENTES:** Puede ser de gran interés en el estudio de la movilidad peatonal o en bicicleta. Podemos obtener información sobre la altitud máxima/mínima, la pendiente (media, máxima, mínima), pendientes en cada uno de los tramos, distancias en ascenso o descenso acumuladas, etc.
- **ANÁLISIS DE TRAMOS:** Una de las ventajas de los track es que toda la información anterior puede obtenerse y analizarse para cada una de las secciones que forman el track, ya sea por tiempo, distancia, velocidades, pendientes, etc.



### 3.3 Resolución espacial de los datos: exactitud posicional

Al capturar localizaciones geográficas, los GPS siempre tienen un pequeño error locacional, normalmente en torno a 2-4 metros, que determina lo que se conoce como exactitud posicional de los datos.

Para determinar las localizaciones geográficas en el globo terráqueo, el sistema GPS se sirve de un mínimo de 3 satélites y un método de trilateración (Figura 14), aunque lo normal es que sean necesarios 4 o más satélites para determinar la posición con cierta precisión. La red de satélites del sistema GPS es variable, pero incluye un mínimo de 24 satélites (actualmente tiene hasta 31) que orbitan a distancias que varían entre los 20 y los 180 km de altura y con trayectorias sincronizadas para cubrir toda la superficie de la Tierra (<https://www.gps.gov/systems/gps/space/>). Cuando un receptor GPS quiere determinar una posición geográfica, localiza ese mínimo de satélites de la red, de los que recibe unas señales indicando la identificación y hora del reloj de cada uno de ellos. A partir de estas señales, el aparato sincroniza el reloj del GPS, calcula el tiempo que tardan en llegar las señales desde cada satélite al equipo y mide la distancia al satélite. Conocidas las distancias, se determina fácilmente la posición relativa respecto a los satélites. Conociendo, además, las coordenadas o posición de cada uno de los satélites, se obtiene la posición absoluta o coordenadas reales del receptor GPS.



**Figura 14:** Método de trilateración para determinar la coordenada geográfica de una localización en el globo a partir de los satélites del sistema GPS. Fuente: Elaboración propia

La precisión en la localización geográfica de los datos GPS depende del número de satélites con los que está trabajando el aparato (los satélites “visibles” en ese momento), la posición respecto a él en la que están (cuanto más dispersos mejor) y la calidad de recepción de las señales. Así, si se capta la señal de entre siete y nueve satélites de referencia y, si estos tienen una distribución

adecuada, (están dispersos), las precisiones que pueden obtenerse son inferiores a 2,5 metros en el 95% de los casos. Estas precisiones pueden ser mejoradas en función de las diferentes tecnologías de los sistemas GPS, consiguiéndose en muchos casos errores inferiores a 0,5 metros (<https://www.gps.gov>). En el apartado 3.5 se profundiza en estos aspectos y se mencionan algunas limitaciones de la tecnología GPS en relación con la precisión en la localización.

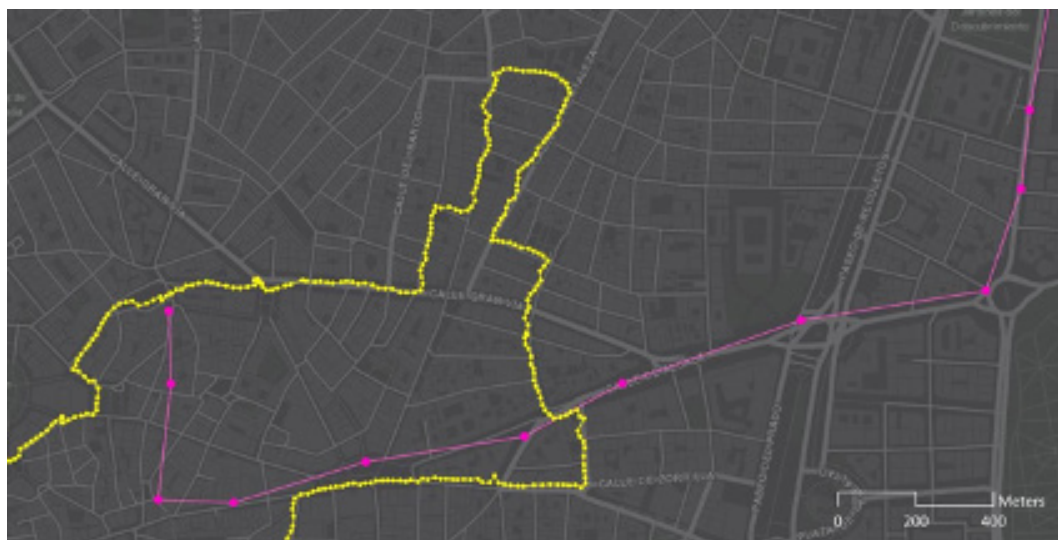


### 3.4 Resolución temporal de los datos

La resolución temporal hace referencia al tiempo transcurrido entre las tomas de las localizaciones de los datos tomados por registros GPS. Cuando hablamos de resolución temporal de los datos podemos hacer referencia a los intervalos de tiempo con los que se ha recogido información por parte de los usuarios. En este caso, la resolución temporal depende de la frecuencia en esas acciones de los usuarios, por ejemplo, la frecuencia con la que se envían tweets geolocalizados (ya sean usuarios compulsivos, frecuentes o esporádicos). En redes sociales de uso frecuente, por ejemplo Twitter, la resolución es mayor que en otras con menos frecuencia de uso, como las de fotografías compartidas.

En los datos de rutas o *tracks* hablamos de resolución temporal para hacer referencia a los intervalos de tiempo regulares para los que se obtiene (de forma automática) la localización de los puntos que definen esa ruta o ese *track*. En este caso, la resolución temporal depende de la configuración que se haya establecido en el receptor GPS para la recogida de los datos. La mayor parte de los sistemas, por ejemplo, los GPS de los teléfonos móviles, tienen una resolución temporal de 2 segundos, pero estos intervalos pueden variar. En general, las resoluciones son más bajas cuando se almacenan datos de forma masiva, por ejemplo, de recorridos de una flota de vehículos. La Figura 15 muestra un track de un ciclista compartido en la plataforma [www.huellaciclistademadrid.es](http://www.huellaciclistademadrid.es) con una resolución temporal de 2 segundos, y un track correspondiente a una ruta registrada por BiciMAD, el Sistema de Bicicleta Pública de Madrid, con una resolución media de 75 segundos.





**Figura 15:** Ejemplos de *tracks* con diferentes resoluciones temporales. *Tracks* con resolución temporal de 2 segundos (amarillo) y 75 segundos (magenta). Fuente: Elaboración propia



### 3.5 Limitaciones de los estudios basados en datos GPS

Como hemos visto, para la correcta geolocalización de un receptor GPS es necesario que al menos 4 satélites mantengan con dicho receptor una “línea de vista” (line of sight) sin obstrucciones. Sin embargo, el sistema de localización GPS tiene aún grandes limitaciones a la hora de abordar estudios de movilidad y transporte. A continuación, describiremos de forma básica las principales limitaciones y cómo estas se están tratando de superar.

#### LIMITACIONES EN TRANSPORTE SUBTERRÁNEO.

Durante el último siglo, el transporte subterráneo ha pasado de ser algo excepcional a tener una gran relevancia, particularmente en las grandes ciudades, donde los sistemas de metro subterráneo o las calles y carreteras soterradas han crecido de manera exponencial.

La tecnología GPS estándar no funciona bajo tierra, por lo que los vehículos que circulan por las redes subterráneas no pueden utilizarla para su geolocalización. Muchos vehículos que normalmente circulan sobre la superficie, como las flotas de autobuses o los vehículos privados, suelen usar sistemas GPS, por lo que su geolocalización se pierde cuando se adentran en túneles o estaciones subterráneas. Los problemas derivados pueden ser la desubicación de enrutadores, que dejan de guiar a los conductores, o la pérdida de los sistemas de información en tiempo real que suelen incorporar la mayor parte de las flotas modernas de autobuses.

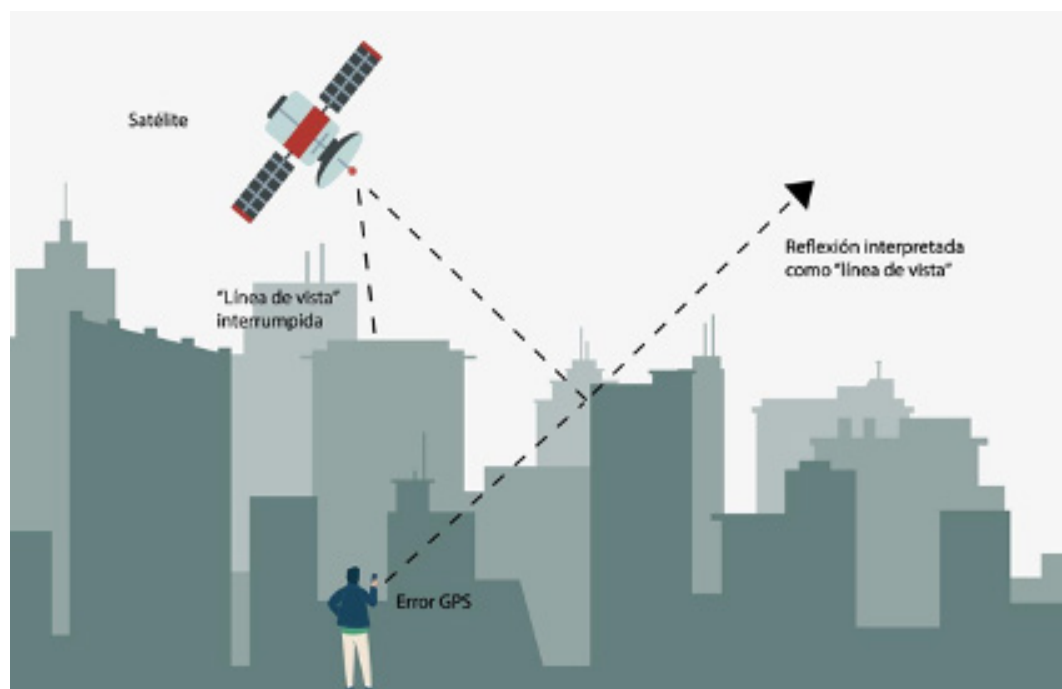
Para hacer frente a este problema, algunas ciudades están implementando sistemas de repetidores de señal GPS en determinados tramos de vía o terminales subterráneas. Ya hay disponibles sistemas comerciales de

repetidores de señal satelital tanto para GPS como para GLOANSS, Galileo o Beidou.

#### LIMITACIONES EN ENTORNOS URBANOS DENSOS.

Cuando un dispositivo GPS se desplaza en entornos urbanos densos, la conexión con la red de satélites puede obstaculizarse fácilmente, dando lugar a pérdidas importantes de precisión en la geolocalización. Para muchos de los usuarios de dispositivos GPS, interesados en registrar sus rutas, esto puede no suponer un problema relevante. Sin embargo, esta falta de precisión puede ser bastante problemática para servicios basados en la geolocalización exacta de personas o vehículos. En un entorno urbano denso, las personas que hacen uso de los servicios de movilidad compartida pueden no encontrar los vehículos en los lugares exactos marcados por los mapas de sus aplicaciones, o los conductores de Uber u otras compañías pueden, igualmente, tener problemas para dar con algún usuario.

Para resolver este problema, grandes empresas tecnológicas, como Uber, han desarrollado sistemas de mejora de la geolocalización basados en algoritmos que corrigen la posición con base en el conocimiento de la posición de los satélites que están en la "línea de vista" (Hawkins, 2018). También se ha trabajado recientemente en el desarrollo de sistemas colaborativos de geolocalización de vehículos, como el sistema CoDrive (Demetriou et al., 2018), que utiliza como fuentes correctoras de geolocalización los datos obtenidos de vehículos que poseen sistemas de navegación más precisos que el GPS (u otros sistemas satelitales), como los que poseen los prototipos de vehículos autónomos.



**Figura 16:** Error típico de geolocalización por GPS en entornos urbanos densos. Fuente: elaboración propia

## LIMITACIONES EN CUANTO A LA PRECISIÓN DE LA LOCALIZACIÓN.

La precisión de la geolocalización mediante tecnología GPS depende de múltiples factores. Dejando a un lado las condiciones relativas al entorno físico, que hemos visto anteriormente, la precisión depende también, entre otros factores, de la localización de los satélites en un momento preciso, de las condiciones atmosféricas, así como de la calidad del sistema receptor de la señal satelital. En un supuesto ideal de localización en un espacio libre de interferencias y sin depender de la calidad del receptor, el sistema GPS llega a garantizar la localización con una precisión de 0,715 metros en un 95% de los casos (cifra denominada URE: Global User Range Error). Sin embargo, para los receptores GPS normalmente integrados en los teléfonos inteligentes, la precisión óptima puede estar en torno a 4,9 metros ([www.gps.gov](http://www.gps.gov)).

Esta precisión, menor en muchos casos en los entornos urbanos densos, puede suponer también una limitación en algunos estudios de movilidad o transporte. Por ejemplo, a partir de un track GPS no podemos conocer con seguridad si un vehículo circula por un carril u otro en una vía de múltiples carriles. No podremos saber si un ciclista está utilizando un carril bici o, por el contrario, circula por la calzada o por la acera.

Nuevos sistemas de posicionamiento por satélite, como el europeo Galileo, tienen, en principio, una precisión de localización mucho mayor que la actual del GPS, de hasta 20 cm en entornos libres de obstáculos. Sistemas basados en Galileo podrían, por tanto, resolver algunos de los problemas actuales de imprecisión, como los citados anteriormente. No obstante, ni siquiera estos sistemas tendrán una precisión suficiente para algunos de los retos más inmediatos, como la geolocalización necesaria para la navegación de vehículos autónomos. Estos funcionan ajustando su trayectoria a través de sistemas de navegación que utilizan múltiples sensores de visión artificial basados en ultrasonidos, láser o en interpretación de imagen de video, así como en sensores de inercia, como los INS - Inertial Navigation Systems (Sistemas de Navegación Inerciales).

## LIMITACIONES EN INTERIORES.

Finalmente, debemos señalar también la limitación que la tecnología GPS tiene en espacios interiores para el análisis de la movilidad o el transporte. Esencialmente, el sistema GPS no funciona en interiores, debido a la obstrucción de las líneas de vista de los satélites.

La geolocalización en espacios interiores está teniendo una importancia cada vez mayor, especialmente en campos como la industria y la logística, que están viviendo una auténtica revolución debido al desarrollo de nuevos sistemas de producción y distribución.

La geolocalización en los interiores es más complicada debido a su susceptibilidad frente a los ruidos y rebotes de las señales electromagnéticas (en las que influyen tanto los muros como el mobiliario), los dispositivos electrónicos e incluso, en ocasiones, la presencia de seres humanos. En estos espacios, los sistemas de posicionamiento por satélite son sustituidos por los denominados IPS -Indoors Positioning Systems (Sistemas de Posicionamiento en Interiores), basados en distintas tecnologías (Regus et al., 2019), como infrarrojos (IR), Wireless Local Area Network - WLAN (Redes Inalámbricas Locales), ultrasonidos, Radio Frecuencia (RFID), Bluetooth, Ultra-wideband -UWB (Banda Ultra Ancha) y otros.



# 4

## **Análisis de Movilidad a partir de Datos de Localización Puntuales**



Diferentes fuentes proporcionan datos de localización puntual geolocalizada a partir de GPS que pueden ser usados en estudios de movilidad. En este apartado nos vamos a centrar en la información proporcionada por las redes sociales. Conociendo la localización cambiante de los usuarios de estas redes sociales, a partir de su huella digital, se pueden analizar sus pautas de movilidad.

Los datos de redes sociales, como en general la mayor parte de los datos asociados al Big Data, no son generados con el fin de analizar la movilidad urbana. En consecuencia, antes de utilizar datos de localización puntuales procedentes de redes sociales es necesario tener muy presente la importancia de los procesos de limpieza y pre-proceso de datos. La base de datos con la que se va a trabajar tiene que contener únicamente datos con los que sea posible obtener información fiable sobre la movilidad de los usuarios, seleccionando una muestra de usuarios válidos para el análisis y eliminando registros que puedan producir ruido en los resultados finales.

Es importante tener siempre en cuenta las limitaciones de estas fuentes de datos. Las redes sociales presentan, como vimos, una resolución temporal baja (entendida como el tiempo medio entre cada actividad realizada por cada usuario). Esta baja resolución dificulta, por ejemplo, obtener datos de lugares de las actividades distintas a la residencia y el lugar de trabajo o estudio. La baja resolución temporal obliga también a trabajar con datos de periodos temporales amplios. Cuando se usan datos de telefonía en los estudios de movilidad es posible trabajar con bases de datos recogidas en periodos cortos (un mes o, incluso, una semana), mientras que con redes sociales necesitamos muestras de periodos mucho mayores (de más de uno o dos años). Por otro lado, mientras que los datos de telefonía móvil se refieren a muestras muy grandes de usuarios, en los datos de redes sociales las muestras son menores y pueden presentar mayores sesgos, al concentrarse los usuarios de estas redes sociales en determinados grupos sociodemográficos.

Finalmente, el uso de datos de redes sociales en estudios de movilidad requiere de un proceso de validación de los resultados. En ese sentido, es importante validar resultados a distintos niveles de agregación espacial, con el objetivo de ver hasta qué nivel podemos llegar en la obtención de información de movilidad, como por ejemplo matrices origen-destino de viajes.

En cualquier caso, los datos puntuales procedentes de redes sociales presentan algunas ventajas sobre otras fuentes de datos, como los registros de telefonía móvil. Entre otras, la facilidad de acceso, pues mientras las compañías de telefonía móvil son reacias a ceder sus datos, y cuando lo hacen es con un



coste muy elevado, la mayor parte de los datos de redes sociales pueden obtenerse de forma gratuita. Por eso, los datos de redes sociales se presentan como una alternativa a los datos de telefonía a la hora de estudiar la movilidad y crear matrices origen-destino de viajes. Una segunda ventaja es la mayor precisión en la localización de los usuarios, lo que, mediante el cruce de los mismos con los de otras fuentes de datos geolocalizados, facilita las tareas de enriquecimiento de datos.



## 4.1 Metodología a seguir para el cálculo de Matrices OD de viajes a partir de datos de localización puntual

La metodología para obtener información de movilidad y matrices de viajes a partir de datos puntuales de GPS, como los obtenidos de redes sociales como Twitter, en líneas generales es similar a la usada con datos de telefonía (Figura 17):

- Una primera fase debe estar dedicada a la recogida de los datos y su preparación para los análisis posteriores. Se incluyen aquí las tareas de descarga de datos, limpieza y filtrado de usuarios no válidos, selección de la muestra final y, en su caso, las tareas de enriquecimiento de la información de la red social con otros datos, como información de los usos del suelo o la unidad administrativa sobre las que se ha realizado el tweet.
- En una segunda fase es necesario definir el lugar de residencia de los usuarios y los lugares donde realizan las actividades frecuentes (normalmente, el lugar de trabajo o estudio).
- En la tercera fase, a partir del lugar de residencia y de actividades frecuentes, pueden obtenerse matrices de viajes con distintos niveles de agregación espacial (zonas de transporte, municipios, áreas metropolitanas, etc.). Estas matrices de viajes pueden ser posteriormente expandidas.
- Finalmente, es necesaria una etapa de verificación de los resultados que permita validar el grado de fiabilidad de los datos obtenidos y las posibles limitaciones de los datos.



**Figura 17:** Metodología a seguir para el cálculo de matrices de viajes OD a partir de Twitter, adaptada de Osorio-Arjona y García Palomares, 2019

## RECOGIDA, LIMPIEZA Y ENRIQUECIMIENTO DE LA BASE DE DATOS

Los datos de redes sociales son recogidos a través de sus interfaces de programación, las conocidas como APIs (*Application Programming Interface*). En el caso de Twitter, por ejemplo, la descarga se hace en streaming (de forma continua) y usando un código de programación en Python (ver Anexo). En todo caso, sea cual sea el lenguaje de programación utilizado, en estas operaciones se definen los parámetros básicos para la descarga. Es necesario, por un lado, establecer el espacio geográfico de referencia (según queremos descargar la actividad de la red social en un país, región o ciudad concreta) y, por otro, seleccionar los campos de información que queremos incluir en los datos descargados. Datos sobre el identificador del usuario (anónimo), las coordenadas geográficas y el momento temporal, serán necesarios para después obtener información sobre movilidad.

Una vez se inicia el proceso de descarga, los datos son recopilados normalmente en un gestor de base de datos. Hay distintas herramientas para ello, una habitual es, por ejemplo, MongoDB. La base de datos descargada se suele importar en un SIG, en el que se representan los datos puntuales según sus coordenadas geográficas.

Después de almacenar los datos ya geolocalizados, se procede a filtrarlos para obtener una muestra de usuarios válidos. Tomando como referencia la información de Twitter, entre los procesos de filtrado más habituales están los siguientes (Osorio-Arjona y García Palomares, 2019):

- Selección de tweets publicados en día laborable, eliminando normalmente los viernes por la tarde (en horas posteriores a las 14:00), fines de semana y días festivos.
- Limpieza de tweets producidos desde cuentas bots, de publicidad o usuarios compulsivos sin movilidad. En las redes sociales hay mucha actividad que se genera desde máquinas, empresas o servicios y que, por tanto, carecen de interés en los estudios de movilidad. Para realizar esta limpieza se eliminan aquellas cuentas de usuarios con un volumen alto de tweets generados desde una misma localización.
- Filtrado de usuarios con una actividad menor, pero cuyos tweets tienen siempre localizaciones muy similares. Para ello se miden las distancias entre los tweets de cada usuario y se eliminan aquellos usuarios que, por ejemplo, tienen una distancia media menor a 100 metros en la localización de todos sus tweets, puesto que son usuarios que no cambian de localización durante el periodo de datos recogido.
- Filtrado de usuarios con una variación temporal reducida. En este caso se eliminan aquellos usuarios que tienen todos sus mensajes concentrados en un periodo inferior, por ejemplo, a dos semanas seguidas. En muchos casos se trata de posibles visitantes en la ciudad como turistas y que tienen una movilidad diferente a la de los residentes. Cuando el objetivo del estudio es conocer la movilidad de los turistas, esos usuarios constituirán la muestra para el análisis.
- Finalmente, de los usuarios restantes se seleccionan aquellos que han publicado un número mínimo de tweets durante el periodo analizado. De esta forma se trabaja con usuarios activos, que han dejado una huella lo suficientemente amplia para conocer sus espacios frecuentes. Es importante insistir en que los usuarios son identificados por identificadores anónimos y sus datos de geolocalización agregados espacial y temporalmente, de forma que se respeta el anonimato de los mismos durante todo el proceso.

Una vez realizada la limpieza, e identificados los usuarios por sus IDs, es posible y recomendable ampliar los datos recogidos para los mismos. En redes sociales como Twitter, existe la posibilidad de ampliar el número de tweets de esos usuarios, descargando los últimos mensajes publicados por cada uno de ellos. El objetivo es aumentar la resolución espacial y temporal de la base de datos. Esta ampliación es también gratuita y permite descargar los últimos 3.200 tweets de cada usuario con el objetivo de obtener mensajes geolocalizados que no se hubiesen captado en *streaming*.

Finalmente, en el SIG es posible realizar tareas de enriquecimiento de los datos. Este proceso de enriquecimiento permite añadir nueva información a la base de datos, con la que se podrán realizar análisis posteriores. Es posible, así, incorporar información sobre el uso del suelo en el que se localizan los registros puntuales. Los datos puntuales localizados en espacios residenciales pueden ayudar a identificar el lugar de residencia de esos usuarios, mientras que los localizados en espacios de oficinas u otro tipo de actividades pueden vincularse al lugar de trabajo. En otros casos, una vez identificados los lugares de residencia de los usuarios, se pueden asociar características sociodemográficas de esos lugares a esos usuarios. Es lo que hicieron, por ejemplo, Netto et al. (2015) para determinar la clase social de los usuarios de Twitter en Rio de Janeiro y conocer cómo se mueven los distintos grupos sociales por la ciudad.



Figura 18: Ejemplo de la detección del hogar a partir de la localización de los tweets nocturnos y el parcelario residencial. Fuente: Osorio-Arjona y García Palomares, 2019

### IDENTIFICACIÓN DEL LUGAR DE RESIDENCIA Y LUGAR DE TRABAJO.

Una vez limpiada y enriquecida la base de datos, es posible identificar el lugar de residencia y trabajo o estudio de cada uno de los usuarios. La actividad de los usuarios (los tweets en el caso de Twitter) se diferencia según sea realizada en horario diurno (relacionada con el trabajo) o en horario de noche (vinculada al hogar). Para ello se utilizan los horarios habituales en el lugar donde se realice el estudio. Estos pueden ser, por ejemplo, los comprendidos entre las 08:30 y las 17:30 para el horario de trabajo y los restantes clasificados como nocturnos.

Para depurar los posibles usuarios que escriben desde lugares de trabajo o zonas de ocio en esa franja horaria nocturna, se pueden cruzar las localizaciones obtenidas con la distribución de los usos del suelo dominantes, y seleccionar los tweets que se encuentran en parcelas con un uso mayoritario de suelo residencial. Y viceversa, a la hora de determinar la localización de los lugares de trabajo o estudio se cruzan los tweets enviados durante día con los datos de uso del suelo donde el uso predominante está vinculado al trabajo.

Si la base de datos es suficientemente grande y se cuenta con una muestra amplia de usuarios activos, es posible tratar de identificar otro tipo de actividades. Por ejemplo, se puede trabajar con actividad en horario lectivo e información de espacios universitarios para identificar la movilidad por motivos de estudio. En otros horarios, por ejemplo, de tarde, y en espacios vinculados a actividades comerciales o de ocio, se puede asociar los viajes a ese tipo de motivos. Eso es lo que hicieron, por ejemplo, Salas-Olmedo y Rojas (2017) para conocer la movilidad en diferentes espacios públicos en la ciudad de Concepción (Chile).

### **CALCULO DE MATRICES ORIGEN-DESTINO Y EXPANSIÓN DE LOS DATOS**

Una vez definidas las localizaciones del lugar de residencia y del trabajo, y utilizando los IDs de los usuarios, es posible calcular matrices de relaciones entre el lugar de residencia (origen) y el lugar de trabajo (destino). El resultado es una matriz de relaciones según las unidades espaciales consideradas (distritos, municipios o zonas metropolitanas).

La siguiente fase es la expansión de las matrices. Para ello, se calcula un factor de expansión para cada unidad espacial. Es posible considerar distintos factores de expansión. Lo más habitual es expandir los datos a partir del lugar de residencia, obteniendo un cociente entre la muestra de usuarios residentes identificados con la red social y el total de habitantes en esa unidad espacial. Es frecuente trabajar con la población residente en un cierto rango de edad, por ejemplo, de 19 a 65 años. Otros factores de expansión pueden ser calculados a partir de los datos de lugar de trabajo o estudio (destinos). Aquí, el valor de expansión de cada unidad espacial se obtiene dividiendo el número de trabajadores identificados a través de la red social entre el número de trabajadores según los datos oficiales. Este segundo factor es usado, normalmente, cuando lo que se está estudiando es la movilidad atraída por un determinado espacio, por ejemplo, los viajes atraídos por centros universitarios o por determinadas zonas de actividad.

A partir de los datos expandidos, las matrices de relaciones entre distritos y municipios pueden representarse a esa escala o agregadas en unidades espaciales de mayor tamaño (Figura 19).



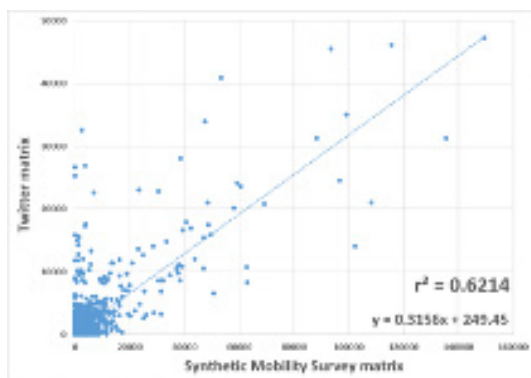


**Figura 19:** Ejemplo de matrices de viajes residencia-trabajo, según municipios y distritos (izquierda) y grandes zonas metropolitanas (derecha). Obtenidas a partir de datos de Twitter en el área metropolitana de Madrid. Fuente: Osorio y García-Palomares, 2019

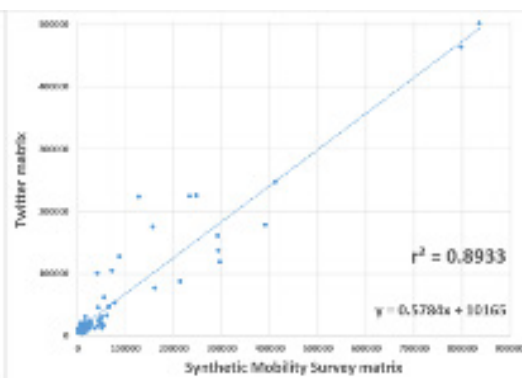
## VERIFICACIÓN DE LOS DATOS

Tras la obtención de la información de movilidad, es conveniente realizar una tarea de verificación de los resultados obtenidos. Para ello, es posible realizar comparaciones entre nuestros datos y los obtenidos a partir de métodos tradicionales. Un proceso sencillo es utilizar correlaciones bivariadas entre las matrices obtenidas, tanto a nivel de distrito y municipio como a nivel agregado, con las distribuciones de viajes de encuestas de movilidad. Este método, habitual en los estudios de movilidad a partir de datos de telefonía, permite conocer la bondad de ajuste entre nuestros datos y los datos de las encuestas domiciliarias. Además, es posible cartografiar los residuos de las diferentes regresiones para visualizar las relaciones que presentan las mayores desviaciones entre las matrices obtenidas y las procedentes de encuestas.

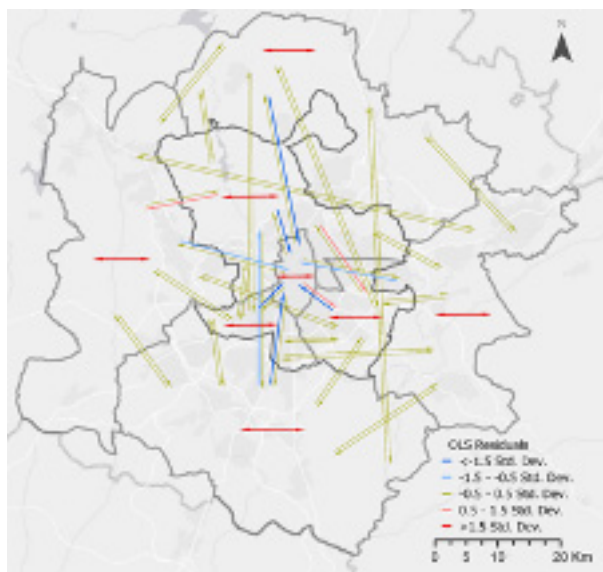
### Datos a nivel de distrito y municipio



### Datos a nivel de zonas metropolitanas



**Figura 20:** Ejemplo de correlaciones bivariadas entre las matrices obtenidas con Twitter y las obtenidas a partir de encuestas domiciliarias en la ciudad de Madrid. Fuente: Osorio y García-Palomares, 2019



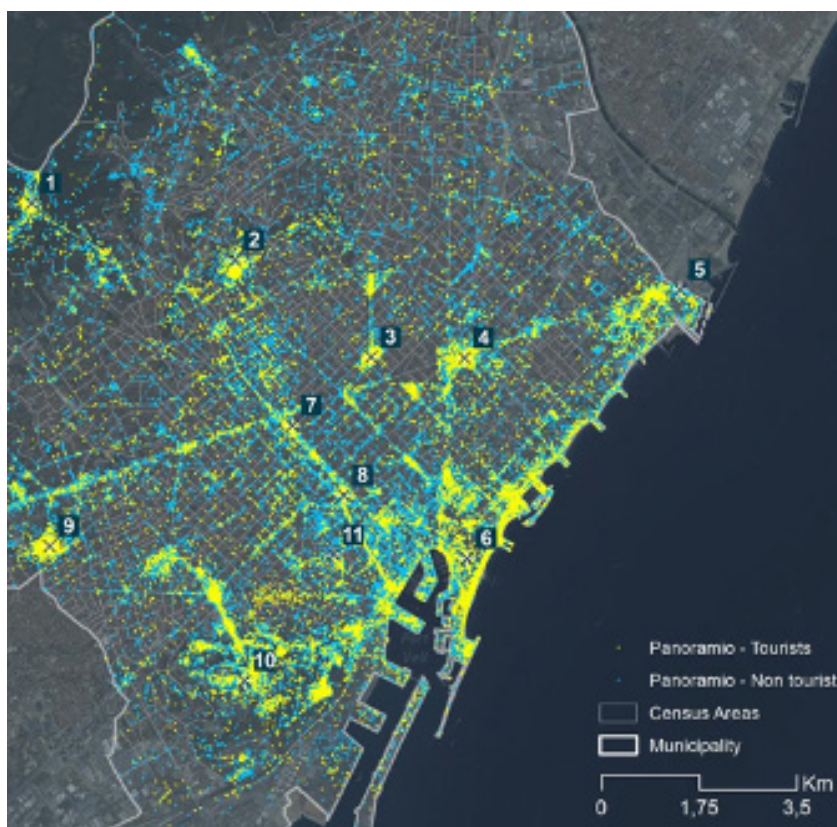
**Figura 21:** Mapa de residuos de las regresiones bivariadas entre las matrices obtenidas con Twitter y las obtenidas a partir de encuestas domiciliarias en la ciudad de Madrid a nivel de grandes zonas metropolitanas. Fuente: Osorio y García-Palomares, 2019



## 4.2 Análisis de patrones de movilidad de los turistas utilizando localizaciones puntuales de fotografías geolocalizadas

Otras redes sociales que producen datos puntuales geolocalizados mediante GPS y pueden ser usados para el estudio de la movilidad son las redes sociales de fotografías compartidas, por ejemplo, Panoramio, Flickr o Instagram. En este caso, sus datos pueden ser utilizados para obtener información de la movilidad de los turistas o visitantes en un determinado espacio. Esta información es de gran utilidad, puesto que apenas existen datos de cómo se mueven los turistas en sus diferentes destinos y el impacto que los mismos tienen sobre la movilidad de la población local.

Para utilizar los datos de estas redes sociales el tratamiento es similar al mostrado en el apartado anterior. En primer lugar, los datos son descargados a través de sus respectivas API. Aunque con una lógica parecida, cada una de las API presenta ciertas singularidades. En el caso de Flickr, por ejemplo, la descarga es relativamente sencilla y puede hacerse para las fotos subidas durante un periodo amplio de años. La descarga de datos de Instagram, por el contrario, es mucho más compleja. Una vez descargados, los datos son depurados y procesados para la selección de una muestra adecuada de usuarios. En este caso, los usuarios seleccionados serán aquellos que tienen actividad en el espacio de estudio durante un corto periodo de tiempo, menos de una semana o de 15 días (García-Palomares et al., 2015). Como puede verse en la Figura 22, la cartografía de las fotografías realizadas por turistas permite conocer bien los espacios más visitados de la ciudad.



**Figura 22:** Fotografías realizadas por usuarios de Panoramio en Barcelona, en amarillo las fotografías realizadas por posibles turistas. Fuente: Gutiérrez Puebla et al., 2017

Además de analizar la presencia de turistas en el destino turístico, su distribución e incluso su temporalidad a partir del momento del día en el que fue tomada la fotografía, los datos de fotografías geolocalizadas pueden ser usados para conocer las principales trayectorias turísticas a lo largo de la trama urbana (ver, por ejemplo, Girardin et al., 2008 o Okuyama et al., 2013).

En este caso, para el conjunto de fotografías geolocalizadas se trabaja con los siguientes atributos: a) owner-id: identificador del usuario que comparte la fotografía; b) -id: identificador de la fotografía; c) location-longitude: coordenada x; d) location-latitude: coordenada, y; e) dates-taken: fecha y tiempo en el que fue tomada (o tiempo de captura). Con esta información, para cada usuario se dispone de un conjunto de fotografías que pueden ser ordenadas cronológicamente como una secuencia espaciotemporal.

Para identificar los patrones de movilidad turística (rutas seguidas a través de la red de calles), el primer paso es construir, para cada usuario, el grafo/ secuencia entre cada par de fotos consecutivas y, a partir del tiempo de captura de cada una de ellas, extraer el período de tiempo transcurrido. La Figura 23 muestra los lugares donde un determinado usuario tomó fotografías durante un mismo día y la correspondiente secuencia punto a punto (foto a foto) en la ciudad histórica de Toledo (España). Gracias a la información geolocalizada (x, y) y al tiempo (t) de cada foto, es posible establecer un seguimiento/secuencia espaciotemporal de las fotografías tomadas:  $S = [(x_0, y_0, t_0), \dots, (x_k, y_k, t_k)]$ . Esta secuencia es una buena aproximación inicial para identificar las trayectorias seguidas por los turistas.



**Figura 23:** Ruta potencial de un turista a partir de sus fotografías geolocalizadas. Fuente: Mohino et al., 2018

El siguiente paso es la identificación de la ruta real seguida a través del entramado de calles. Para cada par de fotografías es posible obtener diferentes rutas que traten de representar la seguida por los turistas en el callejero de la ciudad visitada. Para esto pueden seguirse diferentes criterios, por ejemplo, buscar la ruta más corta entre cada par de fotos o buscar la ruta turística óptima (maximizando el número de espacios de interés turístico visitados) (Figura 24). A partir de las rutas obtenidas como posibles, se pueden comparar los tiempos de recorrido de esas rutas potenciales con los calculados según los tiempos de captura de las fotos, para, así, tratar de identificar la trayectoria real del turista. Esas trayectorias pueden ser llevadas al callejero para conocer los flujos de visitantes en cada uno de los tramos de calle.





**Figura 24:** Identificación de rutas turísticas a lo largo de la trama de calles. Comparando el tiempo transcurrido entre la captura de dos fotografías consecutivas con el tiempo de viaje de dos rutas posibles entre los puntos donde fueron tomadas ambas fotografías (opción A el camino de mínima distancia y opción B la ruta más probable), se concluye que la opción B es posiblemente la ruta seguida por el turista. Ejemplo en la ciudad de Toledo (España). Fuente: Mohino et al., 2018

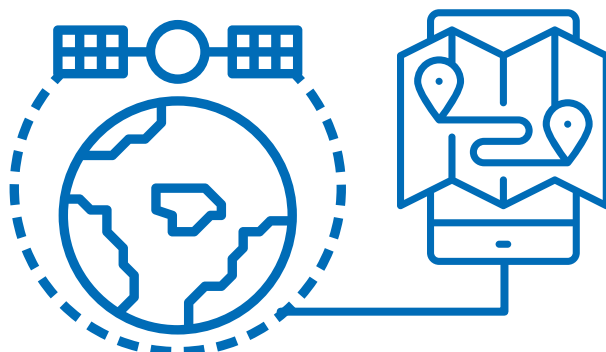






# 5

## **Análisis de Movilidad a partir de Tracks (Recorridos)**



## 5.1 Obtención de datos de recorridos GPS: fuentes y usos potenciales

Los recorridos o *tracks* registrados mediante dispositivos GPS son series ordenadas de puntos (*trackpoints*) para los que se han tomado coordenadas geográficas (longitud y latitud:  $x, y$ ), definiendo la trayectoria realizada por dicho dispositivo. Con frecuencia, estos terminales GPS recogen, además, un registro temporal para cada punto, correspondiente a la fecha y la hora, definidas no en hora local, sino en Tiempo Universal Coordinado (UTC). Los *trackpoints* que incluyen estas coordenadas espacio-temporales se ordenan secuencialmente de acuerdo a este dato temporal. Por otro lado, algunos dispositivos GPS también registran la altitud ( $z$ ), especialmente aquellos asociados a navegación no terrestre, sino aérea.

Aunque los primeros *tracks* GPS registrados con motivos civiles se remontan a mediados de los 1980s, estos no se hicieron populares hasta la década de 1990, cuando comenzaron a integrarse en dispositivos comerciales accesibles para compañías y para un número, aunque reducido, de particulares. Desde entonces, se han utilizado para gestionar el transporte marítimo, aéreo y terrestre. Sin embargo, la verdadera revolución de esta tecnología ha tenido lugar más recientemente. Está directamente relacionada con la aparición de los teléfonos inteligentes en 2007 y su rápida expansión en los años sucesivos. Desde entonces se ha incrementado de manera exponencial el registro de *tracks* GPS relativo a un número creciente de actividades.

Con la finalidad de estudiar el transporte y la movilidad, conviene clasificar los *tracks* GPS como registros de dos tipos de actividad: la actividad de empresas y la actividad de particulares. Por un lado, las compañías registran un número creciente de *tracks* GPS relativos tanto al desplazamiento de mercancía (en el sector de la logística), como al desplazamiento de su personal, en sectores clave como los servicios de emergencias, seguridad, vigilancia y muchos otros. Por supuesto, las compañías de transportes de pasajeros también han registrado de manera creciente *tracks* GPS de sus flotas, ofreciendo una información muy valiosa tanto a los gestores de dichas compañías como a los usuarios, que han visto mejorados los servicios de información de tiempos de viaje y espera en tiempo real.

Por otro lado, los *tracks* GPS relativos a la actividad de individuos o particulares también han crecido en igual o mayor medida. En este caso, además del impulso dado por la rápida expansión de los teléfonos inteligentes, hay que considerar que el uso de pulseras y relojes de actividad que incorporan tecnología GPS también se ha popularizado enormemente, aumentando con ello, aún más, tanto el grado de penetración del uso de esta tecnología como el espectro de actividades cubierto.

Ante este escenario ¿Cómo podemos valernos de la creciente generación de registros de *tracks* GPS relativos a la actividad de individuos y compañías para comprender y planificar mejor el transporte y la movilidad? ¿Cuáles son las fuentes de datos disponibles? ¿Qué tipo de datos nos ofrecen?

Esencialmente, podemos distinguir dos tipos de fuentes de datos: los *tracks* GPS provenientes de compañías o entidades y los capturados por iniciativas específicas con el propósito particular de analizar la movilidad o el transporte. Las diferencias entre los datos facilitados por unas o por otras tienen que ver normalmente con el volumen de datos capturados y con el nivel de información que se registra asociado al recorrido GPS.

### **TRACKS GPS OBTENIDOS DE COMPAÑÍAS**

Como señalamos anteriormente, compañías privadas o entidades públicas registran un creciente número de *tracks* GPS relativo al movimiento de mercancías, de flotas o de personas. En cuanto a los dos primeros, las compañías pueden facilitar datos exactos en la medida en la que sus políticas internas de privacidad lo permitan. Por ejemplo, pueden compartir los *tracks* GPS de los recorridos realizados por sus autobuses, lo cual permite realizar múltiples estudios de gran interés. Sin embargo, las compañías tienen más restricciones en cuanto a los datos relativos a la movilidad de las personas. Las leyes de protección de datos personales limitan, en mayor o menor medida dependiendo de cada país, el tipo de datos personales que pueden capturarse, así como el uso que de estos datos puede hacerse. Por norma general, las compañías no pueden facilitar datos que permitan a terceros identificar a personas en particular. Por su naturaleza, los *tracks* GPS registran coordenadas espaciotemporales con una resolución suficiente que en muchos casos podría servir para identificar a personas concretas. Por este motivo, son considerados datos muy sensibles y, por lo tanto, las compañías no facilitan su descarga de manera masiva. En consecuencia ¿Qué datos podemos obtener a partir de las compañías?

En primer lugar, aunque las compañías no pueden compartir con terceros datos sensibles sin autorización explícita de sus clientes, algunas compañías propietarias de aplicaciones de registro de *tracks* GPS, como *Endomondo*, *Garmin* o *Wikiloc*, permiten la descarga de *tracks* de usuarios cuando estos los hacen públicos para compartirlas con otros usuarios. Esta opción es común en las aplicaciones que funcionan también como redes sociales que facilitan este intercambio de información, a través de un consentimiento explícito de los usuarios. Dado que la descarga de estas rutas es posible y legal, distintos estudios se han basado en ellas para analizar la movilidad de personas en campos tan distintos como el del turismo o la movilidad urbana.

Un estudio ejemplar en este sentido es el realizado por Barros et al. (2019) con base en las rutas GPS de *Wikiloc* en el Parque Nacional del Teide, permitiendo mapear la intensidad de uso de la red de senderos que forman parte de la

infraestructura del parque (Figura 25). La intensidad de uso se midió de acuerdo a la cantidad de usuarios que recorrieron cada uno de los 35 senderos que forman parte de la infraestructura del Parque Nacional del Teide, permitiendo identificar qué senderos están sobre utilizados y subutilizados.

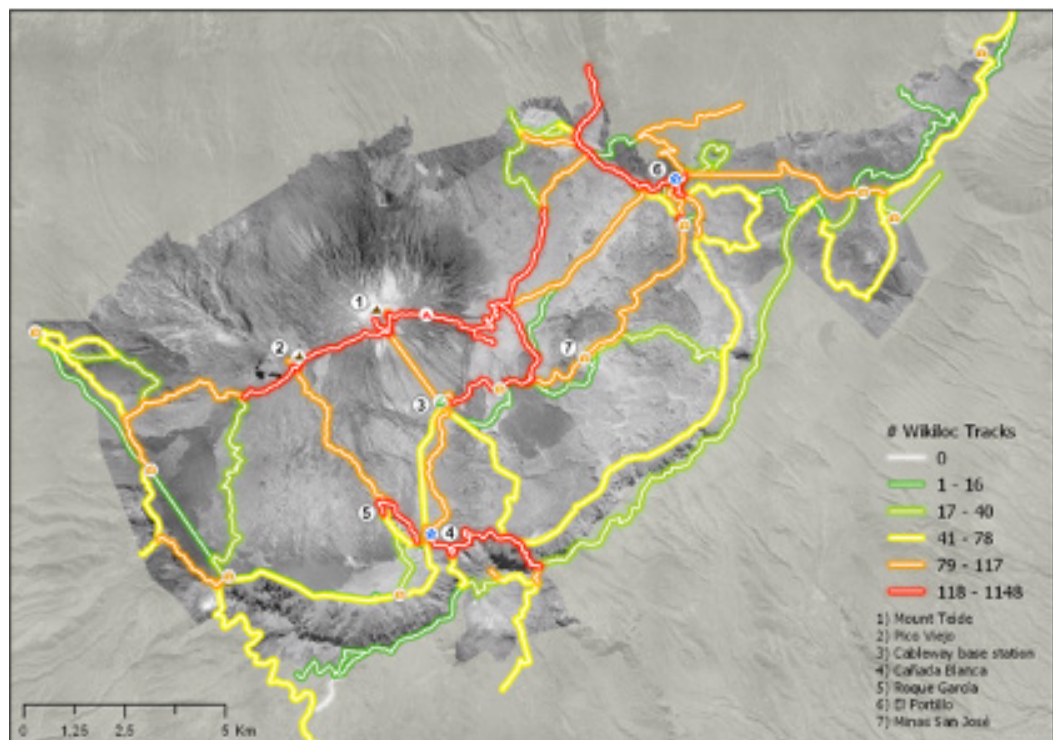
En segundo lugar, las compañías pueden facilitar datos tratados de manera que se proteja el anonimato de las personas. En este caso hemos de distinguir entre los datos correspondientes al propio *track* GPS y los datos asociados a cada *track* GPS. Si el *track* GPS no sirve para identificar al usuario (su posible domicilio o lugar de trabajo), como ocurre con los usuarios de los Sistemas de Bicicleta Compartida, este puede compartirse, junto con otros datos asociados a la ruta, como el género o la edad, normalmente definida dentro de un rango de edad como medida adicional de protección del anonimato. Este tipo de dato fue el utilizado en el estudio de la movilidad ciclista generada por el Sistema de Bicicleta Pública de Madrid, BiciMAD, realizado por Romanillos, Gómez-Moya, Zaltz-Austwick, y Lamíquiz-Daudén (2018), ejemplo desarrollado como estudio de caso en el presente manual.

Cuando los *tracks* GPS registrados por las compañías puedan servir para identificar a sus clientes, estas no podrán compartirlo sin su consentimiento. Por este motivo, las compañías que ofrecen datos de rutas, lo hacen de manera agregada (un número de rutas con origen y destino común) o, más comúnmente, facilitan el dato del número de rutas que pasan por un determinado arco de la red de transporte. Este es el dato que, por ejemplo, ofrece *Strava*. A través de su división comercial, *Strava Metro*, la compañía ofrece, además, datos agregados del número de viajes realizados por trabajo, ocio o deporte, así como si la ruta se ha realizado caminando, corriendo o en bicicleta.

Aunque en este último caso las compañías no ofrezcan el dato individual de ruta o datos concretos asociados a cada ruta, estas fuentes ofrecen datos de enorme valor, debido a dos cuestiones esenciales. La primera tiene que ver con la muestra de rutas recopilada, que en el caso de las aplicaciones más populares, como TomTom o Waze, se cuenta por millones. La segunda ventaja esencial de este tipo de aplicaciones es que no ofrecen una foto fija de la movilidad, sino que ofrecen un registro continuo de datos a lo largo del tiempo, lo que permite realizar estudios dinámicos para analizar los cambios producidos por la implementación de distintas políticas o el desarrollo de diversas infraestructuras.







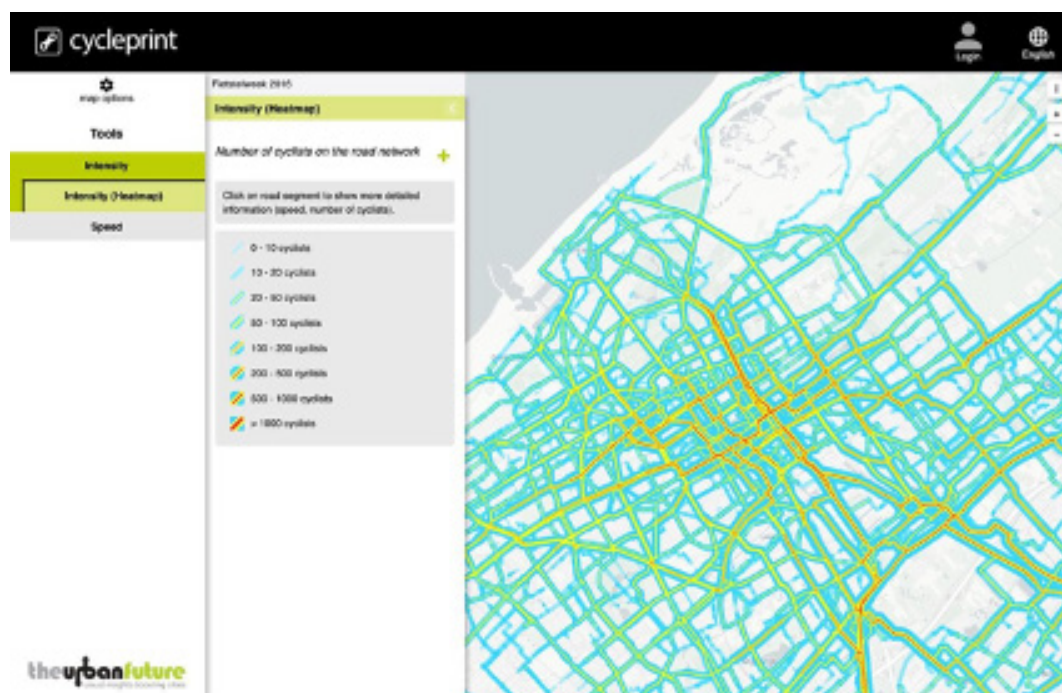
**Figura 25:** Mapa de densidad de rutas realizadas por usuarios de Wikiloc en el Parque Nacional del Teide, en la isla de Tenerife (España). Barros et al., 2019

### TRACKS GPS CAPTURADOS POR INICIATIVAS ESPECÍFICAS

La segunda fuente de datos de *tracks* GPS más relevante es la conformada por distintas iniciativas lanzadas con el objetivo específico de recopilar rutas de individuos e información asociada a dichas rutas, con la finalidad de realizar estudios e investigaciones concretas. Estas fuentes de datos presentan ventajas y desventajas en relación con las anteriores.

En cuanto a las desventajas, la primera es que la muestra de datos normalmente dista mucho de la recogida por las grandes compañías globales señaladas anteriormente. La segunda desventaja es que muchas de estas iniciativas capturan datos por un periodo de tiempo limitado y no de manera continua, como suelen hacer las compañías que citamos anteriormente. Por ello, su capacidad para monitorizar cambios en los patrones de movilidad se ve a menudo más restringida. En cualquier caso, se han lanzado iniciativas que, gracias al apoyo continuo de distintas instituciones, se han prolongado en el tiempo y han obtenido, además, una muestra muy relevante de datos. Sirve como ejemplo la iniciativa *Cycleprint*, (Bussche y Van de Coevering, 2015), lanzada en Holanda con el apoyo del gobierno nacional y la Unión Europea, que desde 2015 ha conseguido capturar las rutas ciclistas de miles de voluntarios en distintas ciudades (Figura 26).





**Figura 26:** Mapa de densidad de rutas ciclistas registradas por la iniciativa *Cycleprint*. Fuente: <https://theurbanfuture.com>

En cuanto a las ventajas, destacan normalmente una mayor riqueza de datos y un mayor nivel de desagregación de los mismos. Por lo general, estas iniciativas llaman a la participación de voluntarios que dan su consentimiento para que sus datos sean utilizados con un objetivo concreto. Los datos recopilados suelen corresponderse, a nivel espacial, con *tracks* GPS individuales para cada ruta realizada. Por otro lado, los participantes suelen aportar información relativa a sí mismos y a la ruta que realizan, como la edad, el género o el motivo del viaje. Ventajas adicionales derivadas del tipo de dato recopilado por estas iniciativas son, por ejemplo, la posibilidad de obtener los tiempos, distancias y velocidades de viaje, o de tramos particulares del viaje, que pueden analizarse en detalle.

La posibilidad de trabajar con rutas individuales permite estudiar patrones individuales de movilidad que pueden determinar patrones de movilidad a nivel urbano o territorial. Por ejemplo, en el análisis de elección de la ruta se puede estimar qué prioridades tienen los viajeros en cuanto a los distintos modos de transporte o qué atractivo tienen para los usuarios determinadas infraestructuras frente a otras. Cuál es el impacto de hacer transbordos, qué infraestructuras prefieren los ciclistas o cuál es el impacto de la densidad de tráfico rodado en su elección de ruta diaria. En los apartados siguientes hablaremos más en detalle sobre el potencial de estos datos para el análisis de la movilidad.

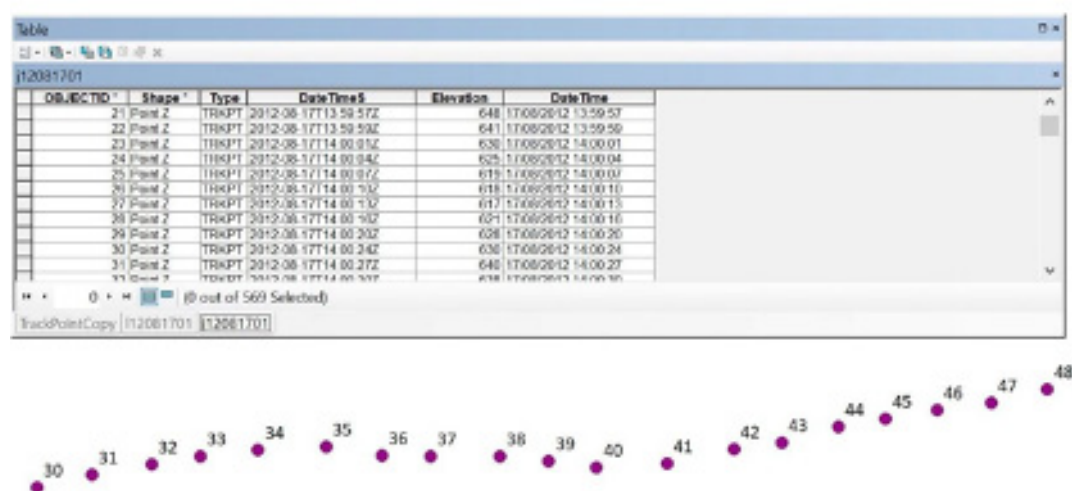


## 5.2 Metodología para el tratamiento y enriquecimiento de los datos

### TRATAMIENTO BÁSICO DE *TRACKS* GPS

Como ya señalamos, los recorridos o *tracks* registrados mediante dispositivos GPS son series ordenadas de puntos para los que se han tomado coordenadas geográficas (longitud y latitud) y, con frecuencia, un registro temporal para cada punto, correspondiente a la fecha y la hora. Es común que algunas aplicaciones registren la altitud (coordenada z). Estos son normalmente los datos mínimos de partida almacenado por los dispositivos GPS en los ficheros GPX.

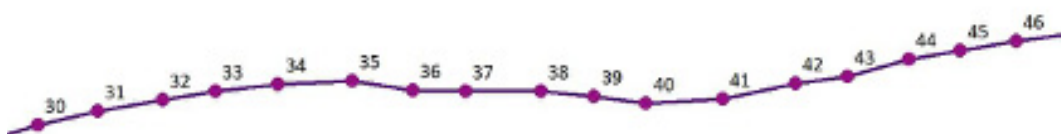
A partir de las coordenadas almacenadas en el fichero GPX podemos generar una capa de puntos geolocalizados, ordenados conforme al registro temporal. La Figura 27 visualiza un conjunto de *trackpoints* y la tabla de datos asociada, generada mediante el procesamiento del fichero GPX con un SIG (ArcGIS en este caso).



**Figura 27:** Tabla de datos relativa a un track GPS original y visualización de los *trackpoints*. Fuente: Elaboración propia

A partir de la capa de puntos geolocalizados podemos generar la línea que los une, conformada por los segmentos que enlazan cada punto con el siguiente. Estos segmentos suelen denominarse *tracklines*. La Figura 28 muestra los *tracklines* correspondientes a la serie anterior de *trackpoints*, y la tabla de datos asociada. Como puede observarse en la tabla, para cada segmento ya se ha calculado su longitud, duración del recorrido a lo largo de dicho segmento y la velocidad del viaje, calculada a partir de las variables anteriores. Dado que tenemos registrado el dato de altitud para cada punto, también podríamos obtener la pendiente de cada *trackline*, así como el desnivel. Si la altitud no hubiese sido registrada por el dispositivo GPS, podría estimarse a partir de un Modelo Digital del Terreno (MDT) mediante la aplicación de algunas herramientas disponibles en SIG convencionales, si este modelo estuviese disponible con una resolución suficiente y ajustada a la del *track* GPS.

OID	Shape	From_ID	To_ID	Start_Time	End_Time	Distance	Duration_S	Speed_KPH	Shape_Length
30	Polyline	30	31	2012-06-17T14:00:24Z	2012-06-17T14:00:27Z	15.68257	3	18.819084	15.68257
31	Polyline	31	32	2012-06-17T14:00:27Z	2012-06-17T14:00:30Z	16.842307	3	20.216709	16.842307
32	Polyline	32	33	2012-06-17T14:00:30Z	2012-06-17T14:00:33Z	13.647601	3	19.376428	13.647601
33	Polyline	33	34	2012-06-17T14:00:33Z	2012-06-17T14:00:36Z	16.015003	3	19.216075	16.015003
34	Polyline	34	35	2012-06-17T14:00:36Z	2012-06-17T14:00:39Z	18.83344	3	22.600128	18.83344
35	Polyline	35	36	2012-06-17T14:00:39Z	2012-06-17T14:00:41Z	15.671705	2	28.266068	15.671705
36	Polyline	36	37	2012-06-17T14:00:41Z	2012-06-17T14:00:43Z	13.472773	2	24.256962	13.472773
37	Polyline	37	38	2012-06-17T14:00:43Z	2012-06-17T14:00:46Z	19.2503	3	23.10996	19.2503
38	Polyline	38	39	2012-06-17T14:00:46Z	2012-06-17T14:00:48Z	13.533796	2	24.368632	13.533796
39	Polyline	39	40	2012-06-17T14:00:48Z	2012-06-17T14:00:50Z	13.344249	2	24.019629	13.344249



**Figura 28:** Tabla de datos relativa a un track GPS original y visualización de las *tracklines*. Fuente: Elaboración propia

La suma de los *tracklines* conforma la ruta GPS. La longitud total de la ruta podrá calcularse como la suma de las longitudes de los *tracklines*. También es posible calcular el desnivel acumulado, como suma de los desniveles de cada *trackline*, y estadísticas generales de la ruta como la velocidad media, la pendiente media o el desnivel total.

Esta información es muy valiosa para realizar estudios de transporte y movilidad, aunque el trabajo con las rutas GPS definidas por los *tracks* puede presentar algunas limitaciones importantes. En primer lugar, una ruta GPS definirá la trayectoria real seguida por el dispositivo con mayor o menor precisión dependiendo de la resolución temporal de los *trackpoints*. En el caso que hemos tomado como ejemplo, el dispositivo GPS ha registrado la localización con intervalos de dos a tres segundos de duración. Esta resolución temporal es muy común en los *tracks* registrados con aplicaciones de registro de actividades de ciclismo o running, y permite reproducir la trayectoria con bastante precisión. Si el intervalo fuera mayor, quizá no estaríamos registrando trayectorias reales entre dos *trackpoints* alejadas de la línea recta que definen los *tracklines*. Estaríamos, por tanto, simplificando la trayectoria, reduciendo la longitud estimada respecto a la real y, dado que el intervalo temporal es el mismo, aumentando la velocidad estimada respecto a la real.

En segundo lugar, los *tracks* GPS pueden presentar lo que se denomina comúnmente como "ruido", que es cierta imprecisión en la geolocalización de algunos de los *trackpoints*. Estas imprecisiones suelen darse cuando la posición del dispositivo GPS no se ha triangulado correctamente a partir de los satélites, debido a una deficiente cobertura. Esta puede aparecer, por ejemplo, con la presencia de obstáculos, como pueden ser los edificios en un ámbito urbano muy denso. *Tracks* "ruidosos" también pueden dar lugar a estimaciones de rutas poco precisas, normalmente con el alejamiento de *trackpoints* con respecto a la localización real del dispositivo GPS, dando lugar a trayectorias no sólo imprecisas, sino también de una longitud estimada superior a la real y una velocidad estimada, en consecuencia, superior a la real. Para evitar este tipo de problemas, las rutas GPS suelen tratarse mediante un proceso denominado "*mapmatching*".

## EL PROCESO DE MAPMATCHING

Con el término *mapmatching* se hace referencia al ajuste de las rutas GPS originales (cuya geometría se ha obtenido directamente a partir de la unión de los *trackpoints*) a la geometría de una red de transporte, de carreteras o calles.

El objetivo principal del *mapmatching* es evitar las imprecisiones debidas a los posibles errores anteriormente mencionados (baja resolución temporal o ruido en los *tracks*). El resultado del proceso de *mapmatching* será una ruta ajustada a la geometría de la red. Si esta geometría tiene una buena resolución y el ajuste ha sido bueno, la ruta ajustada (o “mapmatched”) será más precisa que la anterior, tendrá una longitud estimada más cercana a la real y unas velocidades estimadas también más precisas.

Existe, además, un segundo objetivo en el proceso de *mapmatching*. Al llevarnos las rutas a la red de transporte, podemos enriquecer la información de las rutas a partir de los datos disponibles para cada arco de dicha red (ver en la siguiente sección este proceso de enriquecimiento de datos).

En los últimos años se han aplicado distintos métodos de *mapmatching*, con mayor o menor éxito, medido a partir del porcentaje de rutas que se han logrado ajustar a la red satisfactoriamente, es decir, resultando fieles al recorrido real del dispositivo. A modo de ejemplo, la Figura 29 ilustra de manera sintética la manera de proceder de uno de estos algoritmos, el utilizado por Romanillos y Gutiérrez (2019) en un estudio que analiza rutas ciclistas GPS en Madrid.

La metodología seguida en este ejemplo se ha desarrollado en un SIG, tomando como datos de partida los puntos de origen y destino de la ruta GPS, el track GPS y la red de transporte a la que queremos ajustar el track. El algoritmo desarrollado genera, en primer lugar, un buffer a 25 metros de distancia del track GPS. A continuación, recorta la red vial de acuerdo a los límites del polígono generado como buffer, y trata de encontrar una ruta posible entre los puntos inicial y final a través de la red de transporte recortada. Si el proceso se realiza con éxito, se dará como buena la ruta generada, y se exportará la línea como nueva ruta ajustada a la red. Si, por el contrario, no se ha podido encontrar una ruta posible, el proceso comienza de nuevo, en esta ocasión generando un buffer a una distancia superior, de 50 metros de distancia. Cuanto mayor sea el ruido del track GPS, mayor será el buffer y mayor la posibilidad de generar una ruta diferente a la real en alguno de sus segmentos, especialmente allí donde la densidad de calles es mayor y existen varios posibles caminos entre dos puntos. La mejora de este algoritmo viene dada por el hecho de que no se aplica una única distancia para el buffer, sino que éste se define para cada ruta a partir de la distancia menor, de manera que se ajuste en cada caso con la mayor precisión posible. Esta mejora supone un incremento significativo en el porcentaje de rutas finalmente ajustadas con éxito.

Cuando la resolución temporal de los *tracks* GPS es muy baja, suelen utilizarse otras metodologías más apropiadas, ya que el buffer generado a partir de *tracklines* de mucha longitud podría dar lugar a errores significativos. Por ejemplo, el estudio de las rutas de los usuarios del Sistema de Bicicleta Pública de Madrid realizado por Romanillos et al. (2018) partía de un conjunto de *tracks* GPS registrados con una resolución temporal media de 75 segundos. Esta resolución no permite aplicar la metodología anterior con garantías de éxito, por lo que se optó por aplicar un algoritmo sencillo que busca siempre la ruta entre un punto y el sucesivo por el camino más corto posible a través



de la red vial a la que se quiere ajustar el track. Aun siendo un sistema menos preciso que el anterior, por lo general, las rutas finalmente obtenidas se ajustan bien a las originales, teniendo en cuenta que a menudo no existen muchas alternativas de ruta posibles entre dos puntos distanciados 75 segundos, salvo cuando la red de calles sea muy densa y la velocidad elevada, algo que, por lo general, suele no ir ligado.



**Figura 29:** Ilustración que sintetiza el proceso de *mapmatching* con una versión mejorada del algoritmo de Dalumpines y Scott. Fuente: Romanillos y Gutiérrez, 2019

## ENRIQUECIMIENTO DE LOS DATOS

Como hemos visto, los *tracks* registrados por dispositivos GPS permiten obtener información muy valiosa de cara al estudio de la movilidad o el transporte: la definición de la trayectoria espacial seguida, la estimación de distancias, tiempos y velocidades de viaje, parciales o totales, e incluso los desniveles totales o acumulados. Ahora bien, podemos enriquecer los datos de nuestro track GPS a partir de otras fuentes. A continuación, se describen distintas formas de enriquecer los datos de los *tracks* GPS.

Por un lado, debemos considerar que los datos de *tracks* GPS normalmente son registrados por distintas aplicaciones a través del uso de dispositivos GPS. Estas aplicaciones, además de registrar los *tracks*, pueden registrar otros datos adicionales, enriqueciendo la muestra del track. Los datos con los que enriqueceremos la muestra serán de una naturaleza u otra dependiendo del enfoque de cada aplicación. Por ejemplo, algunas aplicaciones de compañías como Garmin registran si el track se realizó en un día lluvioso, nublado o soleado, sin necesidad de que el dispositivo en el que está integrado el GPS tenga sensores que registren este dato. La aplicación busca, de manera paralela, la información de meteorología asociada a la localización y el momento del registro del track GPS, y la registra en un nuevo campo asociado a la ruta, que ya forma parte del fichero GPX a descargar de la aplicación. Otros dispositivos GPS, alojados, por ejemplo, en sistemas con sensores de temperatura o humedad, pueden registrar también estos datos asociados a cada trackpoint.



Por otro lado, podemos enriquecer los datos de los *tracks* GPS a partir del proceso de *mapmatching*. Al ajustar el track GPS a una red vial, podemos asignar a la ruta ajustada o a los segmentos que la componen la información que tengamos sobre cada segmento de la red vial. Por ejemplo, podemos tener, para un segmento de la ruta GPS, no sólo la información de la distancia, la duración, la velocidad o el desnivel de ese tramo de la ruta, que obtuvimos a partir del track, sino también datos relativos a cuestiones como la densidad del tráfico rodado, el tipo de calle, la existencia o no de algún tipo de infraestructura, la existencia de semáforos, la distancia a cruces, el número de carriles, el sentido de circulación de la vía, etc. Mediante un proceso de “unión espacial” de los datos, la ruta ajustada capturará los datos de la red sobre la que se ajusta. Esta unión espacial puede realizarse fácilmente con algunas herramientas comunes de los SIG.

El enriquecimiento de los *tracks* GPS abre la puerta a un número mucho mayor de potenciales análisis a realizar, tanto relativos a los patrones de movilidad de las personas como a aquellos de la movilidad a escala urbana o territorial.



### 5.3 Análisis de patrones de movilidad de personas a partir de *tracks* GPS

El análisis de los patrones de movilidad de personas es uno de los campos de estudio más relevantes con base en los datos de *tracks* GPS. Existe una gran variedad de análisis potenciales a realizar, y su posible desarrollo dependerá, esencialmente, de la naturaleza de los datos disponibles, tanto a nivel del track GPS en sí mismo, como a nivel de los datos asociados a dicho track. Por supuesto, será importante atender otras cuestiones, como el tamaño y distribución de la muestra de datos, con el fin de validar positivamente su representatividad.

En primer lugar, debemos distinguir los posibles tipos de análisis a realizar en función de si los datos disponibles son rutas agregadas a nivel de segmento de red de transporte o son *tracks* GPS relativos a trayectorias individuales de viaje.

Si la muestra de datos de *tracks* GPS está agregada, el tipo de análisis es más limitado. En este caso, los datos ofrecidos pueden ser los relativos a tiempos y distancias medias de viaje, horarios de salida y llegada, de acuerdo a perfiles personales caracterizados de manera muy básica, usualmente en cuanto a género, grupos de edad y alguna característica más concreta en función del conjunto de datos específico.

Si, por el contrario, la muestra de datos se corresponde con *tracks* GPS relativos a trayectorias individuales de viaje, las oportunidades de análisis son mucho mayores. Por ejemplo, a partir del análisis detallado de los *tracks* GPS podemos estudiar los tiempos y las velocidades de viaje a nivel de la línea del track (trackline), que es la mínima unidad de recorrido de la que tenemos información, es decir, el nivel más desagregado de información a nivel espacial. El estudio de los *tracks* GPS con este nivel de detalle puede dar información acerca del impacto que tienen en los tiempos y velocidades de viaje variables

como la pendiente de la calle, la proximidad a intersecciones, la presencia de semáforos o el tipo de infraestructura por la que se viaja.

Como ejemplo, la Figura 30 muestra una serie de *tracks* GPS correspondientes a rutas ciclistas, representados a nivel de trackline, de acuerdo a la velocidad de viaje. La figura representa los recorridos de ciclistas a lo largo de una calle con una pendiente relevante. Los ciclistas que recorren esta calle con pendiente positiva, circulando en sentido de izquierda a derecha, por el carril derecho, lo hacen con una velocidad reducida, entre los 5 y los 15 km/h. Los ciclistas que descienden por esta calle, circulando de derecha a izquierda por el carril derecho, comienzan con velocidades intermedias y aceleran hasta llegar a velocidades de 30-35 km/h. Además, la figura representa también la reducción de velocidad de algunos ciclistas cuando estos se aproximan a intersecciones o realizan cambios bruscos de dirección. El trabajo con miles de *tracks* GPS a este nivel de detalle permitió realizar un análisis cuantitativo del impacto que estas variables (pendiente y proximidad a intersecciones) y muchas otras (tipo de infraestructura, tipo de calle, densidad de tráfico rodado, velocidad del tráfico rodado, etc.) tienen sobre las velocidades ciclistas (Gustavo Romanillos y Gutiérrez, 2019). Hasta la fecha, la mayor parte de los estudios que analizaban, por ejemplo, el impacto de la pendiente en la movilidad ciclista, se basaban en mediciones realizadas de manera manual, en un día concreto, en una serie de calles concretas y para un número muy limitado de ciclistas y rutas ciclistas.

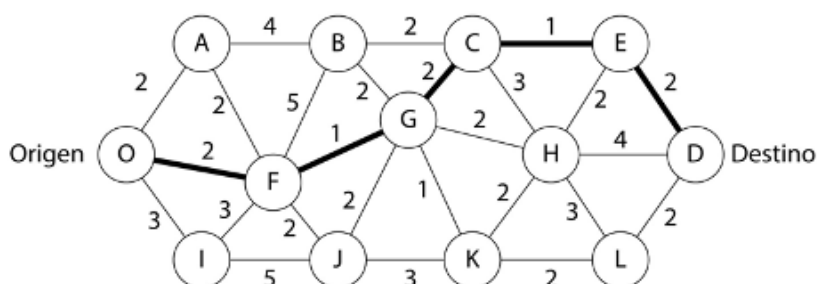


**Figura 30:** Representación de *tracklines* GPS de ciclistas en función de la velocidad. Fuente: Romanillos y Gutiérrez, 2019

El trabajo con *tracks* GPS relativos a trayectorias individuales de viaje permite realizar otro tipo de análisis de gran interés en el campo de la movilidad y el transporte: el análisis de elección de ruta. Este tipo de análisis busca determinar cuáles son las preferencias de los viajeros a partir del análisis de las rutas realizadas frente a otras rutas alternativas descartadas. Los modelos de elección de ruta generan rutas alternativas a la realizada, asignando un peso específico a determinadas variables de viaje o del recorrido. Por ejemplo, podemos generar la ruta que responde al mínimo tiempo de viaje o a la mínima

distancia de viaje, la que minimiza los cambios de línea de autobús o metro, la que minimiza la pendiente en viajes ciclistas o la que prioriza un itinerario sin tráfico excesivo.

La Figura 31 ilustra una red de transporte compuesta por nodos y arcos que unen los nodos. Estos arcos de red contienen un valor de impedancia (la resistencia a moverse a través de la red). Esta impedancia puede representar, por ejemplo, el tiempo de viaje, la distancia, el costo económico del viaje o la comodidad o seguridad del mismo. Los modelos de elección de ruta generan la ruta de menor coste total acumulado entre el punto de origen (O) y el punto de destino (D). Considerando distintas variables y el posible impacto de ellas mediante la asignación de un coste, el modelo generará diferentes rutas. A menudo, los pesos se reparten de manera aleatoria (modelos de elección discreta en transporte con coeficientes aleatorios). Para ciertos contextos, la asignación de costes específicos sí es importante. Un ejemplo es el costo del transporte público en barrios de bajos ingresos económicos. Aquí, un incremento en la tarifa tiene implicaciones mucho más elevadas que el tiempo de viaje o la distancia en términos de accesibilidad. Existen múltiples modelos de elección de ruta y cada uno de ellos genera las rutas alternativas de manera distinta. Finalmente, de la comparación de todas estas rutas con la ruta original, los modelos estimarán un peso específico a cada una de las variables consideradas.



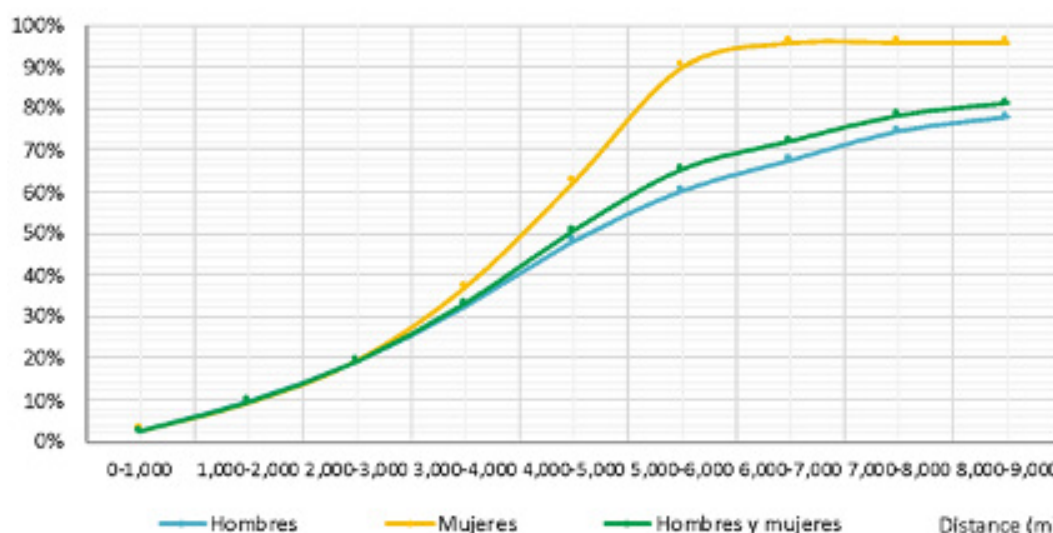
**Figura 31:** Representación de la asignación de costes de viaje a una red de transporte para la generación de rutas entre un punto de origen y un destino y cálculo del camino más corto. Fuente: Elaboración propia

Los análisis de elección de ruta son de enorme utilidad en la planificación de políticas e infraestructuras de transporte con la finalidad de valorar las preferencias de los viajeros y tienen distintas aplicaciones. Por un lado, las estimaciones de los modelos de elección de ruta sirven para alimentar y calibrar modelos de estimación de demanda o reparto modal, que son la base del diseño y planificación de infraestructuras de transporte. En este sentido, los modelos de elección de ruta también son relevantes para hacer estimaciones de demanda a futuro; y nos permitirán contestar algunas preguntas del tipo: ¿Cuántos viajeros utilizarán esta nueva ruta frente a las alternativas? ¿Cuántos ciclistas utilizarán este itinerario ciclista a través de un carril bici aislado que transcurre por un parque, aunque esto suponga un incremento de su distancia y tiempo de viaje?

Por otro lado, las estimaciones de los modelos de elección de ruta son necesarias para la generación de “enrutadores” inteligentes, aplicaciones que

puedan facilitar a los usuarios el conocimiento de la ruta a seguir de acuerdo con distintas preferencias. Por ejemplo: ¿Cuál es la ruta más adecuada para una persona mayor que no prioriza tanto el tiempo de viaje como la comodidad del mismo, minimizando, por ejemplo, el número de intercambios o la distancia caminando?

Si, además, los *tracks* GPS relativos a rutas individuales vienen acompañados de datos asociados a la ruta o el perfil del individuo que realiza el viaje, las posibilidades de análisis se multiplican. Podremos, por ejemplo, analizar los patrones de movilidad de las personas en función de su género, edad o motivo del viaje, cuestiones de gran relevancia en muchos modos de transporte. A modo de ejemplo, la Figura 32 representa el porcentaje de viajes acumulados de acuerdo con la distancia de viaje para ciclistas, según género, en la ciudad de Madrid (Gustavo Romanillos y Zaltz Austwick, 2015). La línea verde, que representa los viajes según distancia sin distinguir entre género, muestra cómo más de la mitad de los viajes ciclistas registrados superan los 5 km de distancia. Sin embargo, si atendemos a la línea amarilla, correspondiente a los viajes ciclistas realizados por mujeres, vemos cómo en este caso tan sólo un 38% de los viajes superan los 5 km de distancia. Esta diferencia se acentúa para distancias superiores, mostrando un desequilibrio importante entre género, que podría pasar desapercibido sin este análisis desagregado.



**Figura 32:** Porcentaje de viajes ciclistas realizados en función de la distancia y según género. Fuente: Romanillos y Austwick, 2015

Todos los estudios mencionados anteriormente, tanto los modelos de elección de ruta como los análisis relativos a la influencia de variables como la pendiente o el tráfico en las velocidades, distancias y tiempos de viaje, pueden realizarse en función de los distintos colectivos en que podamos desagregar la muestra de *tracks* GPS: por género, edad, motivo del viaje, etc. Análisis de este tipo pueden conducir a la promoción de medidas o de políticas de infraestructuras que atiendan a las necesidades específicas de cada colectivo.





## 5.4 Monitorización y análisis de la movilidad en la ciudad y el territorio a partir de *tracks* GPS

En ocasiones, el volumen de *tracks* GPS registrado puede corresponderse con una muestra suficientemente grande como para ser representativa de la movilidad en un modo determinado de transporte, o de un colectivo en concreto, para un lugar específico. En estos casos, podemos pasar de estudiar los patrones de movilidad individuales a analizar los patrones de movilidad de todo un territorio, una ciudad o cualquier área en particular.

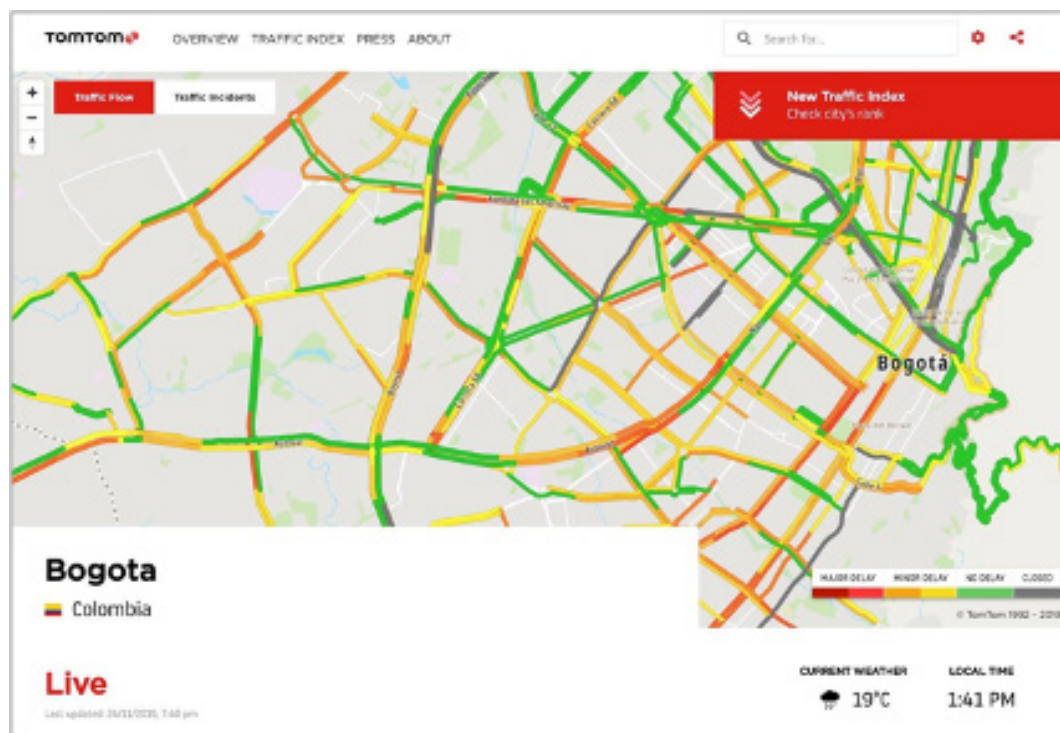
Con la rápida expansión de las aplicaciones de servicios basados en la geolocalización, muchas de ellas ligadas a la movilidad, cada vez es más frecuente tener un registro de datos GPS significativo. Esto abre la puerta a estudiar aspectos de la movilidad que anteriormente no había sido posible explorar con suficiente detalle o profundidad. A continuación, identificaremos y describiremos algunos de estos nuevos análisis.

### ANÁLISIS DE FLUJOS DE TRANSPORTE EN LA RED A PARTIR DE *TRACKS* GPS

En primer lugar, podemos destacar las oportunidades de monitorización de los flujos de la movilidad y el transporte, cuando existe la disponibilidad de un registro de un alto número de *tracks* GPS (como ocurre con TomTom o Waze). Aunque en estos casos la muestra de viajes no es completa, sí es suficientemente grande para realizar análisis relevantes de las redes viales. Sus registros de tiempos de viaje y niveles de congestión son indicadores bastante precisos del nivel de tráfico existente en la realidad. A modo de ejemplo, la Figura 33 muestra la visualización online del estado de congestión de la red de carreteras de Bogotá, en tiempo real, ofrecida por TomTom. A partir de estos datos pueden realizarse novedosos análisis dinámicos de accesibilidad en áreas urbanas (Moya-Gómez, Salas-Olmedo, García-Palomares, y Gutiérrez, 2017).







**Figura 33:** Nivel de tráfico en la red vial, en tiempo real. Fuente: TomTom

También es relevante el trabajo de exploración de la movilidad urbana y metropolitana realizado desde iniciativas y proyectos de investigación, con un objetivo más acotado. Por ejemplo, el análisis de la movilidad ciclista de Madrid realizado por Gustavo Romanillos y Zaltz Austwick (2015) permitió analizar, por primera vez, la huella de los ciclistas urbanos de dicha ciudad. La investigación, basada en el análisis de miles de *tracks* GPS compartidos por ciclistas voluntarios a través de la iniciativa Huella ciclista de Madrid, permitió analizar los patrones de movilidad ciclista en el área metropolitana de Madrid. Gracias a la recopilación de datos asociados a los *tracks* GPS se pudo analizar la extensión de la huella ciclista en la ciudad e identificar cuáles eran las arterias más importantes en términos de flujo ciclista, desagregadas según el motivo de viaje (Figura 34). La iniciativa se abrió posteriormente a la participación de distintas compañías de bici-mensajería, realizando un análisis de los patrones de movilidad de este colectivo emergente, promovido en la actualidad en muchas ciudades para fomentar una logística más sostenible, enfocada en el uso de este transporte para la denominada última milla.



**Figura 34:** *Tracks* GPS de ciclistas voluntarios en el área metropolitana de Madrid. Fuente: Romanillos et al., 2018



**Figura 35:** *Tracks* GPS de BiciMensajeros en Madrid. Fuente: Romanillos et al., 2018

En ocasiones, es posible obtener los *tracks* GPS relativos a la totalidad de los viajes de una determinada flota de transporte, público o privado. Así, la misma iniciativa Huella ciclista de Madrid, incorporó la totalidad de las rutas ciclistas generadas por el Sistema de Bicicleta Pública de Madrid (G. Romanillos, Moya-Gómez, Zaltz-Austwick, y Lamíquiz-Daudén, 2018), dando lugar a una investigación que se presenta como estudio de caso en el capítulo 6.

### **ANÁLISIS DINÁMICOS DE MOVILIDAD A PARTIR DE *TRACKS* GPS**

El interés de este tipo de exploraciones no radica solamente en lo exhaustivo del análisis (la muestra es la totalidad de los viajes), sino en la posibilidad de



monitorizar la movilidad a lo largo del tiempo, siempre y cuando exista un registro de *tracks* GPS continuo.

La monitorización a lo largo del tiempo, permite realizar análisis comparativos del uso de las infraestructuras en distintos momentos: distintas fechas a lo largo del año, en días laborables frente a fines de semana, o incluso la movilidad a lo largo del día. Para realizar este tipo de exploraciones se dispone de herramientas de visualización dinámica. Un ejemplo paradigmático en este sentido, por trabajar con un gran volumen de *tracks* GPS registrados por una flota de transporte público, es el análisis dinámico de la movilidad de más de 170 millones de viajes de taxi realizados en la ciudad de Nueva York a lo largo de un año, desarrollado por el Senseable City Lab del MIT (Szell y Groß, 2014). El estudio permitió analizar la densidad de flujos de taxis en las calles de Nueva York, así como la densidad de los puntos de origen y destino de los viajes, a lo largo del día (Figura 36).



**Figura 36:** Visualización del tráfico de taxis en Nueva York. Proyecto Hubcab. Fuente: Szell y Groß, 2014

La disponibilidad de *tracks* GPS registrados de manera continua por algunos sistemas o compañías, permiten abordar un análisis mucho más detallado de la movilidad y el uso de las infraestructuras de transporte en determinados momentos del día (en hora punta o en hora valle); pero, también, en días señalados en los que los patrones de movilidad puedan ser significativamente distintos. Muchas ciudades cierran al tráfico determinadas calles o áreas urbanas durante los días festivos ¿Cómo afectan estas medidas a la movilidad? ¿Por dónde se desvía el tráfico rodado? ¿Existe un incremento de la movilidad ciclista o peatonal en estas zonas? El análisis con base en los datos relativos a dichas fechas puede responder estas preguntas y permitir tomar decisiones más informadas a futuro.

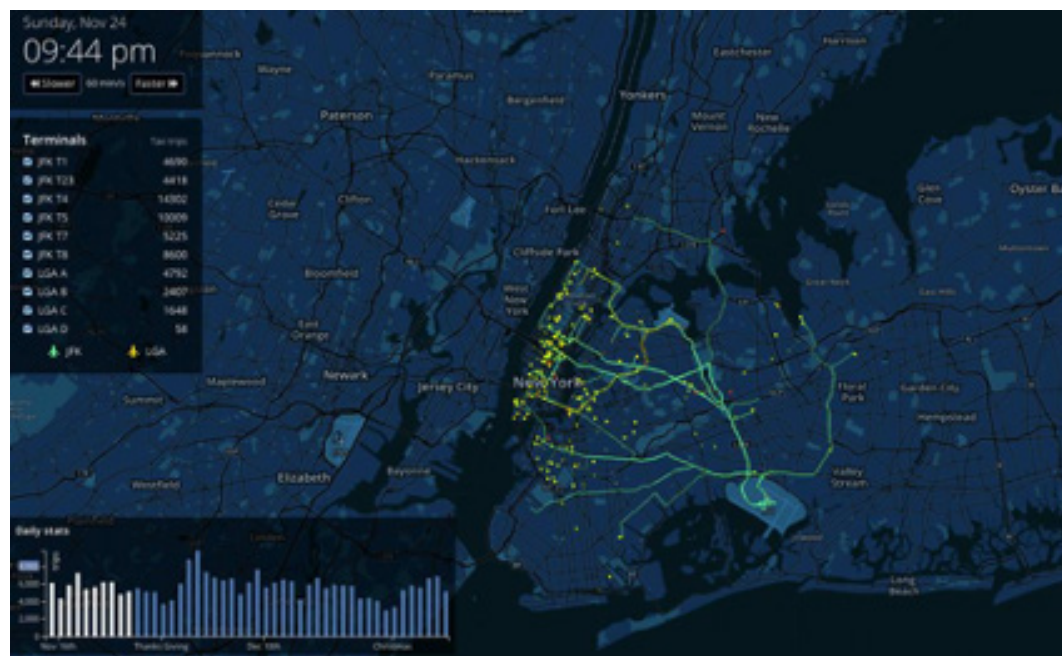
Por otro lado, el análisis de la movilidad a lo largo del tiempo también permitirá

evaluar, de manera rápida, cuál es el impacto que la implementación de determinadas políticas o infraestructuras está teniendo en la ciudad ¿Están siendo utilizados los nuevos carriles ciclistas? Existen diversos ejemplos de evaluación de infraestructuras con base en análisis de *tracks* GPS que han revisado y aportado metodologías en esta línea (Hong, McArthur, y Livingston, 2019).

### ANÁLISIS DEL IMPACTO EN LA MOVILIDAD DE EVENTOS O FECHAS ESPECIALES A PARTIR DE *TRACKS* GPS

Debemos resaltar el potencial de la disponibilidad de *tracks* GPS, registrados de manera continua por algunos sistemas, para el análisis del impacto en la movilidad de eventos especiales o de fechas señaladas. El análisis de los patrones de movilidad en estas fechas es crucial para la planificación de medidas extraordinarias que, por un lado, permitan estimar la demanda y, en consecuencia, la oferta necesaria y, por otro lado, eviten la formación de grandes niveles de congestión.

En este sentido, puede servir de ejemplo ilustrativo el estudio realizado por ImageWork Technologies (Figura 37), enfocado en analizar el incremento del flujo de taxis y tiempos de viaje con destino al aeropuerto en la víspera de un día festivo nacional (Thanksgiving) en la ciudad de Nueva York. El estudio muestra, de manera dinámica, el número y el recorrido de los taxis con destino a las distintas terminales de los aeropuertos más importantes de Nueva York, a lo largo del día.



**Figura 37:** Visualización del tráfico de taxis en Nueva York. Fuente: ImageWork Technologies, 2014



## 5.5 Modelización de la movilidad: simulaciones para escenarios futuros en cuanto a la implementación de políticas e infraestructuras

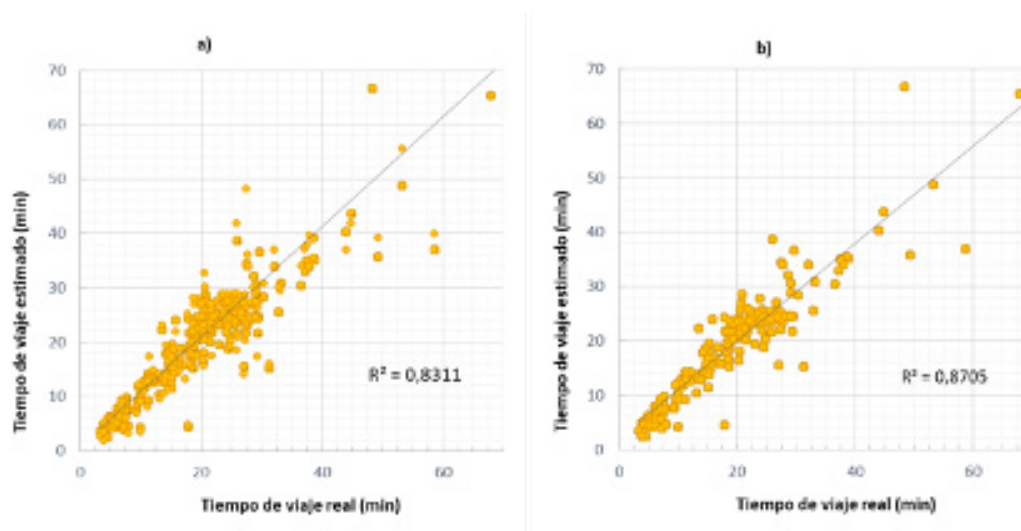
Los recorridos o tracks registrados mediante dispositivos GPS no son tan sólo la base de distintos estudios centrados en la monitorización o el análisis de la movilidad y el transporte. Son, también, la base de múltiples modelos de simulación del transporte que pueden ser utilizados, por ejemplo, para predecir los patrones de movilidad esperables tras la implementación de determinadas políticas o infraestructuras.

El análisis de tracks GPS relativos a viajes individuales nos da información muy valiosa para descubrir los patrones de movilidad de determinados colectivos y modos de transporte. Cuanto mayor sea la información asociada a dichos tracks, más aspectos podremos explorar y mayor será el poder explicativo de nuestros análisis. Este poder explicativo estará directamente ligado a la capacidad de predicción de los modelos derivados de dichos análisis.

Tomemos un estudio de caso que nos ayudará a ilustrar el uso de tracks GPS para la generación de modelos predictivos. La investigación desarrollada por Romanillos y Gutiérrez (2019) estudió más de 6.000 tracks GPS relativos a rutas ciclistas en la ciudad de Madrid y desarrolló varios modelos de regresión que analizaban la velocidad de los ciclistas, en función de un amplio conjunto de variables. Tras procesar los tracks GPS originales, mediante un proceso de *mapmatching* y obtener las trayectorias de los viajes en bicicleta ajustadas a la red de calles, se pudieron incorporar variables relativas a las propiedades de la red vial, como la pendiente, el tipo de vía, la existencia o no de infraestructura ciclista, la densidad de tráfico, la velocidad del tráfico rodado, la densidad de semáforos o la cercanía a intersecciones. Por otro lado, debido a que los tracks GPS fueron recopilados a partir de voluntarios ciclistas que facilitaron datos personales y datos asociados a la ruta, para su trabajo en el marco de un proyecto de investigación, se pudieron también considerar variables como el género, la edad o el motivo de viaje.

Distintos modelos de regresión de mínimos cuadrados permitieron estudiar el impacto que cada una de estas variables tenían en la velocidad de los ciclistas, para cada tramo de calle, de cada uno de los viajes. Posteriormente se realizó otro modelo que estimara los tiempos de viaje de los ciclistas a lo largo de un número de recorridos correspondientes a un porcentaje (10%) de tracks GPS que se habían aislado de la muestra. La Figura 38 muestra la alta correlación entre los tiempos de viaje estimados por el modelo y los tiempos de viaje reales.





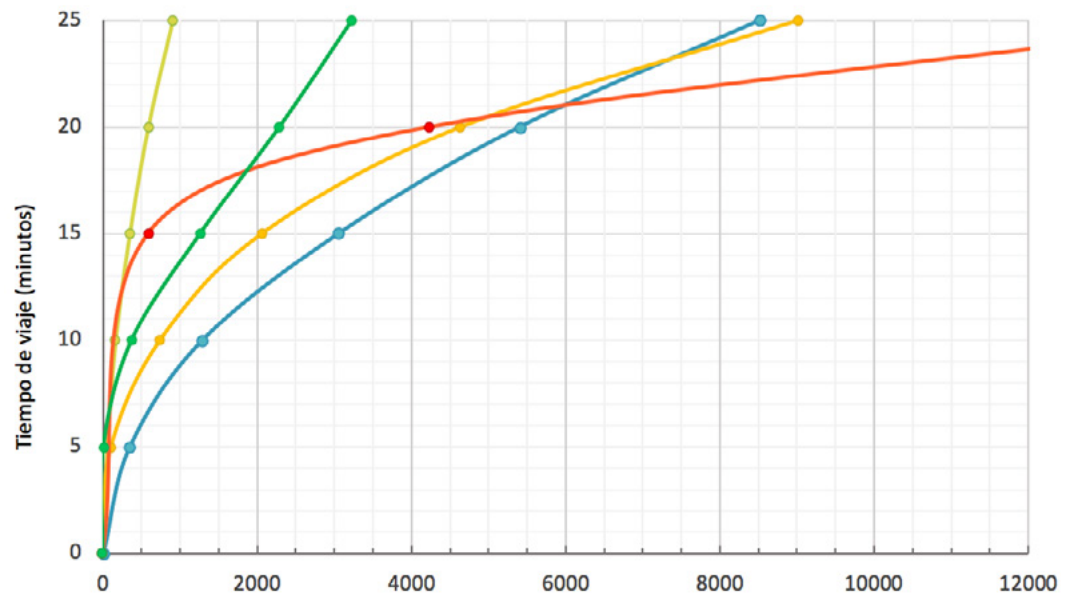
**Figura 38:** Correlación entre tiempos de viaje estimados y reales estimado por los modelos.  
Fuente: Romanillos y Gutiérrez, 2019

La Figura 38a muestra esta correlación para el caso de un modelo en el que sólo se tuvieron en cuenta las variables relativas a la red; la Figura 38b presenta la correlación cuando se introdujeron, además, las variables explicativas relativas al ciclista y al viaje (motivo del viaje). El elevado  $R^2$  obtenido en ambos casos (0,83 y 0,87 respectivamente) evidencia el alto poder predictivo de los modelos.

Una vez calibrados, estos modelos pueden ser utilizados para predecir los tiempos de viaje a lo largo de todos los segmentos de la red, lo cual tiene aplicaciones de mucho interés. En primer lugar, pueden desarrollarse enrutadores, herramientas que dado un origen y un destino pueden indicarnos cuál es la ruta más rápida, estimando, además, cuál será nuestro tiempo de viaje de manera bastante aproximada.

Otra aplicación derivada de la capacidad predictiva de tiempos de viaje de estos modelos es la posibilidad de realizar análisis de accesibilidad. Por ejemplo, podemos generar isocronas que nos permitan estimar qué área de la ciudad podemos cubrir a partir de un punto en un tiempo determinado. El estudio citado anteriormente realizó este trabajo para el caso de Madrid y comparó las isocronas obtenidas desde un punto central de la ciudad relativas a la movilidad ciclista (tanto usando bicicleta particular como las del sistema de bicicleta compartida) con las obtenidas a partir de otros modos de transporte, como el vehículo privado, el transporte público (incluyendo metro, tren de cercanías y autobús urbano) y caminando.

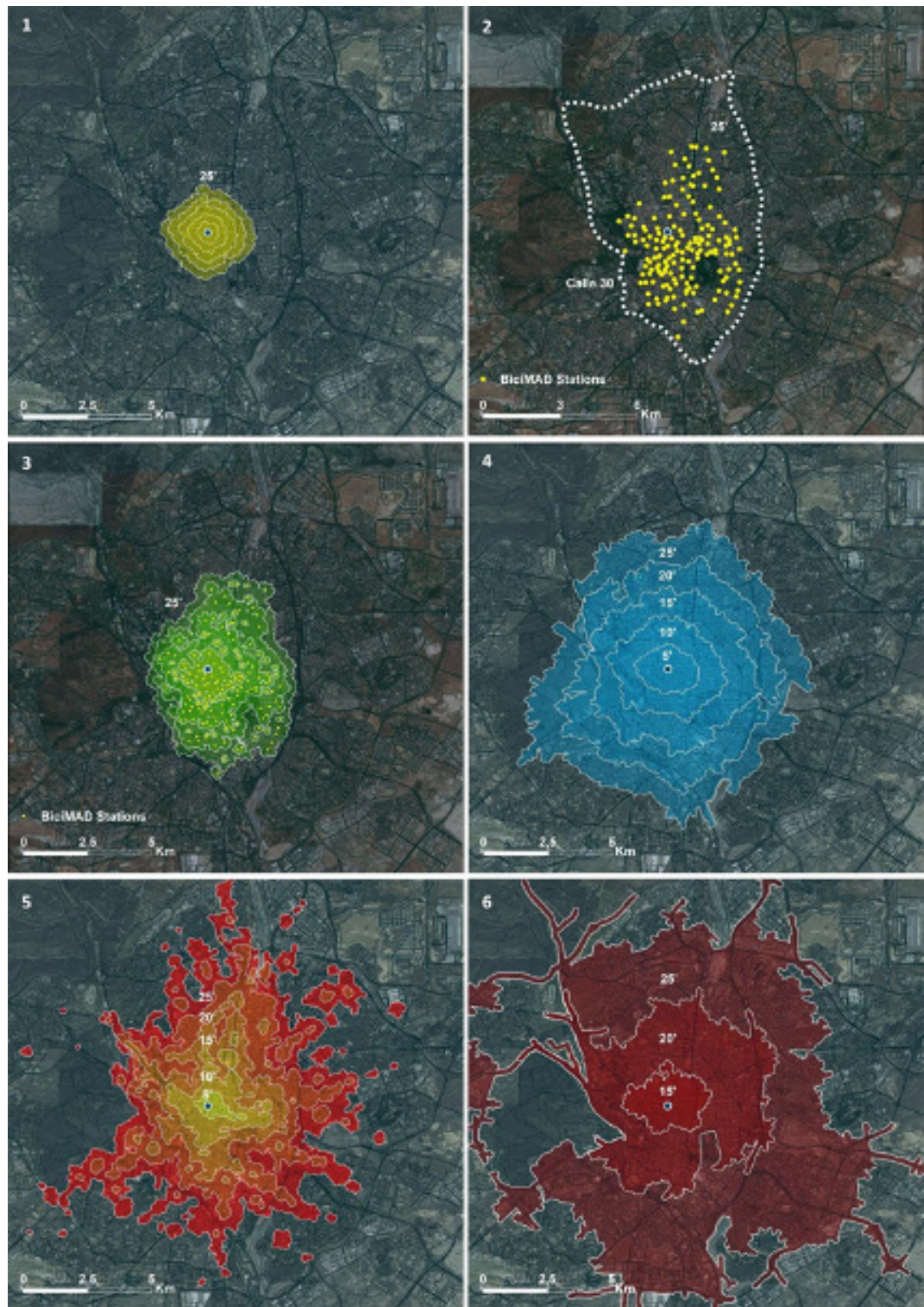
La Figura 39 compara el área cubierta (Hectáreas) de acuerdo con el tiempo de viaje y el modo de transporte, reflejando la competitividad entre los distintos modos de transporte. De la lectura del gráfico podemos destacar que la bicicleta es el modo más competitivo para viajes de hasta aproximadamente 21 minutos de duración. Por otro lado, la Figura 40 muestra las isocronas relativas a las áreas cubiertas de acuerdo con el tiempo de viaje y el modo de transporte.



**Figura 39:** Área cubierta (Ha) de acuerdo al tiempo de viaje y el modo de transporte. Fuente: Romanillos y Gutiérrez, 2019.



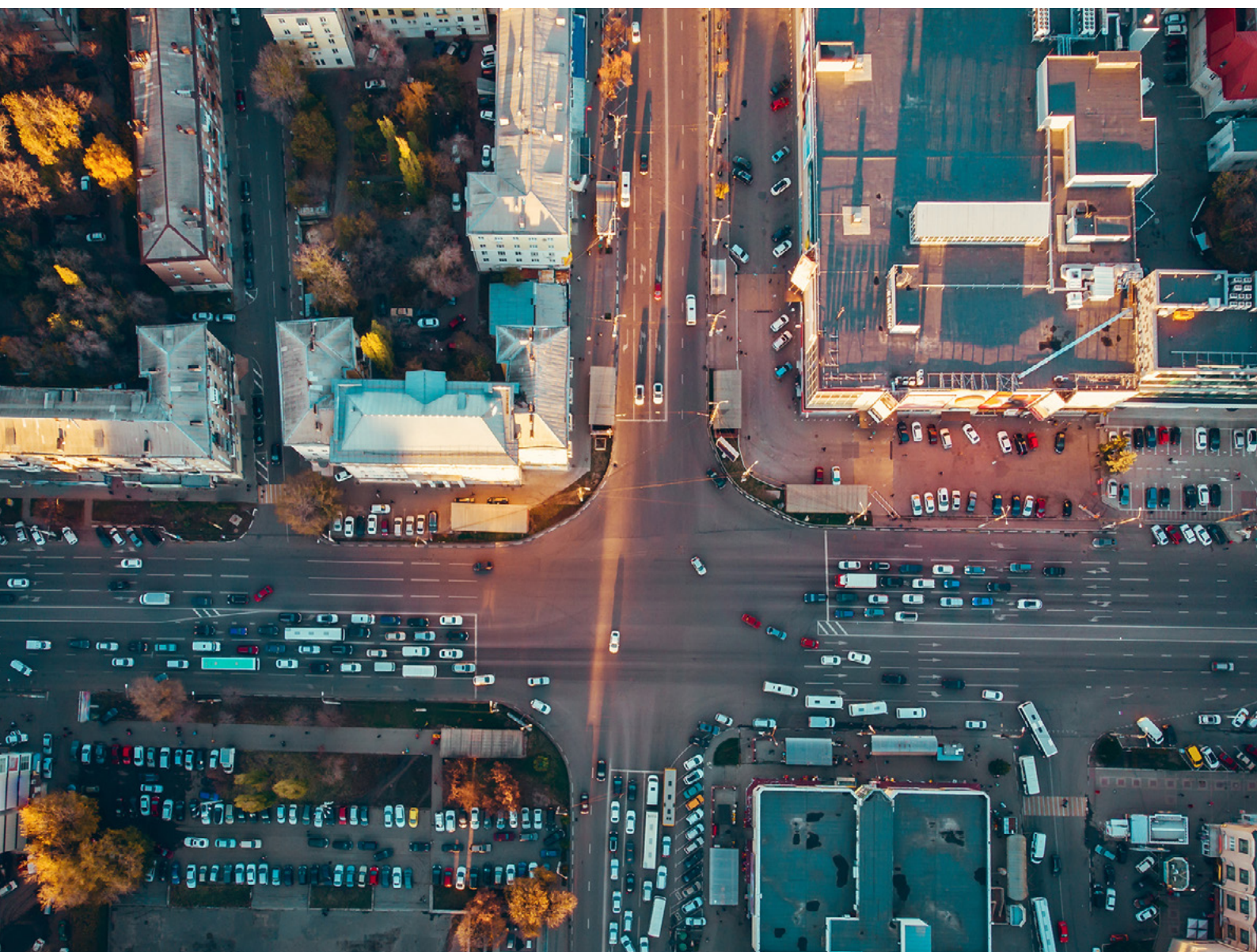




**Figura 40:** Isocronas relativas a las áreas cubiertas de acuerdo al tiempo de viaje y el modo de transporte (1. A pie, 2. Estaciones del Sistema de Bicicleta Compartida, 3. En Bicicleta Compartida, 4. En bicicleta particular, 5. En transporte público, 6. En coche privado). Fuente: Romanillos y Gutiérrez, 2019



Una vez calibrados, este tipo de modelos pueden utilizarse, por ejemplo, para simular los tiempos de viaje en escenarios futuros. Actualizando las propiedades de la red de manera que reflejen cambios relativos al desarrollo de nuevas infraestructuras (como carriles bici) o a la implementación de nuevas políticas (como el calmado o la supresión del tráfico rodado en determinadas vías), podremos correr de nuevo los modelos, cuyos resultados serán una estimación de los tiempos de viaje en ese escenario futuro.

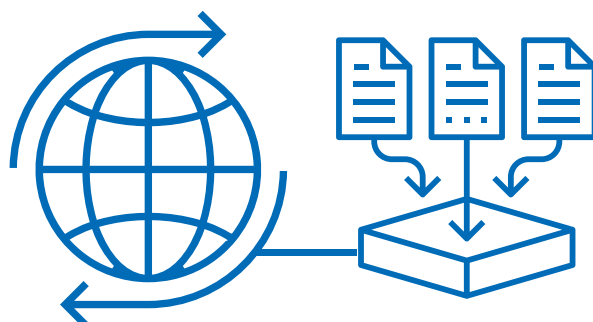




# 6

## Experiencias Internacionales





En este capítulo se incluyen algunos casos de estudio para ilustrar algunas de las posibilidades de esta fuente de datos. Estos casos de estudio, que muestran diversas experiencias en los contextos de Europa y América, se han seleccionado utilizando un triple criterio. Por una parte, debían referirse a distintos modos de transporte, para mostrar que son utilizables en distintos contextos y evitar una visión sesgada de las posibilidades de aplicación. Por otro lado, deberían presentar experiencias tanto con *tracks* (recorridos) como con datos de puntos de interés (datos puntuales). Finalmente, se debían abordar temáticas distintas, más allá de la obtención de matrices OD.

Así, se presentan cinco casos de estudio sobre 1) el análisis de los *tracks* de un sistema de bicicletas público para conocer los patrones de movilidad de sus usuarios, 2) la utilización de registros GPS de las flotas de taxis para estudiar la demanda en este modo de transporte, 3) el uso de datos de TomTom para el análisis dinámico de la congestión y la accesibilidad en las ciudades, 4) la combinación de datos de registros GPS de los autobuses, la tarjeta de transporte y otras fuentes para analizar la movilidad de los usuarios del autobús y 5) la integración de datos de fuentes oficiales y de Waze para el estudio de los accidentes de tráfico. La tabla siguiente presenta los casos de estudio, el modo, la temática, el tipo de datos GPS utilizados y el contexto espacial donde se han aplicado.

CASOS	MODO	TEMÁTICA	TIPO DE REGISTROS	CONTEXTO ESPACIAL
<i>Romanillos, Moya-Gómez, Zaltz-Austwick y Lamíquiz-Daudén (2018)</i>	Bicicleta	Matrices OD, patrones de movilidad	<i>Tracks</i>	Madrid (España)
<i>Yang, Franz, Zhu, Mahmoudi, Nasri y Zhang (2018)</i>	Taxi	Estimación de demanda de taxi	<i>Tracks</i> y otras fuentes	Washington (Estados Unidos)
<i>Moya-Gómez y García-Palomares (2017)</i>	Vehículo privado	Congestión y accesibilidad dinámicas	<i>Tracks</i> agregados	Ocho ciudades europeas
<i>Gschwender, Munizaga y Simonetti (2016)</i>	Autobús	Matrices OD, patrones de movilidad, agrupamiento de autobuses	<i>Tracks</i> , tarjetas y otras fuentes	Santiago (Chile)
<i>Dos Santos y Smarzaro (2017)</i>	Vehículo privado	Accidentes de tráfico	Waze y fuentes oficiales	Belo Horizonte (Brasil)

El primer estudio (Romanillos et al., 2018) analiza los patrones de movilidad de los usuarios del sistema de bicicletas públicas Bicimad de Madrid. Dado que las bicicletas están equipadas con localizadores GPS, cada viaje queda registrado como un *track*. El estudio muestra el proceso completo de tratamiento de los datos, hasta conseguir analizar la distribución de los viajes, las distancias, velocidades y tiempos de viaje, los ritmos temporales, etc., así como la carga de flujo ciclista de la red.

El segundo estudio (Yang, Franz, Zhu, Mahmoudi, Nasri y Zhang, 2018) se centra en la demanda de taxis en Washington. Se utilizan *tracks* de los taxis de la ciudad para geolocalizar los lugares de subida y bajada de viajeros. A continuación, se generan un conjunto de variables explicativas relativas a usos del suelo y oferta de transporte público. Con estos datos se ajustan modelos de regresión múltiple generales (uno para subidas y otro para bajadas) y por horas del día. Se concluye que la demanda de taxis es mayor en las zonas de transporte con mayor densidad residencial y de empleo, mayor tamaño medio de las manzanas y mayor desequilibrio población/empleo. Así mismo, esta demanda aumenta con la accesibilidad a las estaciones de metro (lo que sugiere la existencia de complementariedad entre ambos modos), pero disminuye en zonas con mayor número de paradas de autobús (relación de competencia entre modos).

El tercer caso (Moya-Gómez y García-Palomares, 2017) realiza un análisis comparativo de la congestión y la accesibilidad en ocho ciudades europeas. Para ello se utilizan los datos de TomTom Speed Profiles, que muestran la congestión de cada tramo de la red en intervalos de 5 minutos en las ocho ciudades seleccionadas. Esta alta resolución temporal permite introducir la variable congestión de forma dinámica en los análisis de accesibilidad.

Los indicadores dinámicos de accesibilidad se calculan en intervalos de 15 minutos, lo que permite obtener una visión continua de las variaciones en la accesibilidad debidas a la congestión. Se obtienen perfiles temporales y mapas de accesibilidad en las ocho ciudades.

El siguiente caso (Gschwender, Munizaga, y Simonetti, 2016) analiza la movilidad de la población en autobús en Santiago de Chile a partir de los registros GPS de los vehículos, la tarjeta de transporte y otras fuentes adicionales. Con la combinación de las dos primeras fuentes y una cartografía de las líneas de autobuses se obtuvieron matrices OD. Para caracterizar a los usuarios, se utilizó de forma complementaria información socioeconómica de los lugares de residencia de los mismos, inferidos a partir de sus validaciones en las primeras horas de la mañana. Los *tracks* de los autobuses sirvieron, además, para estudiar los procesos de agrupamiento de autobuses, que tanto afectan a la operación de los mismos.

Finalmente, el quinto caso (Dos Santos, Junior, y Smarzaró, 2017) se centra en el análisis de los accidentes de circulación en la ciudad brasileña de Belo Horizonte. El estudio compara los datos de accidentes de Waze (geolocalizados a través del GPS de los teléfonos móviles de los usuarios que informan de los mismos) con los recogidos en fuentes oficiales (obtenidos a partir de los informes policiales). El número de accidentes reportados por ambas fuentes de datos es parecido, pero muy pocos de esos accidentes son coincidentes. Los datos oficiales recogen mejor los accidentes en los que se ven involucradas las compañías aseguradoras, mientras que los de Waze se refieren más a los accidentes que causan retenciones en el tráfico, particularmente en las principales vías metropolitanas. Se concluye, por tanto, que ambas fuentes de datos son complementarias.



## 6.1 Visualización y análisis de rutas a partir de datos de *tracks* de bicicletas públicas<sup>8</sup>

### **SOBRE EL CASO DE ESTUDIO**

Con el objetivo de promover una movilidad más sostenible, a lo largo de los últimos años muchas ciudades en el mundo, han impulsado la movilidad ciclista, desarrollando nuevas infraestructuras ciclistas, promoviendo nuevas políticas que favorecieran el uso de la bicicleta en viajes más allá de los realizados por ocio o deporte y, en muchos casos, implementando sistemas de bicicleta compartida.

Madrid es una de estas ciudades. Con una cultura ciclista emergente, con relación a otras capitales europeas, e incluso en comparación con algunas ciudades españolas, la ciudad de Madrid ha promovido en los últimos años la construcción de carriles bici y ciclo calles y, ha implementado, un sistema de

<sup>8</sup> Romanillos, G., Moya-Gómez, B., Zaltz-Austwick, M., & Lamíquiz-Daudén, P. J. (2018). The pulse of the cycling city: visualising Madrid bike share system GPS routes and cycling flow. *Journal of Maps*, 14(1), 34-43.

bicicleta compartida. Con ello se ha incrementado considerablemente el uso de la bicicleta, que en los últimos 10 años ha pasado de un 0,22% a un 1,20% en el reparto modal de los viajes (Romanillos, 2018).

El servicio de bicicleta pública compartida de la ciudad de Madrid, llamado BiciMAD, comenzó a operar en junio de 2014, con 1.560 bicicletas y 123 estaciones de aparcamiento. Tras una primera fase de expansión en el año 2015, la ciudad está desarrollando su segunda expansión. Una vez finalizada ésta, BiciMAD contará con más de 200 estaciones y operará con un total de 2.500 bicicletas. El servicio cuenta actualmente con más de 60.000 usuarios activos y registra cerca de 350.000 viajes mensuales.

Por las características de su servicio, BiciMAD es un moderno sistema de bicicleta pública de los denominados de cuarta generación. La totalidad de la flota está constituida por bicicletas de asistencia eléctrica (e-bikes o pedelecs), que pueden impulsar a los viajeros hasta una velocidad límite de 25 km/h. El acceso al servicio se realiza mediante una tarjeta de usuario que puede ser tramitada y recogida en cualquier momento en el tótem de todas las estaciones del sistema. Por otro lado, el BiciMAD consta de una aplicación online que ofrece diversos servicios. A través de ella los usuarios pueden darse de alta en el servicio, comprar su tarjeta y recargarla. También pueden conocer en tiempo real, a través de un mapa online, el grado de ocupación de las estaciones y la disponibilidad de bicicletas, lo que les permite planificar su viaje con mayor facilidad. La aplicación también permite a los usuarios consultar los viajes realizados o el crédito disponible en la tarjeta. Otra de las características de los sistemas de bicicleta pública de cuarta generación que posee BiciMAD es el establecimiento de una política de tarifas que estimule los viajes a las estaciones de mayor o menor ocupación, favoreciendo el reequilibrio del sistema.

## EL DESAFÍO

En su apuesta por la promoción de la movilidad ciclista, Madrid está determinada a seguir invirtiendo en nuevas infraestructuras y políticas que estimulen el cambio modal deseado. Para ello, los planificadores y los tomadores de decisiones necesitan analizar la movilidad ciclista existente y evaluar el impacto que ha tenido la inversión realizada, para así poder promover medidas más eficientes ¿Cuál es el uso que tienen los carriles bici? ¿Cuáles son las arterias más importantes de la ciudad en cuanto a flujo ciclista? ¿Están funcionando las medidas de calmado de tráfico? ¿Qué intensidad de uso está teniendo el sistema de bicicleta pública? ¿Qué estaciones están siendo más o menos utilizadas?

Estas y muchas otras preguntas requieren ser contestadas para diseñar los nuevos planes de movilidad ciclista, planificar las nuevas infraestructuras y plantear nuevas y más eficientes políticas. Ahora bien, ¿tenemos la información necesaria para responder a estas cuestiones?

## LA SOLUCIÓN

Hasta hace pocos años, los únicos datos disponibles en la ciudad de Madrid relativos a la movilidad ciclista consistían en encuestas y aforos, mediciones del flujo ciclista en determinados puntos estratégicos, en días específicos y con poca frecuencia. Con estos datos, era imposible contestar a las cuestiones planteadas, especialmente dada la rapidez con la que se habían producido los

cambios en la ciudad de Madrid, algo que de partida chocaba con el ritmo con el que se tomaban los datos anteriores.

Con la implementación de BiciMAD se abrió una oportunidad única para explorar la movilidad ciclista desde una nueva perspectiva, dado que todas las bicicletas del sistema cuentan con un dispositivo de geolocalización GPS que registra las trayectorias de sus desplazamientos. Considerando que actualmente BiciMAD cuenta con más de 60.000 usuarios activos y que el servicio registra unos 350.000 viajes mensuales, la cantidad de datos disponibles acerca de la movilidad de los ciclistas no tenía precedentes.

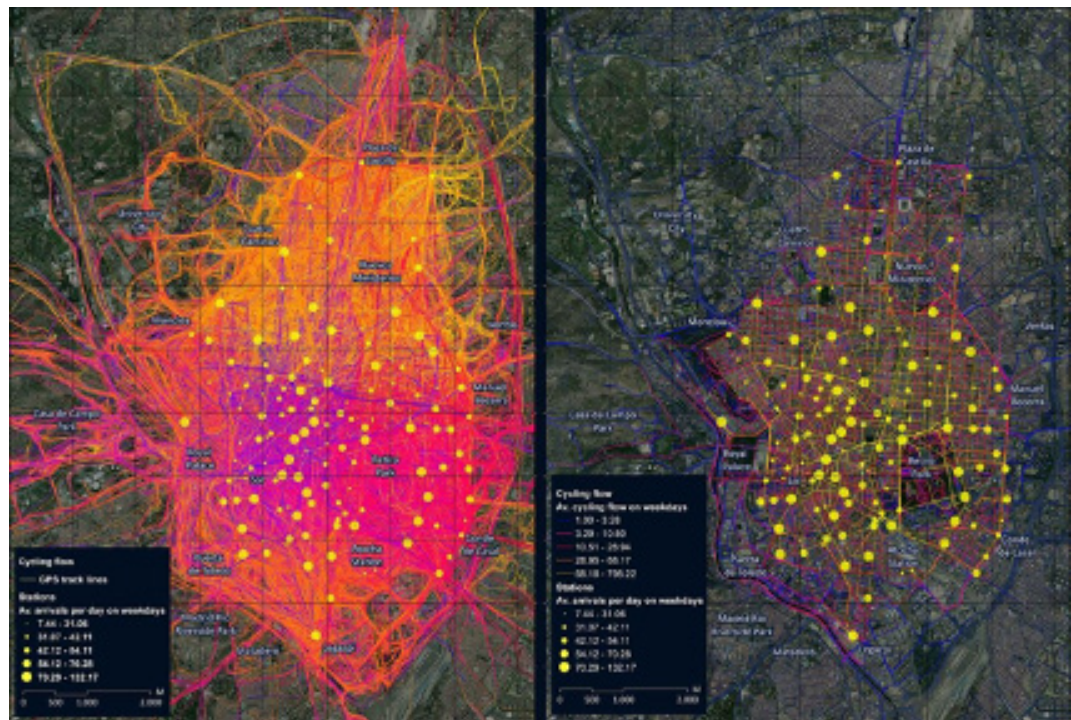
Con base en los *tracks* GPS registrados por BiciMAD, los investigadores (Romanillos et al., 2018) analizaron los patrones de movilidad del sistema y revelaron cuáles eran las vías de mayor flujo ciclista en distintos periodos de tiempo. La muestra de datos seleccionada correspondía con la totalidad de los viajes registrados por el sistema en abril de 2017, lo que suponía un total de 253.556 *tracks* GPS.

Los datos de operación del servicio de bicicleta compartida han sido facilitados por la Empresa Municipal de Transporte de Madrid (EMT), empresa que actualmente gestiona BiciMAD. El conjunto de datos sobre el que se realizó el estudio consiste en un total de 253.556 *tracks* GPS, que se corresponden con las rutas registradas por el sistema en abril de 2017. De ellos hubo que descartar los viajes derivados de la redistribución de bicicletas que realiza la propia empresa, utilizando finalmente para el análisis los 230.238 viajes restantes. De estos viajes, 10.289 fueron realizados por usuarios ocasionales y 215.371 por usuarios frecuentes. Para cada viaje se disponía de información sobre las estaciones y el anclaje en origen y destino, la duración del viaje en segundos, el momento de inicio (agregado por horas, para proteger el anonimato de los usuarios), el tipo de usuario que registra el viaje (ya sea usuario frecuente, usuario ocasional, empleado del servicio o desconocido) y un rango de edad, distinguiendo seis grupos: 0-16, 17-18, 19-26, 27-40, 41-65 y mayores de 65 años.

El primer trabajo fue el de procesamiento de los *tracks* GPS. Los ficheros originales en formato JSON se importaron como GeoJSON en un entorno de SIG, usando el lenguaje de programación Python y el software libre Mongo DB. Después, los *trackpoints* exportados como GeoJSON fueron exportados como entidades de punto con el software libre QGIS, con el que se procedió a obtener las tracklines a partir de los puntos anteriores.

La segunda tarea en cuanto a procesamiento de los datos fue el ajuste de los *tracks* GPS a la red de calles de la ciudad de Madrid (mapmatching). La resolución temporal media de los *tracks* GPS era de 75 segundos, una resolución baja en comparación con la resolución temporal típica en la mayor parte de las aplicaciones GPS de uso deportivo, cuya resolución temporal suele ser de entre 2 y 3 segundos. Por ello, en este caso, el algoritmo de mapmatching consistió en estimar el camino más corto entre cada par de trackpoints. Dado que en 75 segundos de intervalo no hay muchas rutas alternativas entre dos puntos en la ciudad (a menos que la trama urbana sea muy densa y de pequeña escala), el ajuste de los *tracks* a la red vial se produjo con bastante precisión. La Figura 41 muestra en el mapa de la izquierda el conjunto de las rutas representadas según los tracklines originales; el mapa de la derecha muestra, en cambio, las rutas ya ajustadas al viario. Los puntos representan las bases ciclistas, con un diámetro proporcional al número de llegadas medio diario.





**Figura 41:** Rutas ciclistas representadas de acuerdo a los tracklines GPS originales (izquierda) y como trayectorias ajustadas a la red vial tras el proceso de mapmatching (derecha). Fuente: Romanillos et al., 2018

## BENEFICIOS

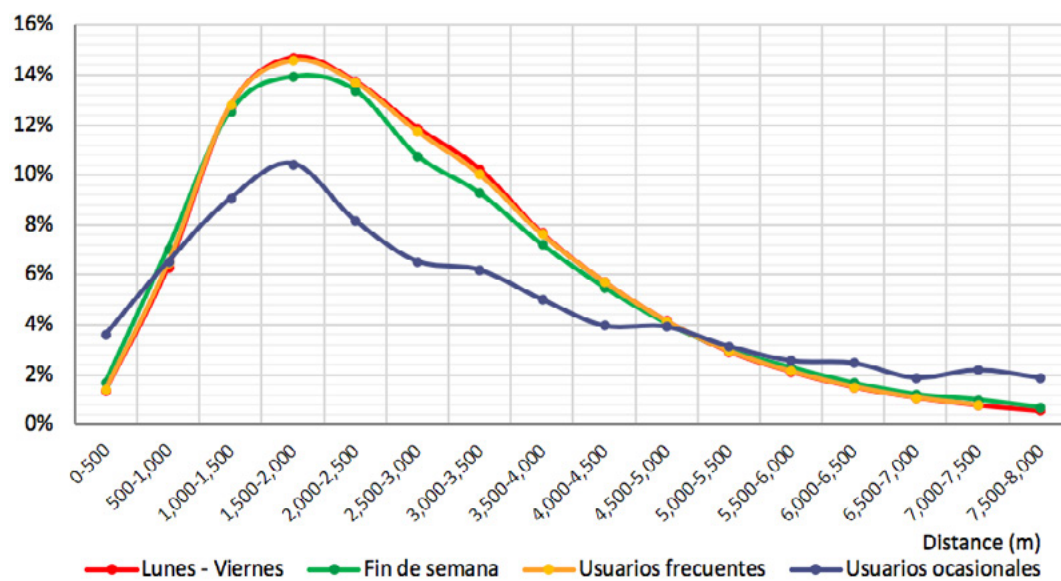
Una vez las trayectorias correspondientes a los viajes ciclistas estaban ajustadas a la red, se pudo realizar un análisis de los patrones de movilidad de los usuarios en cuanto a distancia, tiempo y duración media de los viajes. La medición de estas variables a partir de los tracklines originales habría dado lugar a errores importantes, especialmente teniendo en cuenta la baja resolución temporal de los datos, que daba como resultado representaciones muy simplificadas de los itinerarios reales.

Es importante analizar las distancias medias de los viajes realizados, así como conocer las velocidades de circulación. La Figura 42 muestra estos datos medios para los viajes realizados según el tipo de usuario y el tipo de día en el que se realiza el viaje. Los datos de distancia reflejan un patrón similar al de la duración del viaje, si bien las diferencias son mucho menos acusadas. La distancia media de los viajes realizados por los usuarios frecuentes es de 3.011 metros, mientras que la de los usuarios ocasionales es de 5.518 metros. En este sentido, es importante observar las diferencias en cuanto a las velocidades medias, de 14,29 km/h en los usuarios frecuentes frente a los 8,59 km/h de los ocasionales. Nuestra hipótesis al respecto es que buena parte de los usuarios ocasionales podrían ser turistas, cuyo propósito de viaje no es llegar a un único destino, sino realizar un itinerario con varios destinos para conocer la ciudad.

VIAJES	VELOCIDAD (KM/H)	DISTANCIA (M)
<i>Viajes totales</i>	14,06	3.104
<i>Viajes de usuarios frecuentes</i>	14,29	3.011
<i>Viajes de usuarios ocasionales</i>	8,59	5.518
<i>Viajes de lunes a viernes</i>	14,35	2.993
<i>Viajes en fin de semana y días festivos</i>	13,44	3.256
<i>Viajes de usuarios frecuentes, de lunes a viernes</i>	14,51	2.935
<i>Viajes de usuarios ocasionales en fin de semana</i>	8,55	5.641
<i>Viajes de usuarios frecuentes en fin de semana</i>	13,76	3.120

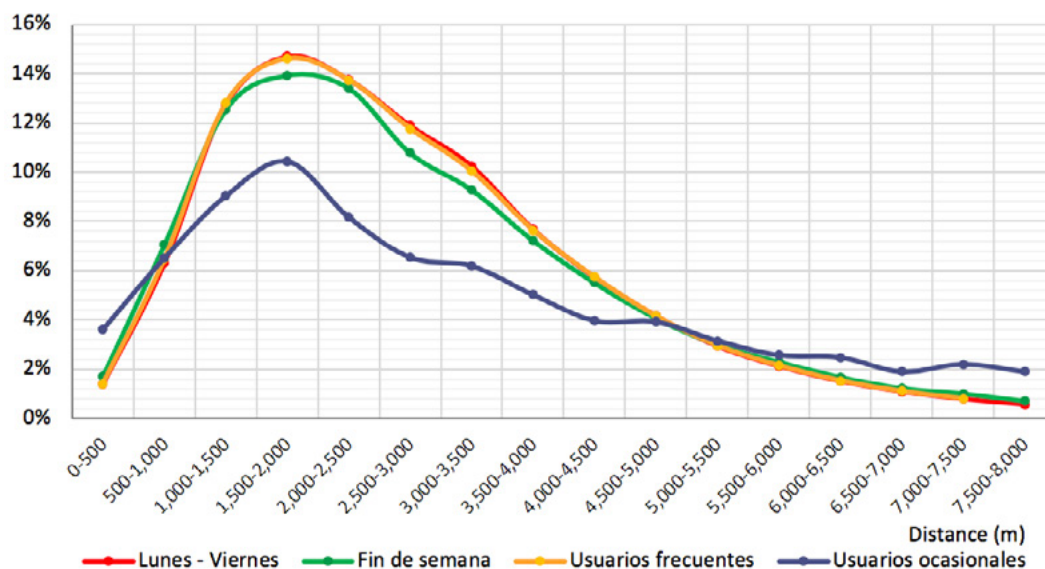
**Figura 42:** Distancias y velocidades medias de los viajes. Fuente: Romanillos et al., 2018

También es importante estudiar la distribución de los viajes de acuerdo a rangos de distancia, lo que permite conocer cuáles son las distancias más o menos comunes realizadas por los usuarios. La Figura 43 ilustra el porcentaje de viajes según distancia; la Figura 44, el porcentaje de viajes acumulados según distancia. Ambos gráficos distinguen entre los viajes realizados por usuarios frecuentes y ocasionales, así como entre los viajes realizados de lunes a viernes o durante fines de semana o días festivos. El primer gráfico refleja que el viaje más habitual entre los usuarios frecuentes tiene una distancia de entre 1 y 3 km. El segundo gráfico muestra que el porcentaje de viajes es muy bajo por debajo de 1 km y por encima de 5 km (en torno al 10% en ambos casos).



**Figura 43:** Porcentaje de viajes según distancia. Fuente: Romanillos et al., 2018





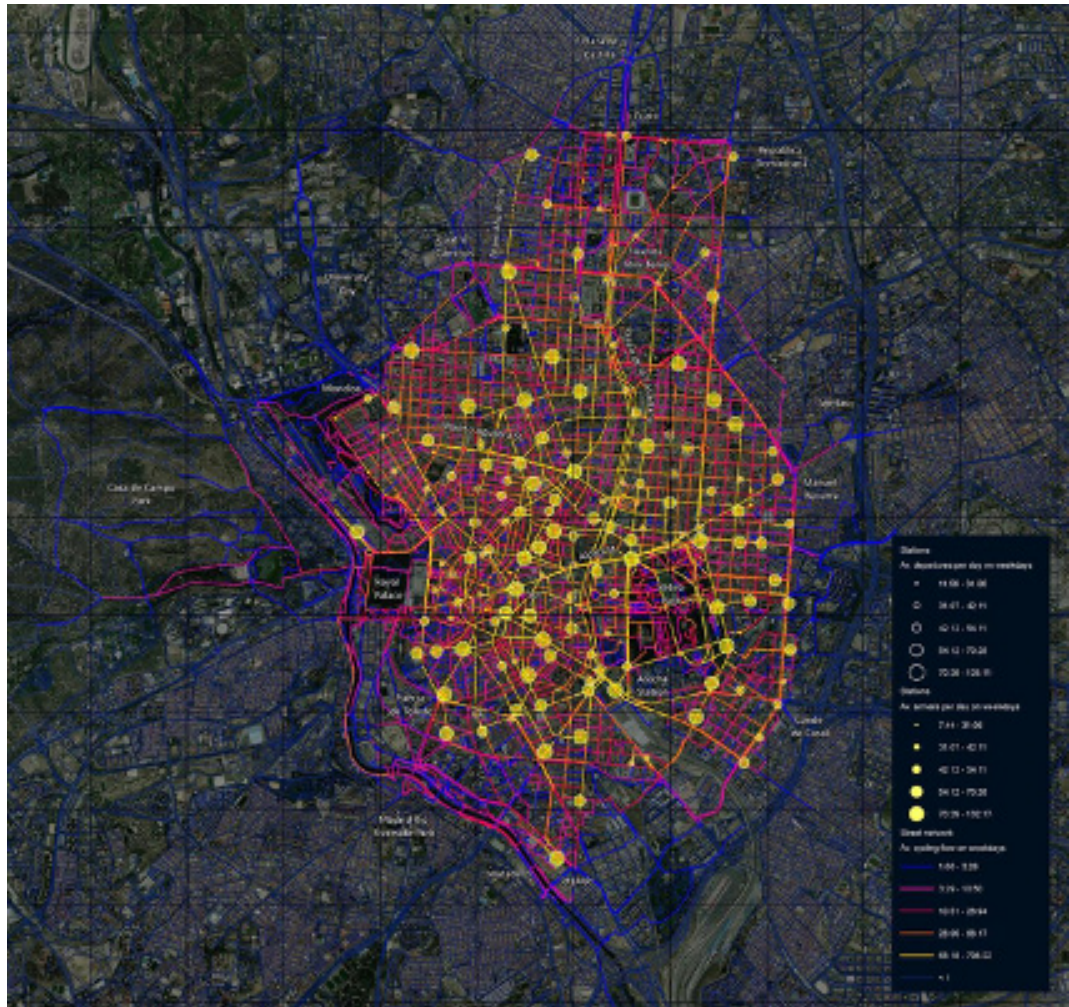
**Figura 44:** Porcentaje de viajes acumulados según distancia. Fuente: Romanillos et al., 2018

Finalmente, a partir de la suma de todos los itinerarios ciclistas ajustados a la red se calcularon los viajes coincidentes en cada tramo de calle de Madrid,





dando lugar a mapas de flujo ciclista medio diario por cada vía. La Figura 45 muestra el flujo medio diario ciclista de lunes a viernes.

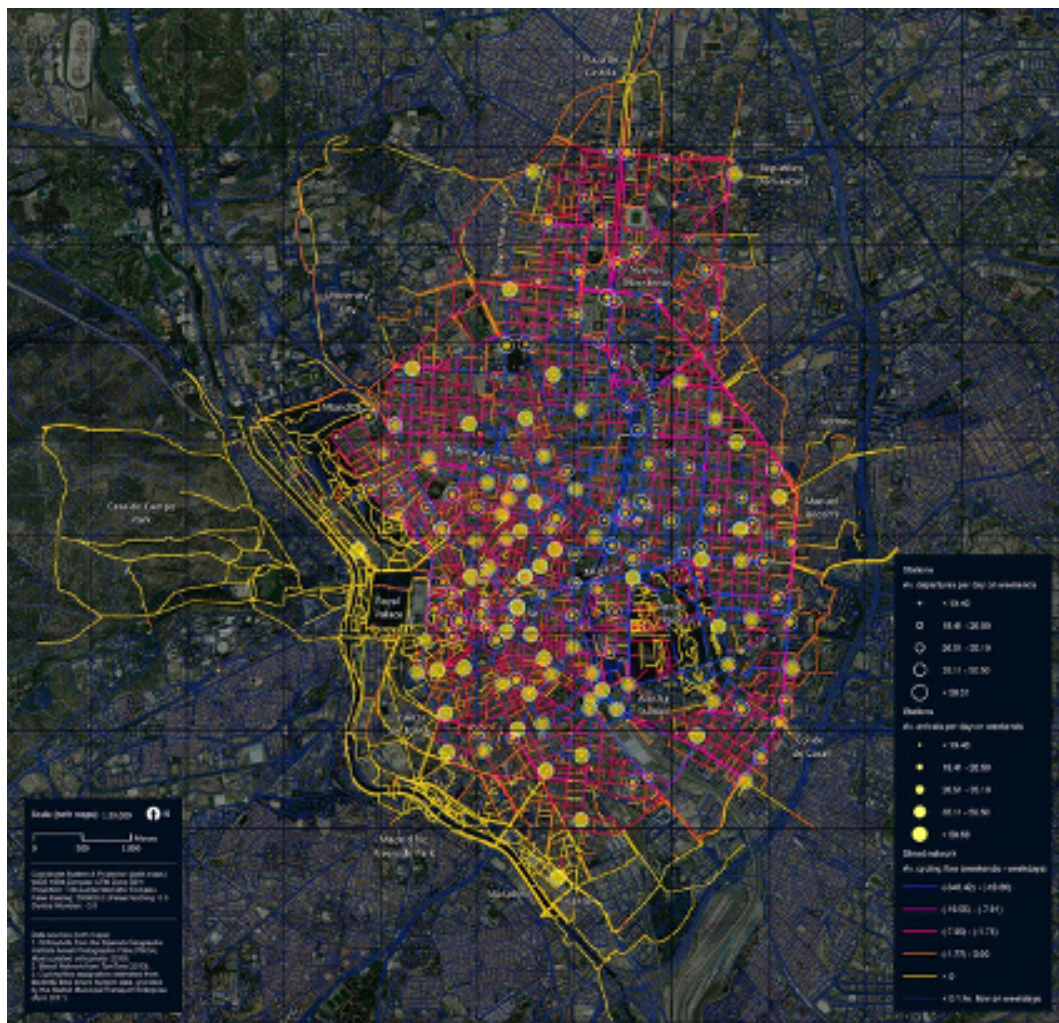


**Figura 45:** Estimación de flujo medio diario ciclista de usuarios de BiciMAD, de lunes a viernes.

Este mapa permite realizar un análisis visual de la distribución del flujo ciclista en la ciudad, identificar cuáles son las principales arterias de la ciudad en términos de volumen de ciclistas (representadas de rojo a amarillo), cuáles son las zonas con menor tráfico (representadas de rojo a azul, las de menor volumen), y tratar de sacar conclusiones con base en las propiedades de las vías. El mapa representa, a la vez, las bases ciclistas del sistema, con un diámetro proporcional al número de anclajes medios diarios, lo que refleja su capacidad de atracción de viajes, y permite ver las conexiones ciclistas de mayor fuerza entre los nodos. Fuente: Romanillos et al., 2018.

Gracias a la disponibilidad de todos los viajes registrados por el sistema durante un mes, se hicieron análisis de movilidad dinámicos, es decir, enfocados en el estudio de la movilidad no en un momento determinado (un análisis estático o una foto fija de la movilidad), sino para distintos tipos de día o para distintos momentos a lo largo del día. Por ejemplo, la Figura 46 permite analizar visualmente el flujo ciclista medio diario de usuarios de BiciMAD, en sábados y domingos, a partir del cálculo de la diferencia entre el flujo ciclista en fin de semana y entre semana. El mapa representa en amarillo las rutas que presentan un mayor flujo durante el fin de semana.





**Figura 46:** Estimación de flujo medio diario ciclista de usuarios de BiciMAD, sábados y domingos (Diferencia de flujo ciclista en fin de semana y entre semana. En amarillo las rutas con mayor flujo durante el fin de semana). Fuente: Romanillos et al., 2018

El análisis visual revela, claramente, un uso diferenciado de la ciudad en cuanto a flujo ciclista, destacándose el tráfico ciclista en el arco suroeste, correspondiente al parque lineal de Madrid Río, y al parque de La Casa de Campo. Por otro lado, se evidencia también la reducción del flujo ciclista al norte de El Retiro, en el barrio de Salamanca y a ambos lados del eje norte-sur del Paseo de La Castellana, que presenta un mayor uso del suelo de oficinas.

Finalmente, el estudio realiza una representación dinámica del flujo ciclista derivado de BiciMAD, a lo largo de un día tipo, a través de un video que ilustra más de 60.000 trayectorias ciclistas en movimiento. El video, que puede visualizarse en la web <http://huellaciclistademadrid.es/>, representa dos tipos de viajes: el de los ciclistas de lunes a viernes (en morado) y el de los ciclistas en fin de semana (en amarillo). Un gráfico dinámico en la parte baja muestra, además, el número de rutas activas para cada franja horaria. La Figura 47 representa un fotograma de dicho video.



**Figura 47:** Fotograma del video que representa el flujo ciclista derivado de BiciMAD a lo largo de un día. Fuente: Romanillos et al., 2018.

Este tipo de visualizaciones dinámicas permiten entender patrones de movilidad en determinados momentos del día, que pueden pasar desapercibidos en representaciones de datos agregados. Por otro lado, generan una mayor consciencia del volumen real de viajeros que se encuentra tras los datos agregados, lo que las convierte en potentes herramientas de análisis visual complementarias a las no dinámicas.



## 6.2 Estimación de la demanda de servicios de taxi<sup>9</sup>

### SOBRE EL CASO DE ESTUDIO

El taxi es un modo de transporte minoritario en el conjunto de la ciudad, pero que continúa teniendo un peso específico considerable en el centro, a pesar de la reciente expansión de compañías competidoras como Uber o Cabify. Además, la constante circulación de los taxis por la ciudad impacta sobre las condiciones del tráfico general. A pesar de la importancia de este modo de transporte, apenas existen modelos de demanda de taxi, posiblemente debido a la falta de datos precisos sobre los patrones espaciales y temporales de los viajes en taxis. Los dispositivos GPS instalados en los taxis proporcionan

<sup>9</sup> Yang, Z., Franz, M. L., Zhu, S., Mahmoudi, J., Nasri, A., & Zhang, L. (2018). Analysis of Washington, DC taxi demand using GPS and land-use data. *Journal of Transport Geography*, 66, 35-44.

una información de gran valor para este tipo de análisis, ya que los registros de cada viaje aportan, entre otras variables, la localización de los puntos de subida y bajada de los viajeros y la hora del día en que se producen.

El caso de estudio fue llevado a cabo en Washington (Estados Unidos). Se disponía de una base de datos proporcionada por el departamento del taxi de Columbia, con 15,2 millones de registros de viajes de taxis, que cubrían un período de un año. Cada registro incluía las coordenadas espaciales y temporales de los puntos de subida y bajada de cada viaje.

Se disponía, además, de una zonificación de transporte. Cada zona de transporte contenía información sociodemográfica y sobre usos del suelo: número de hogares, población residente, empleos (desagregados en cuatro categorías: industria, comercio, oficinas y otros), etc. Así mismo, se disponía de una cartografía con la localización de las estaciones de metro y las paradas de autobús.

### **EL DESAFÍO**

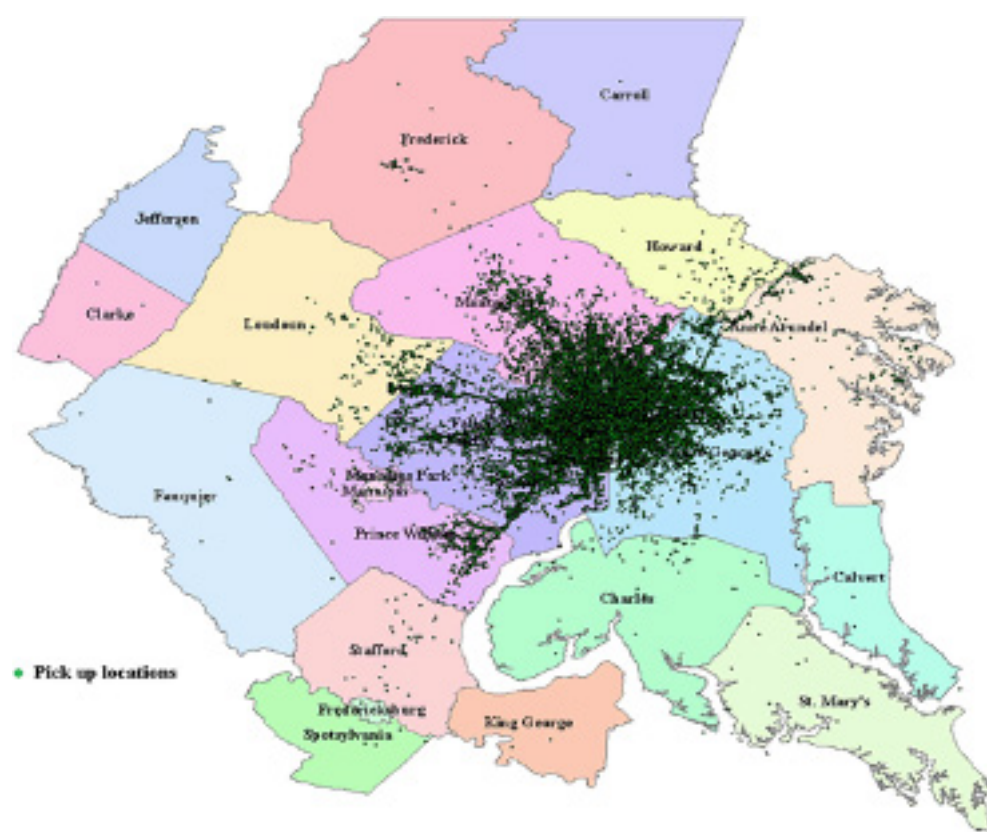
Este caso de estudio busca explicar la demanda de servicios de taxi (puntos de subida y bajada de viajeros) a partir de variables relativas a usos del suelo y accesibilidad a modos competidores. El desafío consiste en generar modelos de demanda de servicios de taxi utilizando datos de registros GPS de los vehículos y de la estructura de la ciudad para explicar la distribución espacial de la demanda y hacer predicciones. Para ello era necesario integrar las distintas fuentes de información disponibles y generar modelos explicativos, para subidas y para bajadas de viajeros, en las distintas horas del día.

### **LA SOLUCIÓN**

La solución se basó en una serie de modelos de regresión múltiple. La geolocalización precisa de los datos de subida y bajada de viajeros permitió cruzar estos datos con información sobre usos del suelo y oferta de modos competidores, utilizando para ello un SIG.

Como el fichero de subidas y bajadas de viajeros disponía de la latitud y longitud de cada una de ellas, con base en esos campos se crearon, con las funcionalidades de un SIG, dos capas de puntos: una para subidas y otra para bajadas (Figura 48). Posteriormente, estos datos de demanda fueron agregados espacialmente según zonas de transporte, de forma que para cada zona se dispusiera del número de subidas y bajadas de viajeros (total y según horas del día) (Figura 49). De esta forma se disponía de las variables dependientes para los modelos explicativos.

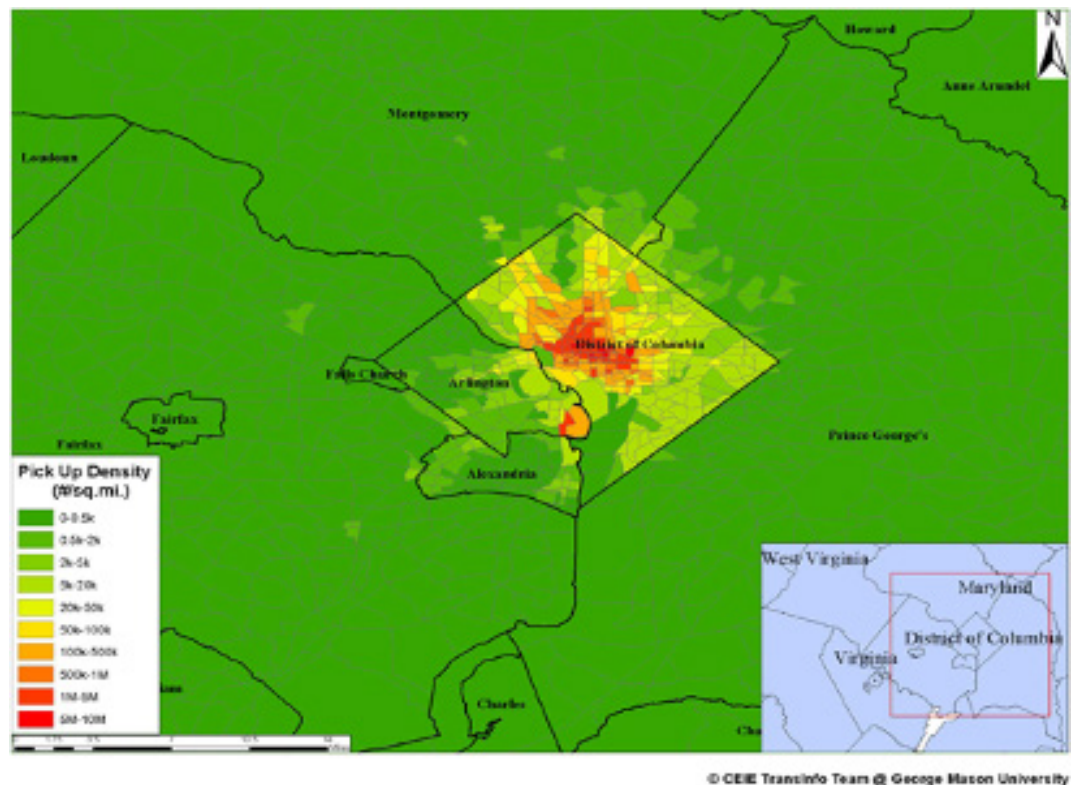




**Figura 48:** Localización de los puntos de subida de viajeros en el área metropolitana de Washington. Fuente: Yang et al., 2018







**Figura 49:** Densidad de subida de viajeros por zonas de transporte. Fuente: Yang et al., 2018

A partir de estos datos, se calcularon variables derivadas como densidades de empleo y población, la distancia de cada zona de transporte al Distrito Central de Negocios o un índice de entropía que reflejaba la mezcla de usos de suelo. Otras variables, como el número de manzanas o su tamaño medio, estaban directamente disponibles en la base de datos. Para calcular la accesibilidad a otros modos de transporte, se obtuvo el número de paradas de autobús dentro de cada zona de transporte y el número de estaciones de transporte comprendidas dentro de la zona, más una franja de media milla. También se consideraron los tres aeropuertos de la ciudad, ya que éstos influyen decisivamente en la demanda de taxi. Así, se creó una variable *dummy* para indicar la existencia de aeropuertos en las zonas de transporte y otra que mostraba la proximidad a los aeropuertos.

Los modelos de regresión múltiple generales (total de subidas y total de bajadas) presentaron unos  $R^2$  moderados (en torno a 0.5), con los signos de los coeficientes esperados. Las variables independientes que entraron en ambos modelos fueron básicamente las mismas. Se verificó que la demanda de taxis aumenta con la densidad residencial y la densidad de empleo (para las cuatro categorías de empleo). Así mismo, la demanda de taxis era mayor en las zonas de transporte en las que había un claro desequilibrio entre población y empleo (necesidad de desplazamientos más largos) y en las que las manzanas tenían mayor tamaño medio (las grandes manzanas son menos amigables para peatones, ciclistas y viajeros que acceden al transporte público). En cuanto a la presencia de oferta de otros modos de transporte, los modelos mostraron que la demanda de taxi era mayor en las zonas de transporte con mayor accesibilidad al metro, pero menor en las que tenían más paradas de autobuses. Ello sugiere que el taxi puede ser un modo complementario al metro, pero competidor con el autobús. Por otro lado, y como era de esperar,

los aeropuertos añaden una gran cantidad de demanda al taxi. La variable distancia al CBD no resultó significativa.

LOCALIZACIONES DE SUBIDA DE VIAJEROS			LOCALIZACIONES DE BAJADA DE VIAJEROS		
	COEFICIENTE	S.D.		COEFICIENTE	S.D.
<i>Manzanas:</i>	26.201*	15.126	<i>Manzanas:</i>	31.502**	10.541
<i>Metro:</i>	9039.917***	873.828	<i>Metro:</i>	7756.717***	608.974
<i>Bus:</i>	- 76.318	50.643	<i>Bus:</i>	- 101.503**	35.294
<i>Aeropuerto:</i>	66,143.91***	17,011.030	<i>Aeropuerto:</i>	166870***	11,855.050
<i>Población:</i>	0.173	0.180	<i>Población:</i>	0.205	0.126
<i>Densidad residencial:</i>	0.2496**	0.106	<i>Densidad residencial:</i>	0.606***	0.074
<i>Densidad comercial:</i>	3.213**	0.272	<i>Densidad comercial:</i>	2.040***	0.190
<i>Densidad de oficinas:</i>	0.006	0.036	<i>Densidad de oficinas:</i>	0.009	0.025
<i>Densidad industrial:</i>	1.052**	0.409	<i>Densidad industrial:</i>	0.720**	0.285
<i>Densidad otros:</i>	1.491**	0.177	<i>Densidad otros:</i>	1.102***	0.124
<i>Entropía empleo:</i>	4416.077	2734.586	<i>Entropía empleo:</i>	2817.835	1905.743
<i>Entropía empleo/ población</i>	- 7846.639**	3446.824	<i>Entropía empleo/ población</i>	- 2942.525	2402.104
<i>Constante</i>	- 5694.444**	2312.916	<i>Constante</i>	- 4678.491**	1611.879
<i>R = 0.4298</i>			<i>R = 0.5236</i>		

\* Estadísticamente significativo al 90%.

\*\* Estadísticamente significativo al 95%.

\*\*\* Estadísticamente significativo al 99%.

**Figura 50:** Modelos de regresión múltiple generales para estimar el total de subidas y el total de bajadas en taxi. Fuente: Yang et al., 2018

Los modelos generados según horas del día añadieron información adicional. En las horas de la noche los modelos presentaban coeficientes de determinación extremadamente bajos, lo que implica que la baja demanda

existente a esas horas tiene un claro componente aleatorio y no está ligada ni al uso del suelo ni a la oferta de transporte público. Como era de esperar, la existencia de aeropuertos era significativa sólo en las franjas horarias en las que estas infraestructuras estaban operativas.

### BENEFICIOS

Los resultados de este análisis pueden ayudar a los planificadores del transporte y a los responsables políticos a comprender mejor la demanda de servicios de taxi y a tomar mejores decisiones. La localización y capacidad de las paradas de taxi se puede adaptar a la demanda existente en función de los datos de registros GPS analizados. Pero, además, se pueden hacer estimaciones futuras sobre la influencia que tendrán en la demanda de taxi los nuevos desarrollos urbanísticos (según sea su densidad y la distribución de sus usos de suelo) o la nueva oferta de transporte (según se trate de modos competidores, como nuevas líneas de autobuses, o complementarios, como estaciones ferroviarias).

Este tipo de estudios podría llevarse a cabo en el futuro para conocer la demanda de los nuevos servicios de vehículos de movilidad compartida que se están expandiendo rápidamente por las ciudades (coches eléctricos, bicicletas *dockless* (sin bases fijas) patinetes y motos eléctricas, etc.), una vez que se pueda disponer de los *tracks* de estos vehículos.



## 6.3 Evaluación de los impactos de la congestión metropolitana en la accesibilidad usando datos de TomTom Speed Profiles<sup>10</sup>

### SOBRE EL CASO DE ESTUDIO

La congestión vial es uno de los grandes retos a los que se enfrentan las ciudades en Latinoamérica. Varias de ellas encabezan los ránquines de las ciudades más congestionadas del planeta. Bogotá, Lima y Ciudad de México aparecen entre las 10 ciudades más congestionadas del planeta, y Recife, Sao Paulo, Rio de Janeiro, Santiago de Chile y Buenos Aires entre las 50 primeras<sup>11</sup>. Realizar un buen diagnóstico de los impactos de la congestión sobre nuestras ciudades es fundamental para planificar actuaciones que permitan mitigarlos. Para ello resulta de gran utilidad la información sobre las redes de transporte privado, disponible a partir de los datos de las compañías de navegadores para automóviles.

Una de estas compañías es TomTom, que tiene una presencia significativa en los países latinoamericanos. TomTom, a través del producto TomTom

<sup>10</sup> Moya-Gómez, B. and García-Palomares, J.C. (2017). The impacts of congestion on automobile accessibility. What happens in the largest European cities?. *Journal of Transport Geography*, 62, 148-159.

<sup>11</sup> [https://www.tomtom.com/en\\_gb/traffic-index/ranking/](https://www.tomtom.com/en_gb/traffic-index/ranking/)

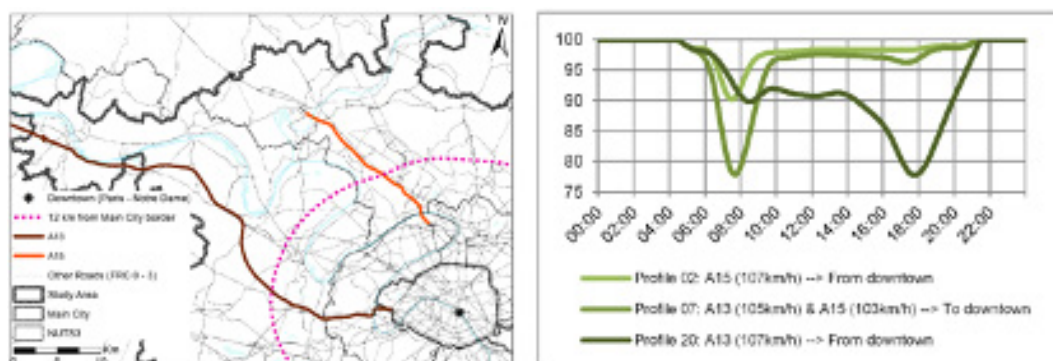
Speed Profiles, proporciona una red de transporte privado digital que puede ser utilizada en los softwares SIG habituales e incluye las velocidades de los vehículos en cada uno de los tramos de la red vial (excepto la local) cada cinco minutos. Los perfiles de velocidad se calculan a partir de millones de datos puntuales sobre velocidades reales tomados sobre los segmentos de las redes de carreteras a lo largo de dos años. Estos datos permiten realizar análisis dinámicos de accesibilidad considerando el efecto de la congestión, algo hasta ahora apenas estudiado. En este caso de estudio, se utiliza esta información para realizar un diagnóstico de los cambios en la accesibilidad en automóvil a lo largo del día. Esto permite incorporar una perspectiva dinámica de accesibilidad para el estudio de la congestión de la red de carreteras. El análisis se realiza en ocho áreas metropolitanas de la Unión Europea: Londres, París, Madrid, Berlín, Barcelona, Roma, Hamburgo y Milán.

### EL DESAFÍO

Este caso de estudio tiene como desafío realizar una propuesta metodológica que permita evaluar los cambios en la accesibilidad a lo largo del día, como resultado de la congestión de la red de carreteras. La accesibilidad cambia a lo largo del día como consecuencia de la congestión del tráfico. Sin embargo, este problema apenas había sido abordado hasta ahora, fundamentalmente debido a las limitaciones de las fuentes de datos tradicionales. Los estudios previos sobre accesibilidad en coche y su variación diaria se limitan al uso de escenarios estáticos, donde se compara una situación en flujo libre con otra de máxima congestión. A diferencia de estos análisis, que sólo abordan situaciones extremas en la operación de la red (hora pico y hora valle), en este trabajo se incorpora una perspectiva dinámica, que proporciona una visión del cambio temporal que tiene lugar a lo largo del día en el rendimiento de la infraestructura. Esta es una visión mucho más realista y precisa de la situación de cada ciudad. La accesibilidad dinámica es importante para las actividades y para la población, especialmente para aquellas actividades o grupos que son menos flexibles con respecto a las consecuencias de la congestión.

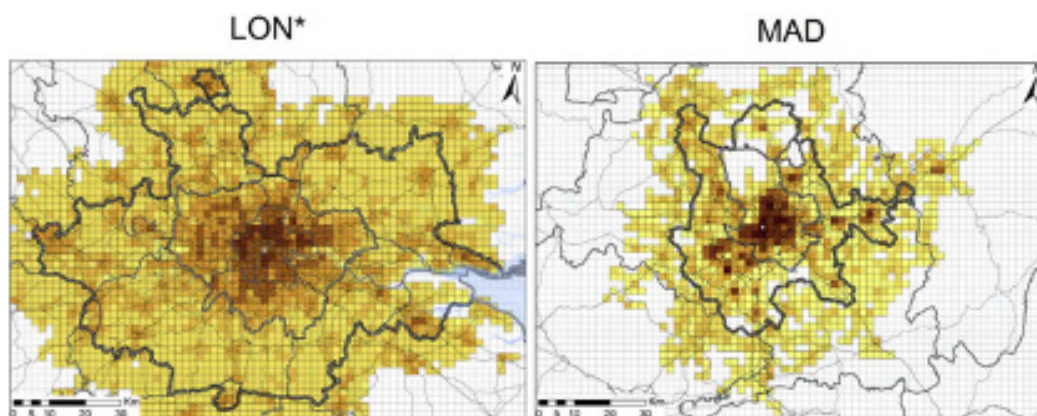
Para realizar el trabajo, se utilizaron los datos de velocidades de cada tramo de la red TomTom Speed Profiles. Los perfiles de velocidad están definidos como un porcentaje respecto a la velocidad del arco en flujo libre cada 5 minutos. Uno de los primeros desafíos es el manejo de una gran cantidad de datos, ya que además de lo detallado de las redes, éstas contienen hasta 98 perfiles predefinidos de las variaciones de la velocidad sobre la situación de flujo libre, que se van aplicando en los arcos de la red. A modo de ejemplo, en la Figura 51 se muestran tres perfiles para dos de las principales carreteras de acceso a la ciudad de París. La gran densidad de las redes utilizadas y el tamaño de las propias áreas de estudio hizo necesario simplificar ligeramente las redes utilizadas, eliminando de las redes la red local de menor importancia.





**Figura 51:** Ejemplo de estructura de los datos de TomTom Speed Profiles tomado de la red París. Fuente: Moya-Gómez y García-Palomares, 2017

Para el cálculo de los tiempos de viaje que alimentan el indicador de accesibilidad se dividió cada área metropolitana en una cuadrícula regular de celdas de  $2 \times 2$  km, que se usaron como las zonas de origen y destino de los viajes. Utilizar datos de cuadrícula regulares permitió llevar a cabo comparaciones de las áreas de estudio sin necesidad de tratamientos previos. A la vez, para cada una de esas cuadrículas se disponía de la distribución de la población en el año 2011, información que fue utilizada en el indicador de accesibilidad para representar el peso de las oportunidades en el destino del viaje. Como ejemplo, la Figura 52 muestra las áreas de estudio de Londres y Madrid y sus cuadrículas con la distribución de la población.



**Figura 52:** Áreas de estudio de Londres y Madrid, con la cuadrícula usada como zonas de origen y destino y la distribución de la población en las mismas. Fuente: Moya-Gómez y García-Palomares, 2017

## LA SOLUCIÓN

En el trabajo se propone el uso del indicador de accesibilidad potencial propuesto por Hansen (1959) para medir los efectos directos de la congestión en la accesibilidad territorial del automóvil, utilizando una función exponencial negativa. La razón para utilizar este tipo de función exponencial fue la transformación de todas las oportunidades en unidades potenciales y evitar, así, el problema del auto potencial. La fórmula utilizada es:

$$A_i^t = \sum_{j \in N} D_j \cdot e^{\beta \cdot c_{ij}^t}; \forall i \in N, t \in T$$

Donde  $A_i^t$  es el valor del potencial de accesibilidad en la zona  $i$  para el momento de salida del viaje  $t$ ;  $D_j$  es el potencial de la zona  $j$  (en este caso, la población total en 2011),  $e^{\beta \cdot c_{ij}^t}$  es una función de caída de la impedancia con la distancia, donde el parámetro  $\beta$  fue calibrado como -0,065.  $N$  es el conjunto de zonas del área de estudio (cada una de las celdas de cada ciudad).

Pero lo más novedoso de la ecuación es el tratamiento que se hace a la impedancia del viaje ( $c_{ij}^t$ ) entre cada zona  $i$  a cada zona  $j$ . En el trabajo la impedancia es el tiempo de viaje en minutos, y el coste de viaje es calculado utilizando la siguiente expresión:

$$c_{ij}^t = \sum_{m \in M} \sum_{e \in E} \alpha_{eij}^{tm} \cdot c_e^m; \forall ij \in G, t \in T$$

En el cual  $c_{ij}^t$  es el tiempo de viaje en minutos de la zona  $i$  a la zona  $j$ , y empezando el viaje en el instante  $t$ ;  $\alpha_{eij}^{tm}$  es una variable binaria que indica si el arco  $e$ , empezando su uso en el instante  $m$ , es usado por el viaje entre la zona  $i$  y la zona  $j$  que ha empezado en el instante  $t$ ;  $c_e^m$  es la impedancia esperada del arco  $e$ , al empezar su uso en el instante  $m$ ;  $T$  es el conjunto de instantes de inicio de viajes;  $G$  es el conjunto de viajes entre cualquier zona del área de estudio;  $E$  es el conjunto de arcos de la red del área de estudio; y  $M$  es el conjunto total de instantes posibles dentro del escenario.

Para realizar comparaciones entre las ciudades, se obtienen perfiles de accesibilidad global de cada una de ellas a partir de los resultados de accesibilidad de cada una de sus celdas. Esta accesibilidad global se calcula como una media ponderada de la accesibilidad en el conjunto de la cuadrícula en cada momento temporal.

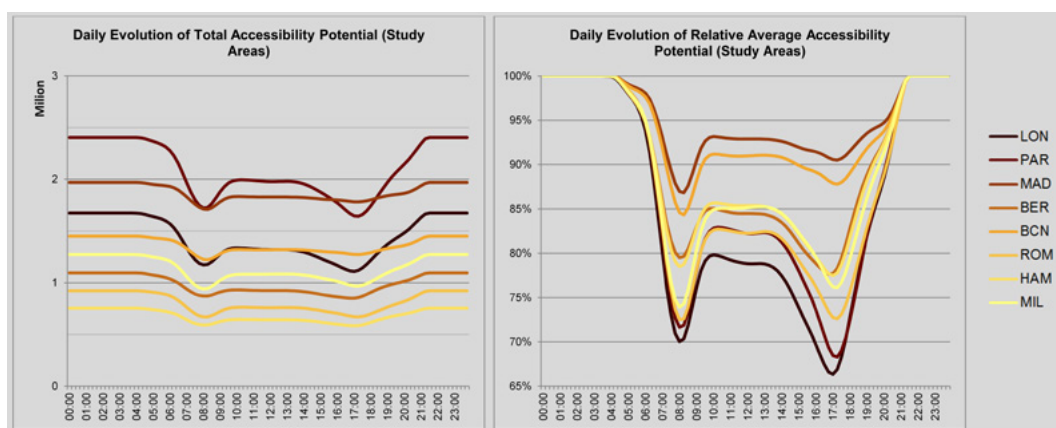
Además, se analiza la distribución espacial de los efectos de la congestión a nivel local en cada área metropolitana. Para ello se hacen mapas de resultados del valor de accesibilidad máximo (en flujo libre), la relación entre el valor de accesibilidad en flujo libre y los valores de las peores situaciones de accesibilidad en los picos de congestión de la mañana y la tarde, o se identifica para cada una de las zonas el momento de menor accesibilidad. Finalmente, se estudia la variación en la accesibilidad según distancia al centro de la ciudad durante los momentos de congestión en la mañana y en la tarde.

Todos los indicadores de accesibilidad se calculan cada 15 minutos en un entorno de SIG, donde se automatizan los análisis de redes y cálculo de accesibilidad, y se mapean los diferentes resultados.

## BENEFICIOS

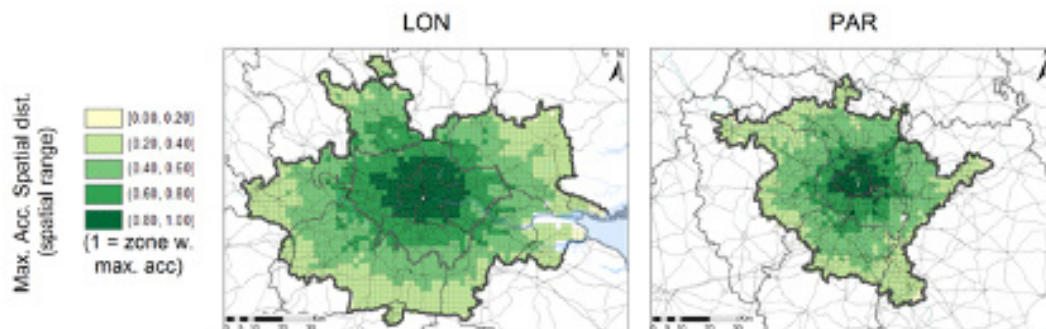
La metodología propuesta y el alto nivel de información de la red vial permiten obtener resultados muy detallados de la distribución de accesibilidad en cada una de las áreas metropolitanas analizadas. Así, es posible realizar comparaciones de los perfiles de accesibilidad diaria en cada ciudad (Figura 53). Estos perfiles permiten conocer el impacto de la congestión vial, mostrando la

intensidad en la pérdida de accesibilidad a lo largo del día. Por ejemplo, los resultados muestran una mayor accesibilidad en París y Madrid, por delante de Londres, a pesar de ser el área de mayor población. El modelo urbano más extenso y disperso de Londres, en contraste con París y Madrid, podrían explicar estas diferencias. Los perfiles de accesibilidad relativa muestran muy bien el impacto de la congestión en la accesibilidad, que es mucho más intenso en las dos grandes áreas metropolitanas europeas de Londres y París, medio en las ciudades italianas y alemanas y mucho menor en las dos grandes ciudades españolas. Así, Londres o París presentan pérdidas muy importantes de accesibilidad por la congestión en la mañana y, sobre todo, en la tarde (con caídas de más del 30% de su accesibilidad) y casi un 20% en las horas centrales del día. Por el contrario, en Madrid y Barcelona las pérdidas se concentran en la mañana, con una caída del 15%. A diferencia de París y Londres, Madrid o Barcelona apenas presentan un pico de pérdida de accesibilidad en la tarde.



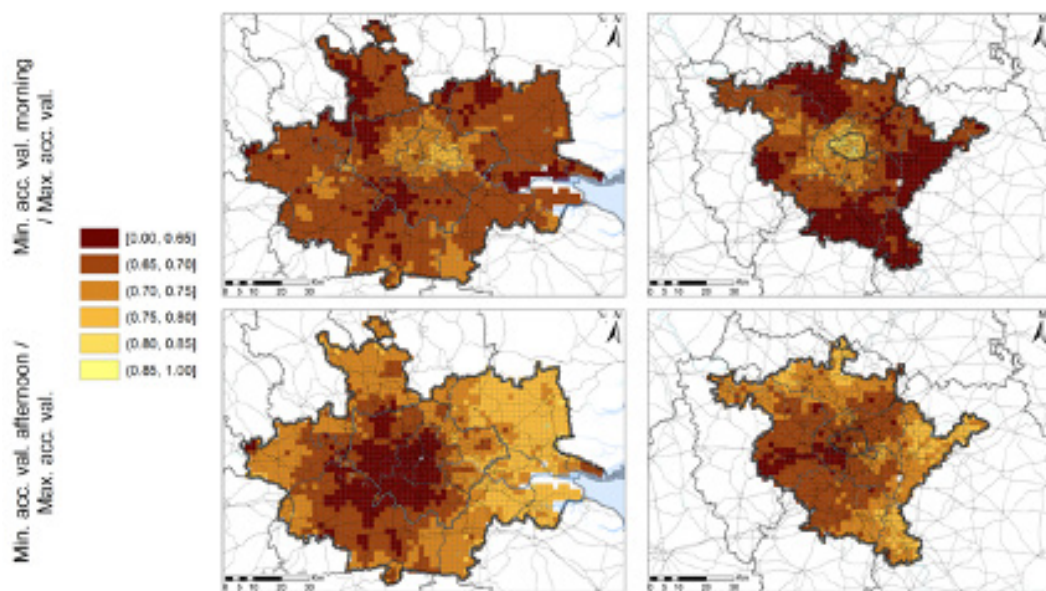
**Figura 53:** Evolución diaria del potencial de accesibilidad promedio (en %). Fuente: Moya-Gómez y García-Palomares, 2017

Además, los cambios en las velocidades de circulación de las vías, debidos a la congestión del tráfico, no tienen un impacto uniforme en el interior de cada una de las áreas de estudio (Figura 54). Para reflejar esos impactos, los autores realizan una cartografía de las distribuciones de accesibilidad en diferentes momentos del día. Se parte de mapas de accesibilidad potencial con flujo libre, que se toman como escenarios base para la comparación de las diferentes situaciones. Como es lógico, estas distribuciones de accesibilidad suelen presentar pautas concéntricas, con mejor situación en el centro y peor accesibilidad en la periferia (ver como ejemplo París y Londres, Figura 54).



**Figura 54:** Situación de accesibilidad potencial con flujo libre en Londres y París. Fuente: Moya-Gómez y García-Palomares, 2017

Tomando como base esa situación de partida, se puede comparar el impacto territorial de los momentos de máxima congestión. Como ejemplo, en la Figura 55 se muestran, nuevamente para Londres y París, las zonas con mayores pérdidas respecto a su situación en flujo libre en los momentos de mayor congestión de la mañana y la tarde. Puede verse cómo las periferias residenciales de ambas zonas metropolitanas sufren caídas de accesibilidad muy importantes en la mañana, mientras que esta situación se invierte en la tarde. Pero también en la periferia hay pérdidas de accesibilidad en la tarde en aquellos espacios en los que se han descentralizado zonas de empleo.



**Figura 55:** Situación de accesibilidad relativa respecto a la situación con flujo libre en Londres y París en los momentos de congestión de la mañana y la tarde. Fuente: Moya-Gómez y García-Palomares, 2017

En definitiva, este caso de estudio muestra las potencialidades del uso de la tecnología SIG y la nueva información proveniente de sensores como navegadores GNSS o de teléfonos inteligentes a la hora de mejorar los análisis tradicionales de la accesibilidad, introduciendo una perspectiva dinámica. Este tipo de análisis permite conocer el funcionamiento real de la ciudad y





el impacto de la congestión, aportando información de gran interés en su ordenación territorial.

## 6.4 Uso de datos de tarjeta inteligente y GPS: el caso del Transantiago<sup>12</sup>

### SOBRE EL CASO DE ESTUDIO

Una de las fuentes de mayor interés para el estudio de la movilidad urbana son las tarjetas inteligentes (smart cards) de transporte público. Numerosas ciudades latinoamericanas han implantado ya su uso. Una de las primeras fue Santiago de Chile, donde el sistema de transporte público de la ciudad las utiliza desde 2007 para administrar el proceso de cobro de tarifas en su famoso Transantiago (rebautizado desde marzo de 2019 como Red Metropolitana de Movilidad o simplemente como Red). Las tarjetas de transporte utilizadas en Santiago tienen el nombre de bip!. Además de los datos de las tarjetas bip!, los autobuses urbanos del sistema van equipados con dispositivos GPS que permiten conocer su ubicación en cada momento. Desde muy pronto, ha existido una colaboración estrecha entre la academia y los técnicos y gestores de la autoridad de transporte público, para desarrollar herramientas basadas en el procesamiento de las bases masivas de datos pasivos, obtenidas del sistema. El objetivo de esta colaboración es ayudar a conocer y analizar la movilidad en la ciudad y realizar una mejor gestión y planificación del sistema de transporte público.

El Red de la ciudad de Santiago es un sistema de transporte público integrado y multimodal (autobús y metro) que en la actualidad sirve a una población de alrededor de siete millones de habitantes. Está coordinado y supervisado por el Directorio de Transporte Público Metropolitano (DTPM), una agencia estatal que depende del Ministerio de Transporte y Telecomunicaciones. La operación de los servicios de autobuses se contrata a diferentes empresas privadas. En 2015 había 7 empresas, con entre 400 y 1.200 vehículos cada una. Metro es operado por una empresa pública. En general, el sistema tiene más de 6.500 autobuses, todos equipados con dispositivos GPS, que operan diariamente en una red con 68 km de vías segregadas, 150 km de calles reservadas o carriles exclusivos y más de 11.000 paradas de autobuses. Por su parte, la red de Metro, tiene actualmente siete líneas, 136 estaciones y una extensión de 140 km. Además, hay otras compañías privadas que prestan servicios complementarios al sistema, como la administración financiera o la propia gestión y venta de las tarjetas inteligentes y los servicios tecnológicos para las compañías de autobuses y de Metro.

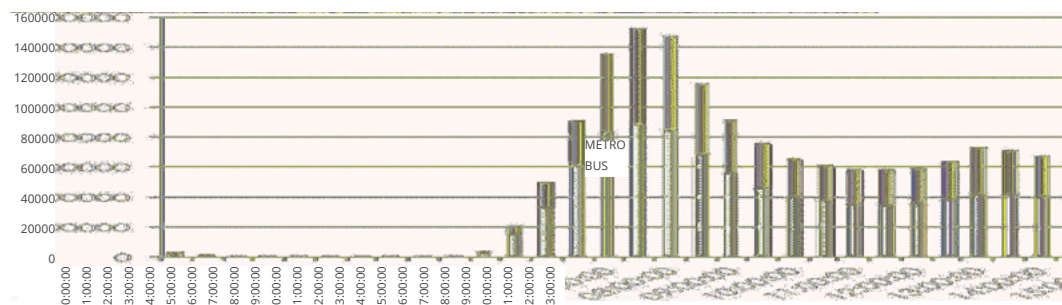
### EL DESAFÍO

Este caso de estudio tiene como desafío el desarrollo de métodos que permitan obtener información de la movilidad en transporte público, utilizando los datos

<sup>12</sup> Gschwender, A., Munizaga, M., y Simonetti, C. (2016). Using smart card and GPS data for policy and planning: The case of Transantiago. *Research in Transportation Economics*, 59, 242-249.

pasivos generados por las tarjetas inteligentes y los GPS de los autobuses urbanos. En concreto, se obtienen matrices de viajes origen-destino, perfiles de velocidad de los autobuses y toda clase de indicadores de calidad de servicio. Para mostrar la utilidad de la información conseguida, en el trabajo se presentan varios ejemplos de su uso específico para la política y planificación del transporte público en Santiago.

En el estudio se usan datos provenientes de dos fuentes principales. Por un lado, se utilizan los datos generados por los dispositivos GPS de cada uno de los autobuses, que envían una observación de su posición cada 30 segundos. En el momento de realización del estudio se generaban entre 80 y 100 millones de observaciones GPS por semana. Por otro lado, se trabaja con los datos provenientes de las tarjetas bip!. Esta tarjeta es la única forma de pago en autobuses y, en el momento del estudio, del 97% de los usuarios del metro. Las validaciones de las tarjetas inteligentes generaban en ese momento entre 35 y 40 millones de observaciones por semana. Además, se utiliza información complementaria, que incluye los recorridos de cada ruta de autobús, asignaciones de ruta, posición de paradas de autobús, posición de estaciones de Metro y posición de las estaciones de autobús. Toda esta información, sin tratar (en bruto), permite observar la demanda de la red a lo largo del tiempo (validaciones de bip!) y de la distribución de la flota en el espacio (movimientos de autobuses) (Figura 56).

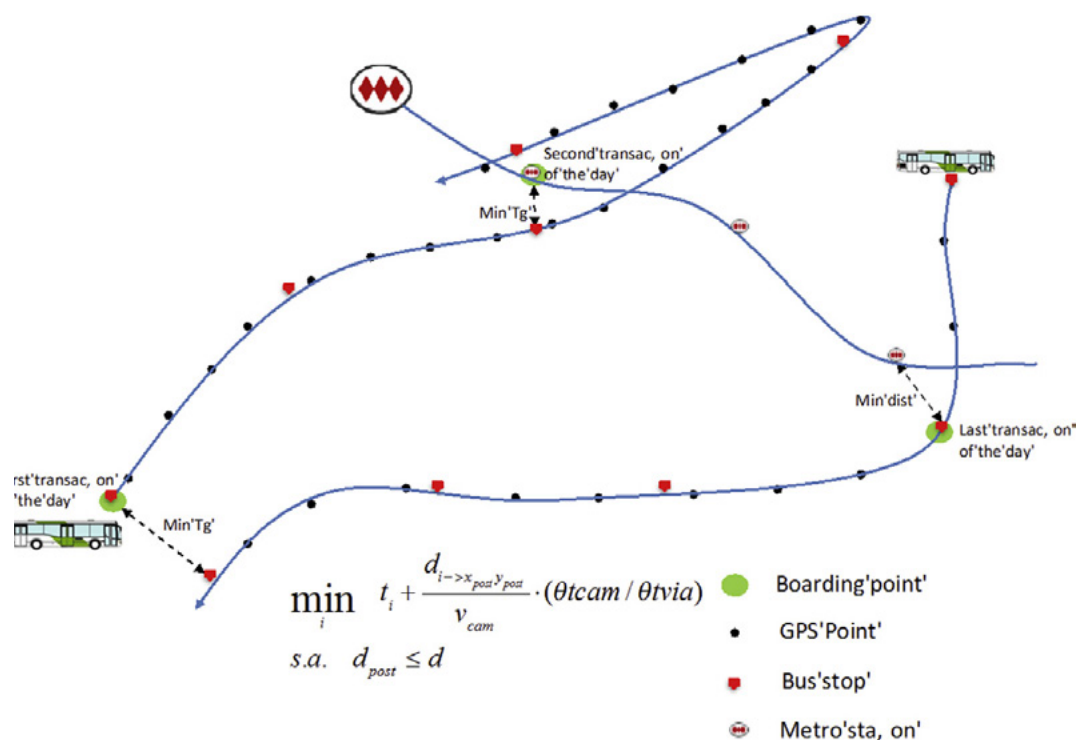


**Figura 56:** Distribución de las validaciones de la tarjeta bip! en días laborables (izquierda) y distribución de autobuses en un determinado momento (derecha). Fuente: Gschwender et al., 2016

## LA SOLUCIÓN

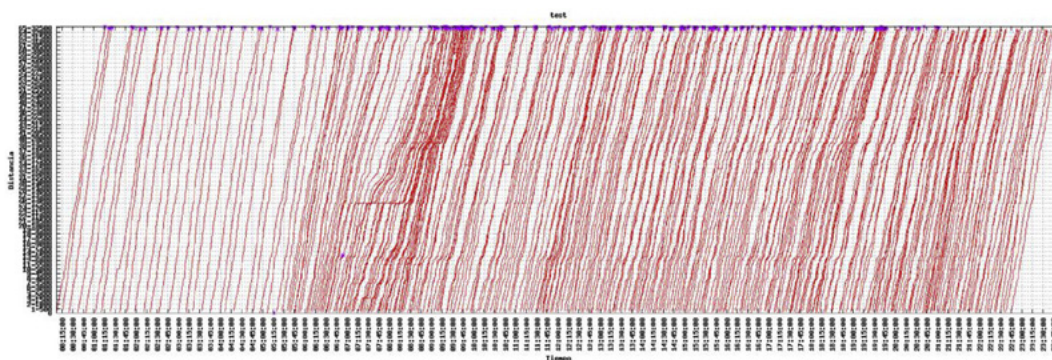
Los primeros procesos necesarios para obtener información de esos datos son los de asignar las localizaciones puntuales del GPS de cada autobús a sus rutas (mediante técnicas de mapmatching) y hacer, después, coincidir las transacciones de tarjetas inteligentes y esas bases de datos de posiciones, utilizando las placas de autobús o los códigos y horarios de las estaciones. Así, al complementar los datos GPS con los de la base de datos del bip!, es posible conocer el número de pasajeros que abordan en cada una de las paradas de cada ruta de autobús según intervalos de tiempo. Además, como sólo se almacena la información de la entrada en el sistema (se valida al subir al autobús, pero no al bajar), en este estudio se utilizó un procedimiento para estimar la parada de autobús o la estación de metro en cada etapa del viaje

desarrollado previamente también en el Transantiago por Munizaga y Palma (2012) (Figura 57). Esta metodología permite obtener matrices de etapas de viaje según paradas de autobús, con las cuales se pueden obtener matrices de viaje origen-destino. En otro trabajo previo realizado por Munizaga et al. (2014) los resultados fueron validados utilizando encuestas y entrevistas personales.

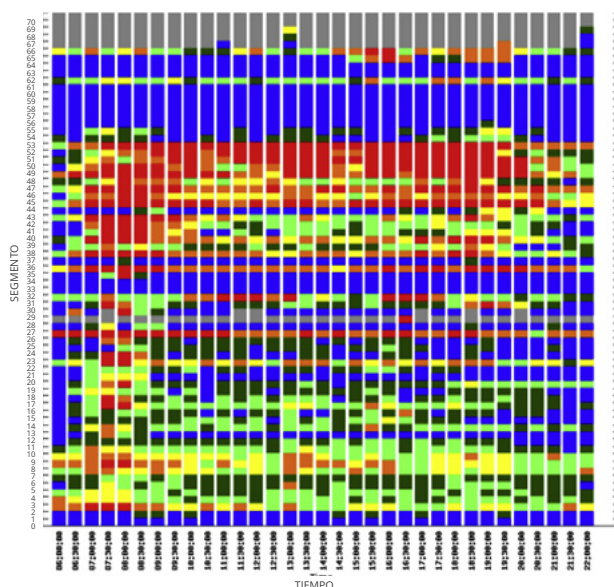


**Figura 57:** Algoritmo propuesto para el cálculo de etapas de viaje a partir de los datos de tarjetas de transporte. Fuente: Gschwender et al. (2016)

A partir de la información del GPS del autobús se desarrollan métodos para visualizar la localización espacial y temporal de cada autobús de una misma ruta (Figura 58) y perfiles de velocidad del autobús desglosados por segmentos espaciales e intervalos de tiempo (Figura 59). Este tipo de visualizaciones permiten detectar problemas como el agrupamiento de autobuses (cuando las líneas en la Figura 58 comienzan a agruparse) o posibles problemas de infraestructura cuando las velocidades observadas del autobús son demasiado bajas (colores rojos en la Figura 59) o su variabilidad es demasiado alta.



**Figura 58:** Diagrama espacio-temporal de la localización de cada autobús en una ruta específica. Fuente: Gschwender et al. (2016)



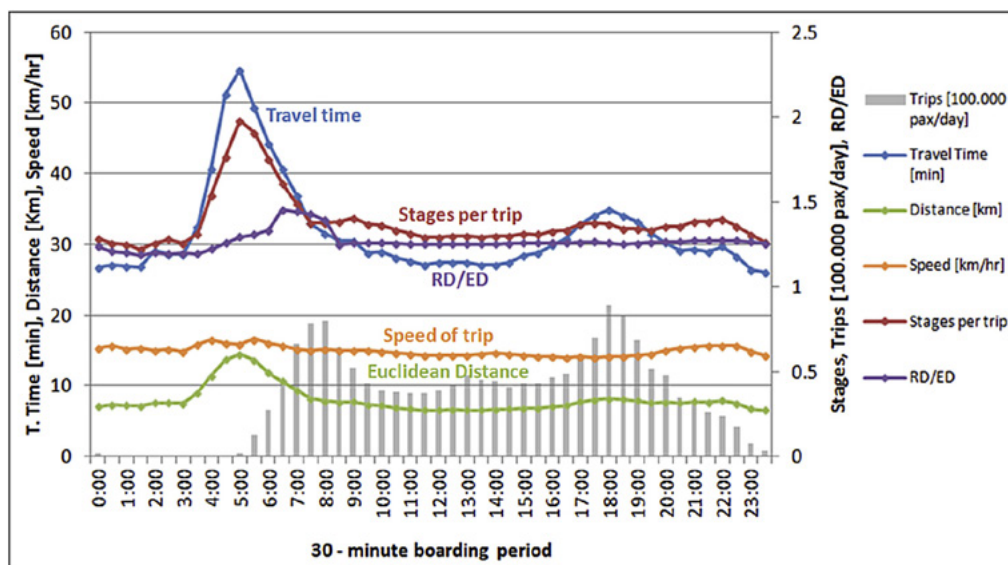
**Figura 59:** Perfil de velocidades de los autobuses de una ruta. En el eje x se representa el momento temporal y en el eje y cada segmento de la ruta (tramos entre dos paradas). La velocidad se muestra en una escala de colores que van desde el rojo (muy por debajo de 15 km/h) hasta el azul (excelente a 30 km/h). Fuente: Gschwender et al. (2016)

Una de las limitaciones de los datos de tarjetas inteligentes es la falta de información socioeconómica. En Santiago, la mayoría de las tarjetas no están personalizadas, y cuando se dispone de alguna información (por ejemplo, tarjetas vinculadas a estudiantes) la información no está disponible por razones de confidencialidad. Para superar esta dificultad, se desarrolló también un método que permite estimar la zona de residencia de los usuarios frecuentes, para asociar después variables socioeconómicas al individuo según las características de su zona de residencia (ver más información en Amaya y Munizaga, 2015). El método usado se basa en la observación de validaciones durante las primeras horas de la mañana y la imputación de la residencia a la zona donde se ha usado la tarjeta bip! en esa franja de mañana durante, al menos, cuatro veces en una misma semana.

## BENEFICIOS

Toda la metodología anterior fue implementada en un software que ha sido protegido y registrado como propiedad intelectual de la universidad y con licencia para DTPM. Este software permite obtener una base de datos muy detallada de los viajes que se realizan en el sistema de transporte público, con información para cada viaje de la parada de entrada y salida del sistema de transporte, la secuencia de rutas tomadas, el tiempo de viaje o los tiempos de transbordo. Con esta información se generan indicadores globales como el tiempo de viaje, número de etapas por viaje, la relación entre la distancia real (RD) y la distancia euclidiana (ED) de los viajes, la velocidad del viaje, etc. Estos indicadores pueden ser analizados temporalmente (Figura 60) o por zona de origen, zona de destino o por operadores.





**Figura 60:** Ejemplo de indicadores globales del servicio. Fuente: Gschwender et al. (2016)

La información obtenida es utilizada regularmente por las autoridades del transporte para gestionar el funcionamiento del sistema. Así, por ejemplo, se utilizan regularmente los datos de los perfiles de localización y velocidad de los autobuses por cada una de las rutas, para identificar ubicaciones donde se requiere una intervención en la infraestructura (por ejemplo, para priorizar el autobús sobre el coche) o la operación del sistema. Los datos de viajes son usados para realizar modificaciones en la red del servicio, para diseñar nuevos elementos del sistema, como estaciones de autobuses, etc. Además del DTPM, toda la información es utilizada por la agencia de planificación de transporte SECTRA y Metro o por los operadores privados de autobuses.



## 6.5 Análisis de los accidentes de tráfico a partir de la integración de datos oficiales y de Waze<sup>13</sup>

### SOBRE EL CASO DE ESTUDIO

Los accidentes de tráfico constituyen un grave problema en todo el mundo. El número de muertos en accidentes de tráfico continúa siendo inaceptablemente alto, a pesar de los constantes esfuerzos de las autoridades por reducirlo. Según la Organización Mundial de la Salud, en el año 2013 se produjeron 1,25 millones de víctimas mortales en todo el mundo por esta causa, afectando

<sup>13</sup> dos Santos, Junior, C. A. D. & Smarzaro, R. (2017). Analyzing Traffic Accidents based on the Integration of Official and Crowdsourced Data. Journal of Information and Data Management, 8(1), 67-82.

<https://periodicos.ufmg.br/index.php/jidm/article/view/379>

particularmente a los países con menores niveles de ingresos, y siendo la principal causa de muerte en el grupo de edad de 15 a 29 años. Además, los accidentes de tráfico afectan gravemente a las condiciones de circulación, provocando retenciones de tráfico que afectan a multitud de conductores, particularmente en horas punta, con un elevado perjuicio para la ciudad, ya que tales retrasos tienen un importante coste económico.

Los datos oficiales de accidentes de tráfico provienen, tradicionalmente, de los informes policiales, pero hoy en día los conductores disponen de teléfonos móviles a través de los cuales pueden informar sobre los accidentes y retenciones de tráfico en tiempo real. Una plataforma muy utilizada para este fin, particularmente en Latinoamérica, es Waze, un sistema de navegación que permite a los conductores calcular la ruta óptima en cada momento del día y que, además, aporta información sobre accidentes de tráfico; recomendando rutas alternativas en caso de que afecten a los niveles de congestión de la zona. Esta aplicación es descargada por los usuarios en sus teléfonos móviles y de esta forma Waze conoce las condiciones del tráfico en tiempo real. Los usuarios de Waze tienen la posibilidad de reportar incidencias, como accidentes o cortes de tráfico. De esta forma comparten información sobre las condiciones del tráfico y pueden evitar los tramos más congestionados. La incidencia reportada queda geolocalizada a través del GPS del teléfono móvil del usuario y se convierte, así, en una fuente de datos de gran utilidad para el estudio de la accidentalidad vial, con la ventaja de que se dispone de los datos en tiempo casi real, mucho antes que en el caso de las estadísticas oficiales.

En este caso de estudio se analizan los datos de accidentes de tráfico en la ciudad brasileña de Belo Horizonte, utilizando tanto datos oficiales como de Waze, con el objetivo de comparar ambas fuentes de datos y explorar posibles complementariedades.

## EL DESAFÍO

Los datos de las estadísticas oficiales constituyen la fuente más fiable para analizar los accidentes de tráfico, pero presentan algunos puntos débiles. En primer lugar, en muchos países las competencias en materia de accidentes están asignadas a distintas administraciones dependiendo del tipo de vía, sin que haya un formato estándar para su registro; esto dificulta su integración en una base de datos única. Por otro lado, la policía no reporta todos los accidentes. Solo registra los accidentes en los que se producen daños personales o materiales que impliquen reclamaciones de las compañías de seguros, pero los accidentes muy leves o que involucren vehículos no asegurados no suelen ser reportados.

Los avisos de accidentes de tráfico de Waze son muy útiles para los conductores porque se producen y comparten casi en tiempo real. Sin embargo, y a diferencia de lo que ocurre con los datos de los informes policiales, un mismo accidente puede estar repetido en varios registros, ya que puede ser informado por varios usuarios. Esto es una ventaja desde el punto de vista de asegurar la fiabilidad de la información del accidente, pero plantea el problema de reducir los diferentes avisos de un mismo accidente a un solo caso.

La cuestión que se plantea es si los datos de accidentes de Waze pueden ser un sustituto de los datos de los informes policiales o si por el contrario aportan información complementaria a los mismos. Y el desafío es doble. Consiste, por un lado, en ser capaces de descargar los datos de Waze y reducir los registros

que informan del mismo accidente a un solo registro por accidente, y, por otro, en conseguir integrar estas dos fuentes de datos, de forma que se pueda discriminar el nivel de coincidencia de ambas. Para este fin se trabajó con los datos de accidentes de ambas fuentes en la misma área de estudio (la ciudad de Belo Horizonte) y en el mismo periodo temporal (del 16 de septiembre al 11 noviembre de 2014).

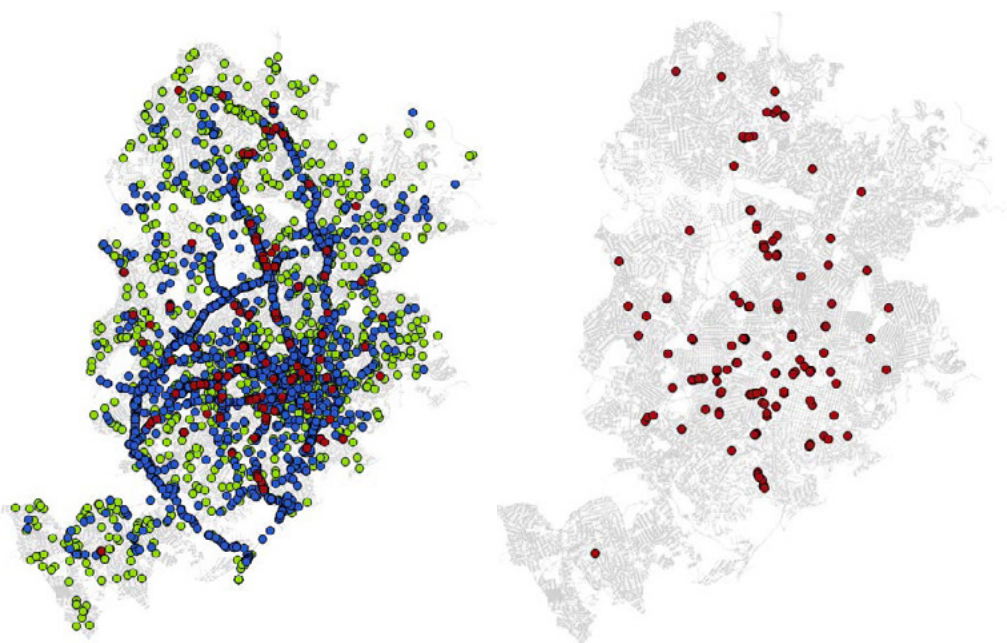
## LA SOLUCIÓN

Los datos oficiales fueron suministrados por las autoridades municipales, cubriendo el periodo anteriormente descrito. Para cada accidente se disponía de la geolocalización, fecha y hora del accidente, tipo de accidente y vehículos involucrados.

En cuanto a los datos de Waze, dado que no se dispone de una API para su descarga, los datos se recopilaron a través de un archivo GeoRSS generado por el Mapa Vivo en el sitio web de la aplicación. Se descargaron en un archivo JSON a intervalos regulares, que contiene características geográficas en tiempo real y ubicaciones de objetos y, en este caso, datos sobre congestión de tráfico y alertas. Una vez descargados, los datos fueron filtrados para seleccionar solamente los del periodo de estudio. Cada registro incluía datos sobre las coordenadas geográficas, fecha y hora del accidente y gravedad del mismo (a juicio del usuario). Los avisos de un mismo accidente reportados por varios usuarios, fueron reducidos a un solo accidente calculando la media de las coordenadas geográficas de esos avisos (utilizando el procedimiento que se describe a continuación para integrar accidentes de las dos bases de datos) y registrando para cada accidente el número de avisos.

Después de los procesos anteriores se obtuvieron dos bases de datos, la de los informes policiales y la de Waze, que contenían 1.434 y 1.543 accidentes, respectivamente. Para la integración de ambas bases de datos, verificando la existencia de eventos superpuestos, se estableció que dos registros de ambas fuentes de datos se refieren al mismo accidente si se informaron dentro de un intervalo de tiempo de una hora y se encontraban a una distancia de menos de 50 metros el uno del otro o dentro de los 150 metros en la misma vía. Al final del proceso, se creó un único conjunto de datos integrado sobre accidentes, que contiene la fuente de cada registro.

El intervalo de una hora, parece razonable para establecer que se trata del mismo accidente en las dos fuentes de datos, ya que podría haber una demora entre el momento en que ocurrió el accidente (informado por las partes involucradas en un informe policial) y el momento en que Waze informó el accidente. El criterio de distancia se adoptó porque podría haber situaciones en las que el usuario de Waze pasa por la ubicación del accidente y lo informa, no exactamente en la localización donde ocurrió. En este caso, si los registros están en la misma vía, el límite de distancia se establece en 150 m, ya que es muy probable que se trate del mismo accidente. Cuando las vías son diferentes, la distancia es más restrictiva (50 metros) para este caso con el fin de evitar falsas coincidencias.

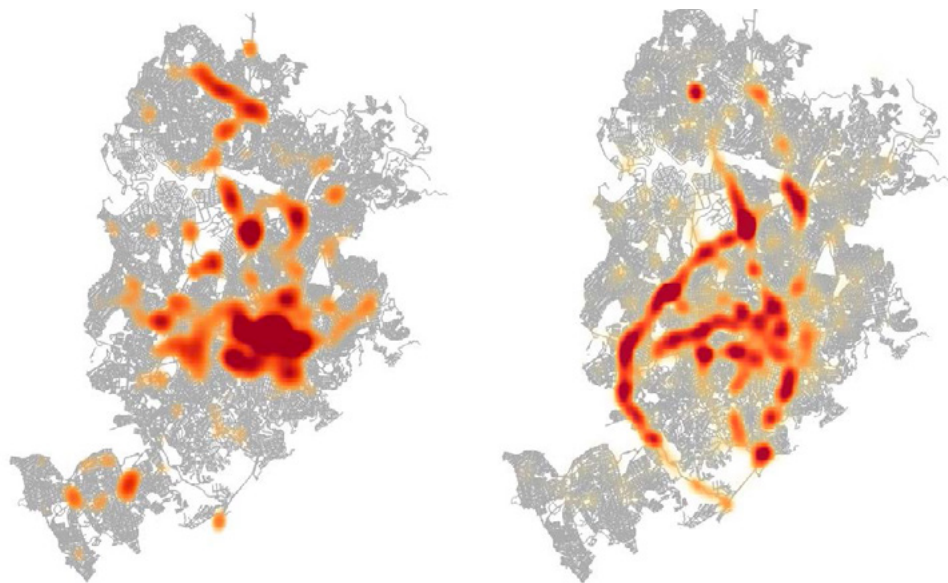


**Figura 61:** Localización de los accidentes según datos oficiales (verde), Waze (azul) y coincidentes (rojo). Fuente: dos Santos et al., 2017

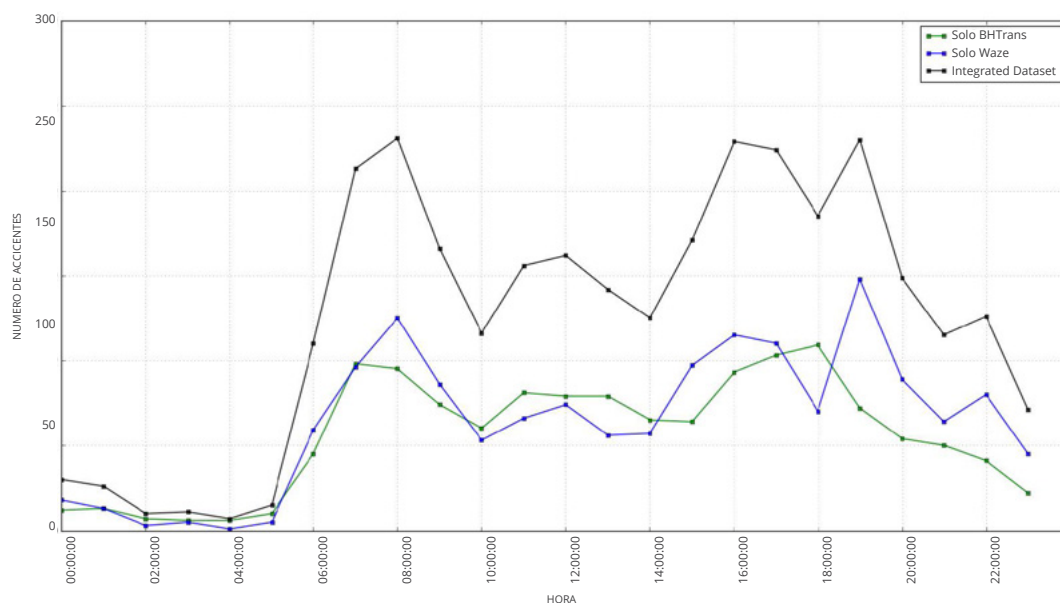
Como resultado del proceso anterior, se obtuvieron 1.333 accidentes informados sólo por los registros oficiales, 1.442 informados sólo por Waze y 101 informados por ambas fuentes (Figura 61). La tasa de coincidencia es, por tanto, muy baja, con lo que el número de accidentes de la fuente oficial se duplicó al incorporar los de Waze no coincidentes. Los registros policiales recogen accidentes en los que se pide una compensación por daños en los vehículos, por lo que es posible que no recojan accidentes leves. En el caso de Waze, las alertas obedecen sobre todo a situaciones en las que los accidentes afectan significativamente a la congestión y los usuarios de esta app comparten la información para evitar retraso.

Los patrones espaciales de ambas fuentes de datos son muy diferentes, como muestran los mapas de calor (Figura 62). Los accidentes de la fuente oficial se localizan sobre todo en el centro de la ciudad, donde hay una gran intensidad de tráfico durante todo el día, y en algunas áreas de actividad de la periferia. Los de Waze se sitúan sobre todo a lo largo de las principales carreteras, en las que las condiciones de circulación se ven muy afectas por este tipo de incidencias, particularmente en horas punta (Figura 63).





**Figura 62:** Mapas de calor de los accidentes no coincidentes: a la izquierda, datos oficiales; a la derecha, Waze. Fuente: dos Santos et al., 2017



**Figura 63:** Distribución de los accidentes según horas del día y fuentes de datos. Fuente: dos Santos et al., 2017

## BENEFICIOS

Waze aporta información en tiempo casi real sobre accidentes que afectan a las condiciones de circulación. Además, se puede acceder a estos datos a través de la web de la aplicación. Esta información puede ser útil para cubrir algunas lagunas de las fuentes basadas en registros policiales. Las fuentes oficiales son fiables y detalladas, pero normalmente es difícil acceder a ellas y, además, subestiman el número total de accidentes, ya que muchos de carácter

leve no son recogidos en los informes policiales. Waze añade accidentes que no han sido registrados por las fuentes oficiales posiblemente por tratarse de accidentes leves.

La integración de ambas fuentes de datos permite, por tanto, enriquecer las fuentes oficiales, incorporando accidentes en general de carácter leve, pero que afectan gravemente a las condiciones del tráfico. Esta base de datos conjunta puede resultar de gran utilidad para las autoridades al momento de tomar acciones conducentes a reducir el número de accidentes de tráfico (y la congestión derivada) en los puntos y tramos de mayor concentración de los mismos, como mejorar la señalización, aumentar los controles de velocidad o llevar a cabo campañas de concienciación.





# 7

## Consideraciones Finales





Los registros GPS constituyen una excelente oportunidad para planificadores y gestores de la movilidad urbana. Ofrecen información muy variada sobre la movilidad en las ciudades. Los datos se generan de forma constante y ello permite monitorizar y analizar procesos. El disponer de un flujo continuo de datos hace posible estudiar la movilidad en días atípicos o el impacto de eventos especiales en la movilidad de las ciudades. Los datos de dispositivos GPS permiten analizar temáticas muy diferentes, como las velocidades de circulación, la distribución de los viajes, la carga de la red, los problemas de agrupamientos de autobuses, los accidentes de tráfico, etc. A diferencia de lo que ocurre con los datos de telefonía móvil, en los que existen serias dificultades para identificar el modo de transporte, en el caso de los registros GPS la identificación del modo está asegurada cuando se trata de datos provenientes de dispositivos instalados en los vehículos: autobuses, vehículo privado, bicicletas, patinetes eléctricos, etc.

Existen diferentes fuentes de datos de utilidad para realizar estudios de movilidad basadas en registros GPS. Algunas de estas fuentes (que se describen en el apartado 2.3) son las siguientes:



	TIPO	UTILIDAD	MODOS	ACCESO
<i>Redes sociales</i>	Localizaciones	Matrices OD, rutas	No diferenciado	API
<i>Tracks vehículos</i>	<i>Tracks</i>	Matrices OD, rutas, tiempos de viaje, velocidades, carga red	Según tipo de vehículo	Administraciones Acuerdos con empresas
<i>Smart cards</i>	Localizaciones	Matrices OD, carga red	Transporte público	Administraciones
<i>TomTom Speed Profiles</i>	Agregados	Tiempos de viaje	Vehículo privado	Compra ficheros
<i>TomTom API</i>	Agregados	Matrices OD, estadísticas de tráfico	Vehículo privado	API
<i>Waze</i>	Localizaciones Agregados	Accidentes, incidentes Tiempos de viaje	Vehículo privado Vehículo privado	API API
<i>Google Maps</i>	Agregados	Tiempos de viaje	Vehículo privado Transporte público Bicicleta Peatonal	API
<i>Strava</i>	Agregados	Carga de la red	Vehículo privado	API

Los registros GPS comparten buena parte de las características de los datos de telefonía móvil, si bien es más difícil hacer generalizaciones por la gran variedad de fuentes de datos existentes:

- **GRAN ALCANCE ESPACIAL:** Muchas de estas fuentes de datos tienen una cobertura prácticamente global, como ocurre, por ejemplo, con Google Maps, Waze, TomTom o Strava. Por lo tanto, es posible analizar la movilidad de los usuarios, tanto dentro, como fuera del área de estudio definida y realizar estudios comparativos entre ciudades. Algunas fuentes de datos de este tipo, en cambio, tienen carácter puramente local, como los datos de transporte público urbano obtenidos a partir de tarjetas de transporte y geolocalizados por el GPS de los autobuses.
- **RECOGIDA PASIVA DE LA INFORMACIÓN:** Al igual que en el caso de la telefonía móvil, y en contraposición con las encuestas, los datos se recogen de forma pasiva, sin que el usuario tenga que responder un cuestionario. La contrapartida es que los registros GPS no aportan información sobre algunas características de los viajes (como el motivo), ni sobre motivaciones y valoraciones de los usuarios, que sí puede obtenerse a través de encuestas. La valoración sobre los servicios de transporte que puede obtenerse a través de redes sociales (como Twitter) es muy sesgada y poco fiable.

- **GRANDES VOLÚMENES DE DATOS:** Como los datos se generan de forma continua, se dispone de grandes volúmenes de datos, que en algunos casos pueden referirse no a muestras, sino a todos o prácticamente todos los viajes realizados en un modo de transporte (por ejemplo, los datos provenientes de las tarjetas inteligentes de transporte público en algunas ciudades son exhaustivos).
- **RESOLUCIÓN TEMPORAL Y ESPACIAL MUY ALTA:** Tanto los *tracks* como los datos puntuales de localizaciones tienen una localización muy precisa, con un error de tan solo unos pocos metros. Los *tracks* tienen, además, una resolución temporal muy alta, generalmente de tan solo unos segundos o, en el peor de los casos, pocos minutos. En el caso de datos de localizaciones puntuales (como los de redes sociales) la resolución temporal es relativamente baja, lo que supone una importante limitación a la hora de estudiar los viajes.

Si se comparan los datos de telefonía móvil con los de registros GPS, se puede afirmar que los primeros son más adecuados para estudiar las características de la movilidad general de las ciudades. Los resultados de los estudios de movilidad basados en datos de telefonía móvil permiten recoger de forma satisfactoria el conjunto de los viajes que realiza la población de una ciudad, mientras que los basados en *tracks* normalmente se refieren a modos específicos. Los datos de localización espacio-temporal obtenidos a partir de redes sociales presentan la debilidad de su menor resolución temporal, si bien pueden aportar una primera aproximación a los patrones de movilidad urbana (viajes recurrentes) e interurbana (viajes por turismo de ocio o de negocio). Esta situación puede cambiar si alguna empresa, como Google, decide dar el salto hacia la producción de matrices OD según modos de transporte, a partir de los datos de los *tracks* de los usuarios de Google Maps.

Las ventajas de los *tracks* quedan patentes al analizar los patrones de movilidad de modos específicos, no sólo porque el modo queda claramente identificado, sino también porque los patrones de movilidad se pueden analizar con mucha más precisión: horas de inicio y fin del viaje, lugar de origen y destino, velocidad y tiempo de viaje, ruta, carga de la red, etc. Esta información es de gran utilidad para los gestores de movilidad a la hora de tratar problemas específicos: congestión viaria, agrupamientos de autobuses, redistribución de bicicletas públicas, etc. Además, existen fuentes que aportan informaciones específicas de gran interés, como es el caso de Waze para los accidentes de tráfico.

Por otro lado, también hay que considerar las limitaciones técnicas de los registros GPS frente a las de los datos de telefonía móvil. La exactitud posicional del GPS es menor en los entornos urbanos densos, con edificios altos, y la señal se pierde en espacios subterráneos (túneles) y en el interior de edificios; a pesar de que se han desarrollado soluciones basadas en otras tecnologías que ofrecen resultados parecidos a los del GPS en interiores. Frente a esto, es raro que se pierda la cobertura de la red de telefonía móvil en las ciudades y ésta puede funcionar satisfactoriamente en túneles y espacios interiores mediante antenas *indoor*.

En cuanto a la accesibilidad a los datos, en el caso de los registros de telefonía móvil se depende siempre de las operadoras, mientras que en el de los registros GPS la situación es muy variable. El sector público produce una enorme cantidad de *tracks* (registros GPS de autobuses, tarjetas de transporte



y de nuevos servicios de movilidad como las bicicletas públicas), pero en otros casos el acceso a los datos depende de las políticas de las empresas que los almacenan, pudiendo ser gratuito para “pequeños volúmenes” de datos o de pago para volúmenes mayores.

En cualquier caso, la cuestión no es pronunciarse sobre qué fuente de datos es mejor para los estudios de movilidad, sino ser capaces de utilizar las más adecuadas en función del objetivo del estudio y, sobre todo, ser capaces de cruzar distintas fuentes de datos para obtener mejores resultados. Afortunadamente, la geolocalización de los datos permite cruzar nuevas fuentes de datos entre sí y con datos de encuestas o de estadísticas oficiales (enriquecimiento de datos). Así, por ejemplo, en un estudio de movilidad general con base en datos de telefonía móvil, los datos de tarjetas inteligentes de transporte pueden ayudar a identificar el modo y los datos de estratos sociales de las estadísticas oficiales permitirán caracterizar a los viajeros.

Los registros GPS aparecen, por tanto, como una interesante oportunidad para planificadores y gestores de la movilidad en las ciudades. Algunas ciudades de América Latina han comenzado a utilizarlos con resultados muy positivos, pero todavía queda mucho camino por recorrer. El potencial de estos datos es enorme y lo será cada vez más en el futuro con la extensión de los nuevos servicios de movilidad. La configuración de equipos multidisciplinarios permitirá sacar cada vez más provecho de estas “nuevas” fuentes de datos.





## Bibliografía

- Alvarado-Sosa, Edgar, Marín-Fernández, Alberto, Mirabal-Martínez, Marbelis, González, Onéximo, y Barinas-Uribe, Marcos (2019). Actualización del diagnóstico y propuesta de gestión de la movilidad urbana de Santiago de los Caballeros. Banco Interamericano de Desarrollo. <http://dx.doi.org/10.18235/0001757>
- Amaya, M., Cruzat, R., y Munizaga, M. A. (2018). Estimating the residence zone of frequent public transport users to make travel pattern and time use analysis. *Journal of Transport Geography*, 66, 330-339.
- Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy*, 15(2), 73-80.
- Barbero, José, y Uechi Luis (2012). Evaluación de la disponibilidad y la calidad de los datos sobre el transporte en América Latina. Banco Interamericano de Desarrollo. <http://publications.iadb.org/handle/11319/5620?locale-attribute=en>
- Barros, C., Moya-Gómez, B., y Gutiérrez, J. (2019). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. *Current Issues in Tourism*, 0(0), 1–20. <https://doi.org/10.1080/13683500.2019.1619674>
- Batty, M. (2013). Big Data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274-279.
- Benevolo, C., Dameri, R. P., y D'Auria, B. (2016). Smart mobility in smart city. In *Empowering Organizations* (pp. 13-28). Springer, Cham.
- Bussche, D., y Van de Coevering, P. (2015). BikePRINT–In depth analysis of cyclist behaviour and cycle network performance using GPS-tracking technology. In *European Transport Conference 2015 Association for European Transport (AET)*.
- Byon, Y. J., Cortés, C. E., Martínez, C., Javier, F., Munizaga, M., y Zuniga, M. (2011). Transit performance monitoring and analysis with massive gps bus probes of transantiago in santiago, chile: Emphasis on development of indices for bunching and schedule adherence (No. 11-2233).
- Cortés, C. E., Gibson, J., Gschwender, A., Munizaga, M., y Zúñiga, M. (2011). Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies*, 19(4), 695-707.
- Demetriou, S., Jain, P., y Kim, K.-H. (2018). Codrive: Improving automobile positioning via collaborative driving. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications* (pp. 72–80). IEEE.
- Dewulf, B., Neutens, T., Vanlommel, M., Logghe, S., De Maeyer, P., Witlox, F., ... Van de Weghe, N. (2015). Examining commuting patterns using Floating Card Data and circular statistics: Exploring the use of new methods and visualizations to study travel times. *Journal of Transport Geography* 48, 41-51.
- Dos Santos, S. R., Junior, C. A. D., y Smarzaro, R. (2017). Analyzing Traffic Accidents based on the Integration of Official and Crowdsourced Data. *Journal of Information and Data Management*, 8(1), 67-82.
- García-Albertos, P., Picornell, M., Salas-Olmedo, M. H., y Gutiérrez, J. (2019). Exploring the potential of mobile phone records and online route planners for dynamic accessibility analysis. *Transportation Research Part A: Policy and Practice*, 125, 294-307.
- García-Palomares, J.C., Gutiérrez, J. y Mínguez, C. (2015). Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63, pp. 408-417.



- García-Palomares, J.C., Salas-Olmedo, M.H., Moya-Gómez, B., Condeço-Melhorado, A., and Gutiérrez, J. (2018). City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities*, 72(B), 310-319.
- Girardin, F., Blat, J., Calabrese, F., Dal Fore, F. Y Ratti, C. (2008) Digital Footprinting: uncovering tourists with user-generated content. *IEEE Pervasive computing*, 7(4).
- Gschwender, A., Munizaga, M., y Simonetti, C. (2016). Using smart card and GPS data for policy and planning: The case of Transantiago. *Research in Transportation Economics*, 59, 242-249.
- Gutiérrez-Puebla, Javier, Benítez, Carolina, Leaño, Juan, García-Palomares, Juan, Condeço-Melhorado, Ana, Mojica, Carlos, Scholl, Lynn, Adler, Verónica, y Vera, Felipe (2019). Cómo aplicar Big Data en la planificación del transporte urbano: El uso de datos de telefonía móvil en el análisis de la movilidad. Banco Interamericano de Desarrollo. <http://dx.doi.org/10.18235/0002009>
- Gutiérrez, J., García-Palomares, J.C., Romanillos, G, and Salas-Olmedo, M.H. (2017). The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tourism Management*, 62, 278-291
- Gutiérrez Puebla, J. (2018). Big Data y nuevas geografías: la huella digital de las actividades humanas. *Documents d'anàlisi geogràfica*, 64(2), 0195-217.
- Gutiérrez Puebla, J., García Palomares, J. C., y Salas Olmedo, M. H. (2016). Big (Geo) Data en Ciencias Sociales: Retos y Oportunidades. *Revista de estudios andaluces*, 33(1), 1-23.
- Hadas, Y. (2013): Assessing public transport systems connectivity based on Google Transit data. *Journal of Transport Geography*, 33, 105–116.
- Hawkins, A. (2018). How Uber moves the 'blue dot' to improve GPS accuracy in big cities. *The Verge*. Retrieved from <https://www.theverge.com/2018/4/19/17252680/uber-gps-blind-spot-shadow-maps>
- Hong, J., McArthur, D. P., y Livingston, M. (2019). The evaluation of large cycling infrastructure investments in Glasgow using crowdsourced cycle data. *Transportation*. <https://doi.org/10.1007/s11116-019-09988-4>
- Jittrapirom, P., Caiati, V., Feneri, A. M., Ebrahimigharehbaghi, S., Alonso González, M. J., y Narayan, J. (2017). Mobility as a service: A critical review of definitions, assessments of schemes, and key challenges.
- Lenormand, M.; Picornell, M.; Cantú-Ros, O.; Tugores, A.; Louail, T.; Herranz, R.; Barthelemy, M.; Frías-Martínez, E. y Ramasco, JJ. (2014). Cross-checking different sources of mobility information. *PLoS ONE*, 9, e105184.
- Luong, T.B.T., Houston, D. (2015). Public opinions of light rail service in Los Angeles, an analysis using Twitter data. *Universidad de California. iConference 2015 Proceedings*.
- Mazloumi, E., Currie, G., y Rose, G. (2009). Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering*, 136(7), 623-631.
- Miralles-Guasch, C., (2012). Las encuestas de movilidad y los referentes ambientales de los transportes. *EURE (Santiago)*, 38(115), 33-45.
- Mohíno, I., Moya-Gómez, B. y García-Palomares, J.C. (2018). Identifying tourist main routes within historic city centres using geotagged photographs. *Congreso de Ingeniería del Transporte*. Gijón, España
- Moya-Gómez, B. and García-Palomares, J.C. (2017). The impacts of congestion on automobile accessibility. What happens in the largest European cities?. *Journal of Transport Geography*, 62, 148-159
- Moya-Gómez, B., Salas-Olmedo, M. H., García-Palomares, J. C., y Gutiérrez, J. (2017). Dynamic

Accessibility using Big Data: The Role of the Changing Conditions of Network Congestion and Destination Attractiveness. *Networks and Spatial Economics*, 2(18), 273–290.

Munizaga, M. A., y Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.

Munizaga, M. A., Devillaine, F., Navarrete, C., y Silva, D. (2014). Validating travel behaviour estimated from smartcard data. *Transportation Research Part C*, 44, 70e79.

Netto, V. M., Pinheiro, M., Meirelles, J. V. y Leite, H. (2015): Digital footprints in the cityscape. *International Conference on Social Networks*, Athens, USA.

Okuyama, K., y Yanai, K. (2013). A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web. In *The era of interactive media* (pp. 657-670). Springer, New York, NY.

Osorio Arjona, J., y García-Palomares, J. C. (2017). Nuevas fuentes y retos para el estudio de la movilidad urbana. *Cuadernos Geográficos*, 56(3), 247-267.

Osorio-Arjona, J., y García-Palomares, J.C. (2019). Social media and urban mobility: Using twitter to calculate home-work travel matrices. *Cities*, 89 (June 2019), 268-280

Pelletier, M.-P., Trepanier, M., y Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C*, 19, 557-568.

Pereira, F.C., Rodrigues, F., Ben-Akiva, M. (2013). Using Data from the web to predict public transport arrivals under special events. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 19(3), 273-288.

Pérez, G. V. A., López, J. C., Cabello, A. L. R., Grajales, E. B., Espinosa, A. P., y Fabian, J. L. Q. (2018). Road Traffic Accidents Analysis in Mexico City through Crowdsourcing Data and Data Mining Techniques. *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering* Vol:12, No:8, 2018

Regus, M., Talar, R., y Labudzki, R. (2019). Indoor positioning and navigation system for autonomous vehicles based on RFID technology. In *IOP Conference Series: Materials Science and Engineering* (Vol. 659, p. 12059). IOP Publishing.

Rendón, José, Hernández, Enrique, y Del Río, Hernán (2019). Nueva generación de modelos de transporte a través del uso de big data: Caso San Salvador. Banco Interamericano de Desarrollo. <http://dx.doi.org/10.18235/0002130>

Romanillos, G., y Zaltz Austwick, M. (2015). Madrid cycle track: visualizing the cyclable city. *Journal of Maps*, 12(5), 1–9.

Romanillos, G.. (2018). The digital footprint of the cycling city GPS cycle routes visualization and analysis. Universidad Complutense de Madrid. Retrieved from <https://eprints.ucm.es/50818/>

Romanillos, G., Moya-Gómez, B., Zaltz-Austwick, M., & Lamíquiz-Daudén, P. J. (2018). The pulse of the cycling city: visualising Madrid bike share system GPS routes and cycling flow. *Journal of Maps*, 14(1), 34-43.

Romanillos, G., y Gutiérrez, J. (2019). Cyclists do better. Analyzing urban cycling operating speeds and accessibility. *International Journal of Sustainable Transportation*, 1–17. <https://doi.org/10.1080/15568318.2019.1575493>

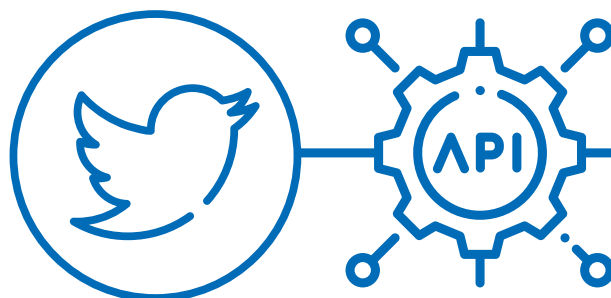
Rubinstein da Silva, Elias, Hanz, Martin, y Hernández, Diego (2018). ¿Qué implica la accesibilidad en el diseño e implementación de políticas públicas urbanas?: Concepto, instrumentos para su evaluación y su rol en la planificación de la movilidad urbana. Banco Interamericano de Desarrollo. <http://dx.doi.org/10.18235/0001469>

- Salas-Olmedo, M.H. y Rojas, C. (2017). The use of public spaces in a medium-sized city: from Twitter data to mobility patterns. *Journal of Maps*, 13(1), 40-45
- Schwab, K. (2017). *The fourth industrial revolution*. Crown Business.
- Shelton, T., Poorthuis, A., y Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142, 198-211.
- Stępnia, M., Pritchard, J. P., Geurs, K. T., y Goliszek, S. (2019). The impact of temporal resolution on public transport accessibility measurement: Review and case study in Poland. *Journal of transport geography*, 75, 8-24.
- Steur, Roeland (2014). Twitter as a spatio-temporal information source for traffic incident management. *GIMA (Geographical Information Management and Applications)*.
- Szell, M., y Groß, B. (2014). Hubcab-exploring the benefits of shared taxi services. *Decoding the City How Big Data Can Change Urbanism*.
- Tao, S., Rohde, D., y Corcoran, J. (2014). Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap. *Journal of Transport Geography*, 41, 21-36.
- Tascón, M., Coullaut, A. (2016). *Big Data y el Internet de las cosas*. Los Libros de la Catarata.
- Trépanier Martin, Morency Catherine, Agard Bruno (2009). Calculation of Transit Performance Measures Using Smartcard Data, *Journal of Public Transportation*, 12(1), pp. 79-96.
- Vallejos, S., Caimmi, B., Alonso, D., Soria, Á., y Berdun, L. S. (2017). Detectando incidentes de tránsito en redes sociales: un enfoque inteligente basado en Twitter vs. Waze. In XVIII Simposio Argentino de Inteligencia Artificial (ASAI)-JAIIO 46 (Córdoba, 2017).
- Wu, L., Zhi, Y., Sui, Z., y Liu, Y. (2014). Intra-Urban Human Mobility and Activity Transition: Evidence from Social Media Check-In Data. *PLoS ONE*, 9(5), 1-13.
- Yang, Z., Franz, M. L., Zhu, S., Mahmoudi, J., Nasri, A., & Zhang, L. (2018). Analysis of Washington, DC taxi demand using GPS and land-use data. *Journal of Transport Geography*, 66, 35-44.
- Yue, Y., Wang, H. D., Hu, B., Li, Q. Q., Li, Y. G., y Yeh, A. G. (2012). Exploratory calibration of a spatial interaction model using taxi GPS trajectories. *Computers, Environment and Urban Systems*, 36(2), 140-153.



## Anexo: Descarga de datos de la API de Twitter





Twitter es una red social con unos 330 millones de usuarios-perfiles activos<sup>14</sup>, en la que sus miembros pueden expresar sus opiniones en mensajes o tweets de hasta 280 caracteres, mencionando a otros usuarios (@nombre\_usuario), etiquetando sus mensajes con el uso de hashtags (#hashtag) e incluyendo imágenes e imágenes animadas (ficheros .gif) y su descripción. Además, los usuarios pueden difundir los tweets de otros usuarios (retwittear) y comentar los comentarios de otros usuarios (retwittear con comentario). No se consideran tweets los mensajes privados, también llamados mensajes directos, que se envían entre dos o más usuarios.

Twitter también permite vincular sus cuentas con cuentas de otras redes sociales, tales como Facebook, Instagram y Foursquare. De esta manera, mensajes que se generan en una red social pueden aparecer en la cuenta de Twitter asociada y viceversa.

Se estima que se emiten unos 6.000 tweets cada segundo en todo el planeta<sup>15</sup>, llegando momentáneamente a puntas de hasta 12.000 tweets por segundo, como ocurrió durante la celebración de la Copa del Mundo de Fútbol de 2014. Por otro lado, se estima que un 1% de los tweets están geolocalizados, ya sea conociendo la posición exacta de su emisión o la ciudad en la que se emite.

La red social Twitter permite la descarga gratuita, a través de la Standard API y previa aprobación de la solicitud, de los datos de hasta el 10% del total de tweets que se generan. También se dispone de la posibilidad de acceder a todos los tweets mediante la compra de accesos a Premium API o Enterprise API. Un dato de tweet es mucho más que sólo 280 caracteres, tal y como se puede observar en los metadatos de los objetos tweet<sup>16</sup>.

En este anexo se explican los pasos a seguir para solicitar las credenciales para el acceso a datos que se pueden obtener de la Standard API de Twitter y se ponen a disposición del público scripts en lenguaje de programación Python, para la descarga de tweets geolocalizados en el entorno QGIS de OSGeo4W mediante el uso de la Standard API de Twitter, fácilmente usable en otro entorno:

<sup>14</sup> Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019. <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>. Visitado el 15 de diciembre de 2019.

<sup>15</sup> The 2014 #YearOnTwitter. [https://blog.twitter.com/official/en\\_us/a/2014/the-2014-yearontwitter.html](https://blog.twitter.com/official/en_us/a/2014/the-2014-yearontwitter.html). Visitado el 15 de diciembre de 2019.

<sup>16</sup> Tweet object. <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object>. Vistado el 15 de diciembre de 2019.

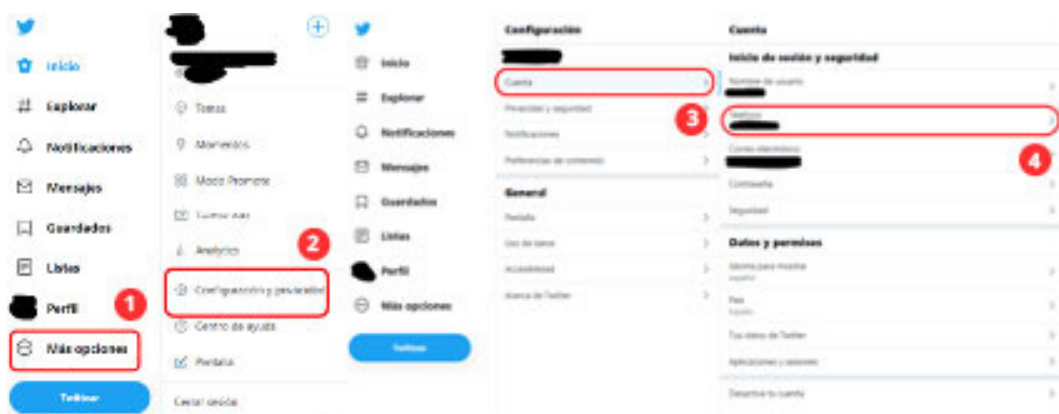
- El primer código (tw1 Descargar twitter HISTORICO – REST) permite descargar tweets de hasta una semana de antigüedad, pudiendo limitar las búsquedas por lengua o texto, limitando el número de tweets a encontrar y el número de solicitudes a realizar a las APIs de Twitter.
- El segundo código (tw2 Descargar twitter STREAMING) permite obtener los tweets geolocalizados en tiempo real, pudiendo parar el proceso cada minuto.
- Ambos códigos generan ficheros de texto plano, con formato geojson de nueva delimitación, que pueden ser abiertos en QGIS.

## A. Obtención de credenciales para la Standard API de Twitter

El proceso de obtención de credenciales para la Standard API puede durar varios días, según cómo se complete el formulario de solicitud.

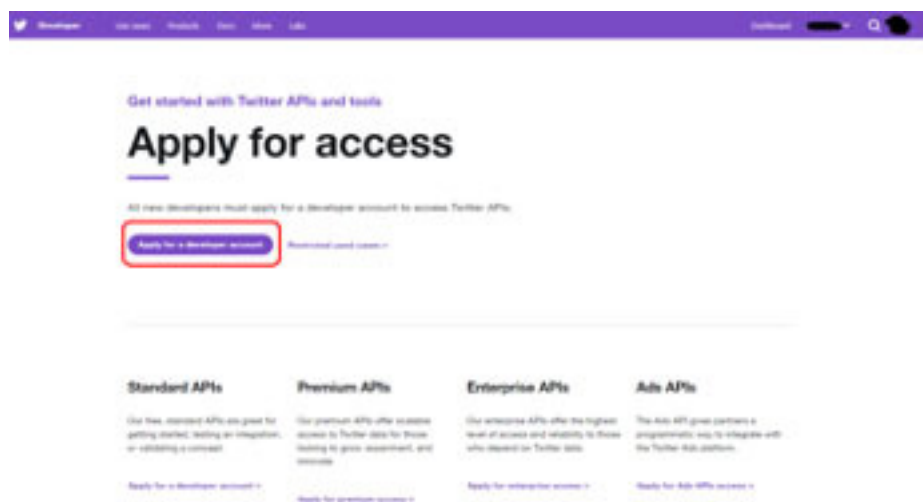
Para poder obtener las credenciales para la Standard API es necesario disponer de una cuenta de Twitter. Pueden generar su cuenta en el sitio web <https://twitter.com/>.

Además de tener una cuenta, ésta debe tener un teléfono móvil asociado. Compruebe que es así y, en caso contrario, vincule un número de teléfono tal y como se muestra en la Figura X1.



**Figura X1:** Pasos para vincular un número de teléfono a una cuenta de Twitter.

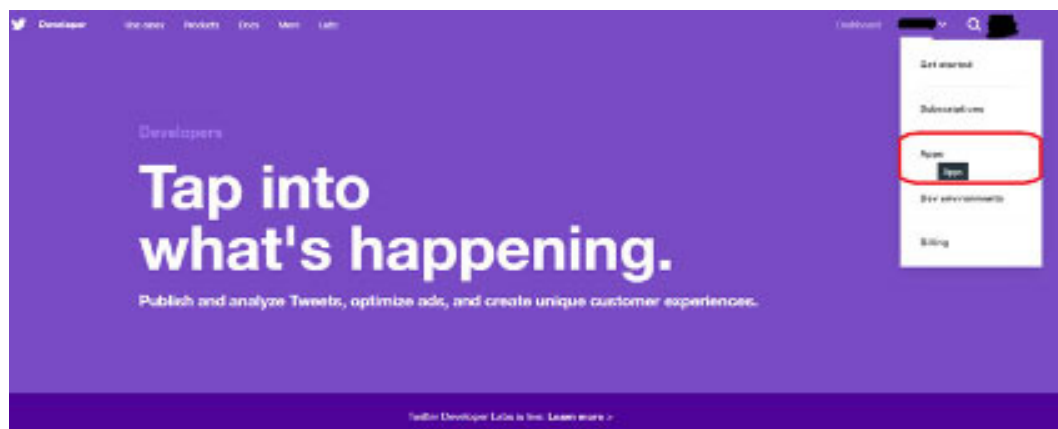
Una vez se tiene una cuenta de Twitter, se debe pedir ser un usuario desarrollador (developer) para acceder a la Standard API. Para ello, se debe visitar el sitio web <https://developer.twitter.com/en/apply-for-access> y clicar en el botón mostrado en la Figura X.2.



**Figura X.2:** Solicitud acceso a Standard API

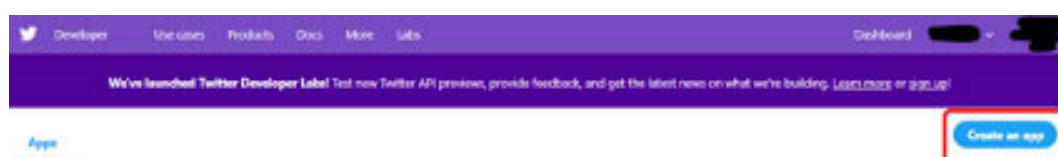
Una vez cliqueado el botón de la Figura X.2, deberá rellenar el formulario que le aparecerá, preferentemente en inglés, y deberá esperar a que le concedan acceso a la API.

Una vez tenga acceso a la API, deberá crear una nueva app, iniciando el proceso tal y como se muestra en la Figura X.3. Cada usuario puede tener varias apps creadas, con las correspondientes credenciales a Standard API de Twitter.



**Figura X.3:** Acceso a entorno de tus APPs de Twitter.

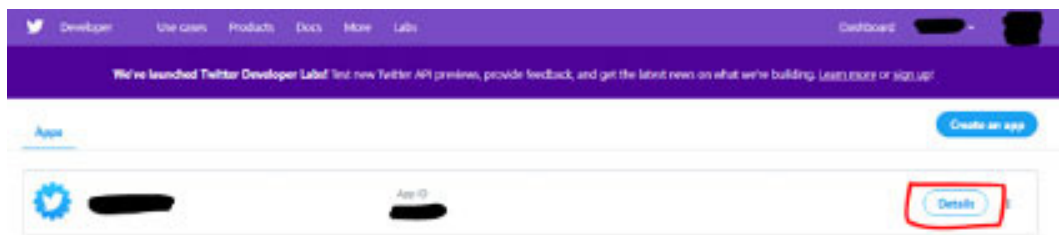
Seguidamente, se solicita crear una nueva app (Figura X.4.)



**Figura X.4:** Creación de una nueva app de Twitter

A continuación, debe rellenar el formulario de definición de su app, debiendo decidir obligatoriamente el nombre de su app, añadir una descripción de lo que hará, una url y para qué usar la app (última pregunta).

Una vez finalizado, volverá a la pantalla de sus apps, en la que aparecerá la nueva app creada (Figura X.5). Deberá clicar en “details” para acceder a sus credenciales.



**Figura X.5:** Pantalla de apps con una app creada

En la ficha de detalles de su app, deberá elegir la pestaña “Keys and Tokens” (Figura X.6.)



**Figura X.6:** Pantalla de la ficha de detalles de su app

Finalmente, tendrá acceso a las credenciales de su app a la Standard API (Consumer API keys, Access token, y Access token secret), tal y como se muestra en la Figura X.7. Estas claves son necesarias para el funcionamiento de los códigos de este anexo.



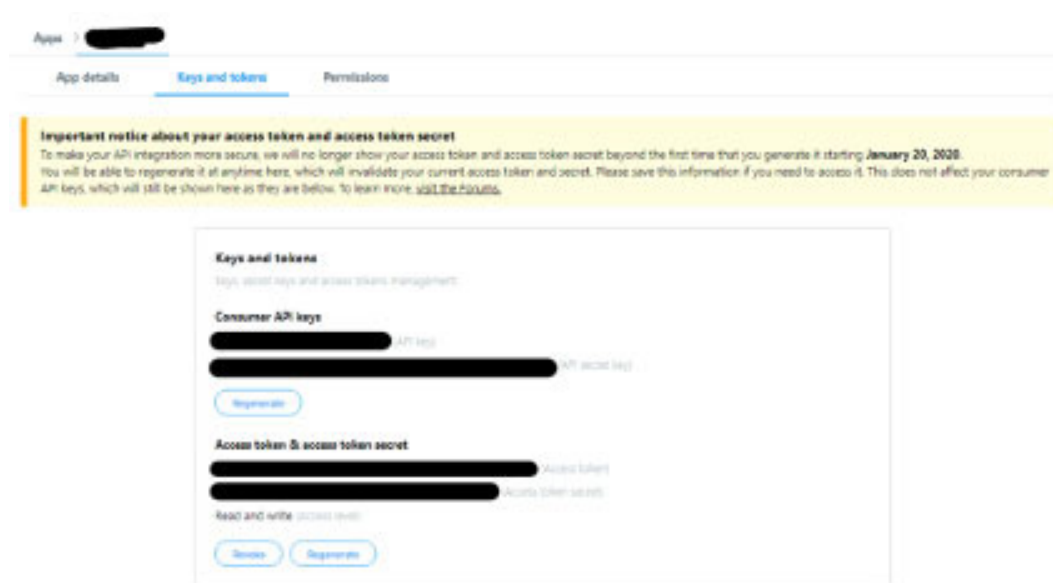


Figura X.7: Pantalla de la ficha de detalles de su app

En caso de no estar generadas, se pueden generar en el correspondiente botón. También se pueden regenerar, si así se considera conveniente.

## B. Códigos de descarga de tweets

Los códigos aquí mostrados están preparados para ser añadidos directamente a la *Caja de herramientas de Procesos* de QGIS 3.10. Si se desea usar el código en otro entorno, deberá copiar y pegar el código en negrita de la función `processAlgorithm` de la clase definida en el código. Deberá asegurarse de que todos los parámetros que se solicitan están correctamente definidos.

El método `feedback.pushInfo` es el equivalente a `print` en las consolas de Python. En caso de usar consola de Python, deberá sustituir `feedback.pushInfo` por `print`. De la misma manera, el método `feedback.isCanceled()` sólo es válido para parar el proceso en entorno QGIS.

### TW1 DESCARGAR TWITTER HISTORICO – REST

Esta herramienta descarga tweets geolocalizados de hasta una semana de antigüedad en el momento de la petición.

Para ello es necesario definir, además de las credenciales, un centro (coordenadas en EPSG:4326) y su área de influencia o buffer (en kilómetros). Opcionalmente se pueden aplicar más filtros para delimitar la búsqueda según palabras que aparezcan en el texto<sup>17</sup>, el idioma<sup>18</sup> del tweet o usuario, o buscar entre dos tweets con sus ids conocidas y emitidos en los 7 días anteriores a la petición.

<sup>17</sup> Véase <https://developer.twitter.com/en/docs/tweets/search/guides/standard-operators>

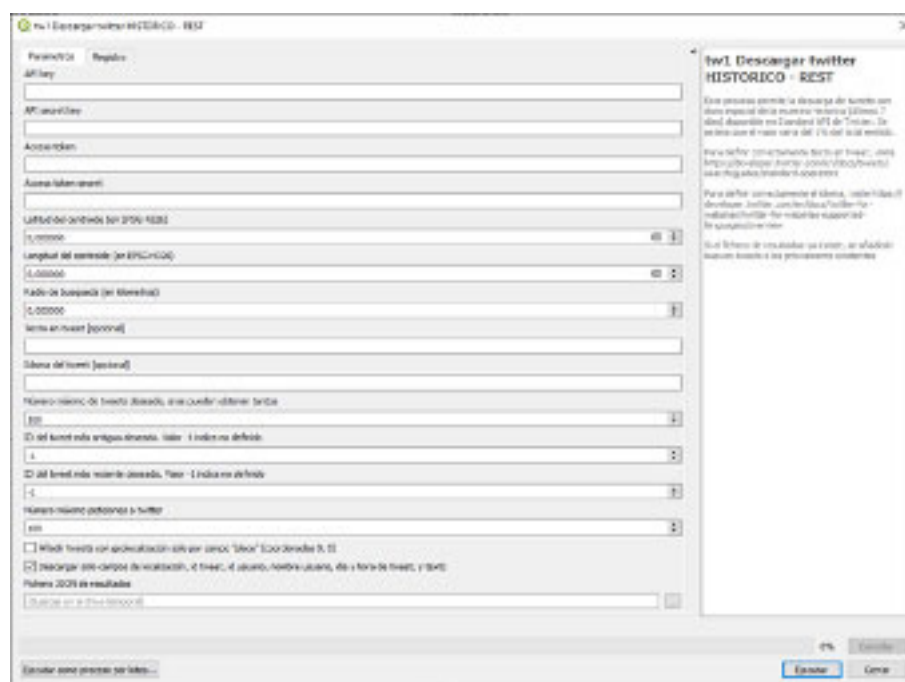
<sup>18</sup> Véase <https://developer.twitter.com/en/docs/twitter-for-websites/twitter-for-websites-supported-languages/overview>

Esta herramienta realiza varias peticiones a la API de Twitter, dentro de los límites establecidos<sup>19</sup>. En cada petición se va a recibir un máximo de 100 tweets, de los cuales sólo se escribirán en el fichero .json (especifíquese .json al final del nombre del fichero) aquellos tweets que tengan coordenadas, o campo place (con coordenada 0, 0) si así se especifica. De la misma manera, se puede solicitar que sólo escriba los tweets en su versión ligera (localización, id del tweet, id del usuario, nombre del usuario en el momento de emitir el tweet, día y hora del tweet<sup>20</sup>, y texto).

Se añaden nuevos tweets al fichero de resultados. En caso de no existir ese fichero, lo crea.

Esta herramienta finalizará al llegar al número de tweets deseados por el usuario, al cumplir el número máximo de peticiones pedidas por el usuario, o en caso de apretar el botón “Cancelar”.

En la Figura X.8 se muestra el cuadro de diálogo de la herramienta tw1 Descargar twitter HISTORICO – REST en entorno QGIS.



**Figura X.8:** Cuadro de diálogo de la herramienta tw1 Descargar twitter HISTORICO – REST

### **Código tw1 Descargar twitter HISTORICO - REST**

<sup>19</sup> Se pueden hacer hasta 180 peticiones cada 15 minutos.

<sup>20</sup> En huso UTC

```
# -*- coding: utf-8-sig -*-

from qgis.PyQt.QtCore import QApplication
from qgis.core import (QgsProcessing,
                       QgsProcessingAlgorithm,
                       QgsProcessingParameterString,
                       QgsProcessingParameterNumber,
                       QgsProcessingParameterBoolean,
                       QgsProcessingParameterFileDestination)

class twitter_download(QgsProcessingAlgorithm):
    def tr(self, string):
        return QApplication.translate('Processing', string)

    def createInstance(self):
        return twitter_download()

    def name(self):
        return 'twitter_download'

    def displayName(self):
        return self.tr('tw1 Descargar twitter HISTORICO - REST')

    def group(self):
        return self.tr('Twitter')

    def groupId(self):
        return 'twitter'

    def shortHelpString(self):
        return self.tr("Este proceso permite la descarga de tweets con dato espacial\
de la muestra historica (últimos 7 días) disponible\
en Standard API de Twitter. Se estima\
que el valor seria del 1% del total emitido.\n\
Para definir correctamente texto en tweet, visite\
https://developer.twitter.com/en/docs/tweets/search/guides/standard-operators\n\
Para definir correctamente el idioma, visite\
https://developer.twitter.com/en/docs/twitter-for-websites/twitter-for-websites-supported-languages/overview\n\
Si el fichero de resultados ya existe, se añadirán\
nuevos tweets a los\
previamente existentes")

    def initAlgorithm(self, config=None):
        self.addParameter(
            QgsProcessingParameterString(
                'INPUT_API_key',
                self.tr('API key'),
            )
        )

        self.addParameter(
            QgsProcessingParameterString(
                'INPUT_API_secret_key',
                self.tr('API secret key'),
            )
        )
```

```
self.addParameter(  
    QgsProcessingParameterString(  
        'INPUT_Access_token',  
        self.tr('Access token'),  
    )  
)  
  
self.addParameter(  
    QgsProcessingParameterString(  
        'INPUT_Access_token_secret',  
        self.tr('Access token secret'),  
    )  
)  
  
self.addParameter(  
    QgsProcessingParameterNumber(  
        'INPUT_coordY',  
        self.tr('Latitud del centroide (en EPSG:4326)'),  
        type = QgsProcessingParameterNumber.Double,  
        minValue=-90,  
        maxValue=90  
    )  
)  
  
self.addParameter(  
    QgsProcessingParameterNumber(  
        'INPUT_coordX',  
        self.tr('Longitud del centroide (en EPSG:4326)'),  
        type = QgsProcessingParameterNumber.Double,  
        minValue=-180,  
        maxValue=180,  
    )  
)  
  
self.addParameter(  
    QgsProcessingParameterNumber(  
        'INPUT_radium',  
        self.tr('Radio de busqueda (en kilometros)'),  
        type = QgsProcessingParameterNumber.Double,  
        minValue=0  
    )  
)  
  
self.addParameter(  
    QgsProcessingParameterString(  
        'INPUT_text',  
        self.tr('Texto en tweet'),  
        optional=True,  
    )  
)  
  
self.addParameter(  
    QgsProcessingParameterString(  
        'INPUT_lang',  
        self.tr('Idioma del tweet'),  
        optional=True,  
    )  
)  
  
self.addParameter(  
    QgsProcessingParameterNumber(  
        'INPUT_max',  
        self.tr('Número máximo de tweets deseado, si se pueden obtener
```



```
tantos'),
        minValue=0,
        defaultValue = 100,
    )
)

self.addParameter(
    QgsProcessingParameterNumber(
        'INPUT_sinceID',
        self.tr('ID del tweet más antiguo deseado. Valor -1 indica
no definido'),
        minValue=-1,
        defaultValue = -1,
    )
)

self.addParameter(
    QgsProcessingParameterNumber(
        'INPUT_lastID',
        self.tr('ID del tweet más reciente deseado. Valor -1 indica
no definido'),
        minValue=-1,
        defaultValue = -1,
    )
)

self.addParameter(
    QgsProcessingParameterNumber(
        'INPUT_iter',
        self.tr('Número máximo peticiones a twitter'),
        minValue=1,
        defaultValue = 100,
    )
)

self.addParameter(
    QgsProcessingParameterBoolean(
        'INPUT_place',
        self.tr('Añadir tweets con geolocalización solo por campo
"place" (coordenadas 0, 0)'),
    )
)

self.addParameter(
    QgsProcessingParameterBoolean(
        'INPUT_lite',
        self.tr('Descargar solo campos de localización, id tweet, id
usuario, nombre usuario, día y hora de tweet, y texto'),
        defaultValue = True
    )
)

self.addParameter(
    QgsProcessingParameterFileDestination(
        'OUTPUT_file',
        self.tr('Fichero JSON de resultados'),
        createByDefault = False
    )
)

def processAlgorithm(self, parameters, context, feedback):
    #Captura de valores de la interfaz grafica
    ckey = self.parameterAsString(
```

```
        parameters,  
        'INPUT_API_key',  
        context  
    )  
  
    csecret = self.parameterAsString(  
        parameters,  
        'INPUT_API_secret_key',  
        context  
    )  
  
    atoken = self.parameterAsString(  
        parameters,  
        'INPUT_Access_token',  
        context  
    )  
  
    asecret = self.parameterAsString(  
        parameters,  
        'INPUT_Access_token_secret',  
        context  
    )  
  
    coordX = self.parameterAsDouble(  
        parameters,  
        'INPUT_coordX',  
        context  
    )  
  
    coordY = self.parameterAsDouble(  
        parameters,  
        'INPUT_coordY',  
        context  
    )  
  
    radium = self.parameterAsDouble(  
        parameters,  
        'INPUT_radium',  
        context  
    )  
  
    text = self.parameterAsString(  
        parameters,  
        'INPUT_text',  
        context  
    )  
  
    language = self.parameterAsString(  
        parameters,  
        'INPUT_lang',  
        context  
    )  
  
    max_tweets = self.parameterAsInt(  
        parameters,  
        'INPUT_max',  
        context  
    )  
  
    since_tweet = self.parameterAsInt(  
        parameters,  
        'INPUT_sinceID',
```

```
        context
    )

    last_tweet = self.parameterAsInt(
        parameters,
        'INPUT_lastID',
        context
    )

    max_iter = self.parameterAsInt(
        parameters,
        'INPUT_iter',
        context
    )

    place = self.parameterAsBool(
        parameters,
        'INPUT_place',
        context
    )

    lite = self.parameterAsBool(
        parameters,
        'INPUT_lite',
        context
    )

    result_file = self.parameterAsString(
        parameters,
        'OUTPUT_file',
        context
    )

    feedback.pushInfo(result_file)

    #Esta parte del código se puede ejecutar en cualquier consola de
Python, indicando correctamente los parámetros anteriores
    import datetime, dateutil.parser, datetime, json, tweepy, time

    def prepare_tweet(data_in, fields_in):
        json_new = {"type": "Feature", "geometry": data_in["coordinates"],
                    "properties": {}}

        data.pop("coordinates", None)

        if fields_in == []:
            json_new["properties"] = data_in
        else:
            for i in fields:
                j = i.split(".")
                try:
                    if len(j) == 1:
                        json_new["properties"][i] = data_in[i]
                    elif len(j) == 2:
                        json_new["properties"][i] = data_in[j[0]][j[1]]
                except:
                    pass

        json_new = json.dumps(json_new)

        return json_new
```

[illegible]



```

while add_another_tweet and goon:
    tweet = tweets[num_tweet]
    last_tweet = tweet["id"] - 1
    num_tweet += 1

    if isinstance(tweet["coordinates"],dict)
or    (place    and    isinstance(tweet["place"],dict)    and    not
isinstance(tweet["coordinates"],dict)):
        if place and isinstance(tweet["place"],dict)
and not isinstance(tweet["coordinates"],dict):
            tweet["coordinates"] ={"type": "Point",
"coordinates": [0, 0]}

                                new_date = dateutil.parser.
parse(unicode(tweet["created_at"]))
                                tweet["created_at"] = new_date.strftime("%Y-
%m-%d %H:%M:%S")

                                if lite:
                                    toJSON = prepare_tweet(tweet, ["id_str",
"user.id_str", "user.screen_name", "created_at", "full_text", "place.name",
"place.country"])
                                else:
                                    toJSON = prepare_tweet(tweet, [])

                                if count_tweets > 0:
                                    results.write("\n")

                                results.write(toJSON)

                                count_tweets += 1

                                if num_tweet >= total_tweets:
                                    add_another_tweet = False

                                if count_tweets >= max_tweets:
                                    goon = False

                                feedback.pushInfo(str("\t{}. Vuelta {}/{ } buscando
tweets. Encontrados { } tweets con dato espacial".format(datetime.datetime.
now(), iteration, max_iter, count_tweets)))

                                if feedback.isCanceled(): #SOLO EN QGIS, Abortar proceso
                                    goon = False

                                else:
                                    goon = False

                                if iteration >= max_iter:
                                    goon = False

                                except:
                                    pass

                                feedback.pushInfo(str("\nSE HAN ENCONTRADO { } TWEETS CON DATOS
ESPACIALES".format(count_tweets)))

                                return {}

```

## TW2 DESCARGAR TWITTER STREAMING

Esta herramienta descarga tweets geolocalizados en el momento en que se realizan mediante una conexión al servicio de Streaming de Standard API.

Para ello es necesario definir, además de las credenciales, un área de búsqueda rectangular, delimitada por las coordenadas (en EPSG:4326) de la esquina inferior izquierda y la esquina superior derecha.

Solo se escribirán en el fichero *.json* (especifíquese *.json* al final del nombre del fichero) aquellos tweets que tengan coordenadas, o campo place (con coordenada 0, 0) si así se especifica. De la misma manera, se puede solicitar que solo escriba los tweets en su versión ligera (localización, id del tweet, id del usuario, nombre del usuario en el momento de emitir el tweet, día y hora de emisión del tweet<sup>21</sup>, y texto).

Se añade nuevos tweets al fichero de resultados. En caso de no existir ese fichero, lo crea.

Esta herramienta finalizará cuando el usuario aprete al botón “Cancelar”, realizando la cancelación en los siguientes 60 segundos<sup>22</sup>. El código se desconecta cada 60 segundos para saber si el usuario ha solicitado cancelar la búsqueda. En caso de ser así, el proceso de para; en caso contrario, se vuelve a reconectar.

En la Figura X.9 se muestra el cuadro de dialogo de la herramienta tw2 Descargar twitter STREAMING en entorno QGIS.

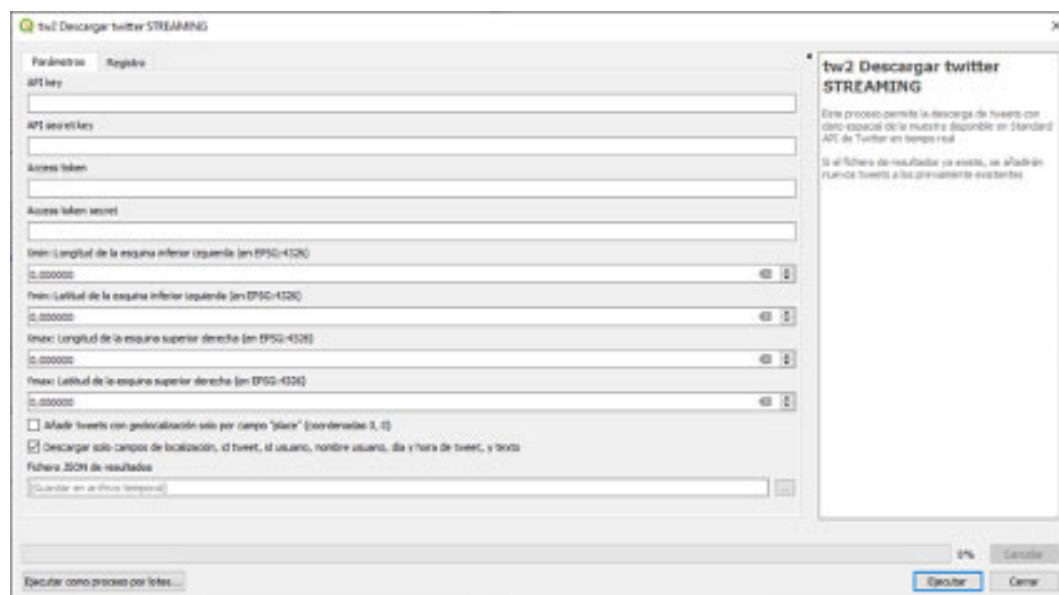


Figura X.9: Cuadro de dialogo de la herramienta tw2 Streaming

### Código tw2 Descargar twitter STREAMING

<sup>21</sup> En huso UTC

<sup>22</sup> Se pueden realizar un total de 15 conexiones al servicio de Streaming de Twitter cada 15 minutos

```
# -*- coding: utf-8-sig -*-

from qgis.PyQt.QtCore import QApplication
from qgis.core import (QgsProcessing,
                       QgsProcessingAlgorithm,
                       QgsProcessingParameterString,
                       QgsProcessingParameterNumber,
                       QgsProcessingParameterBoolean,
                       QgsProcessingParameterFileDestination)

class twitter_streaming(QgsProcessingAlgorithm):
    def tr(self, string):
        return QApplication.translate('Processing', string)

    def createInstance(self):
        return twitter_streaming()

    def name(self):
        return 'twitter_streaming'

    def displayName(self):
        return self.tr('tw2 Descargar twitter STREAMING')

    def group(self):
        return self.tr('Twitter')

    def groupId(self):
        return 'twitter'

    def shortHelpString(self):
        return self.tr("Este proceso permite la descarga de tweets con dato  
espacial\nde la muestra disponible en Standard API de Twitter  
en tiempo real\n\nSi el fichero de resultados ya existe, se añadirán  
nuevos tweets a los\previamente existentes")

    def initAlgorithm(self, config=None):
        self.addParameter(
            QgsProcessingParameterString(
                'INPUT_API_key',
                self.tr('API key')
            )
        )

        self.addParameter(
            QgsProcessingParameterString(
                'INPUT_API_secret_key',
                self.tr('API secret key')
            )
        )

        self.addParameter(
            QgsProcessingParameterString(
                'INPUT_Access_token',
                self.tr('Access token')
            )
        )

        self.addParameter(
            QgsProcessingParameterString(
```

```
        'INPUT_Access_token_secret',
        self.tr('Access token secret')
    )
)

self.addParameter(
    QgsProcessingParameterNumber(
        'INPUT_coordX_min',
        self.tr('Xmin: Longitud de la esquina inferior izquierda (en
EPSG:4326)'),
        type = QgsProcessingParameterNumber.Double,
        minValue=-180,
        maxValue=180,
    )
)

self.addParameter(
    QgsProcessingParameterNumber(
        'INPUT_coordY_min',
        self.tr('Ymin: Latitud de la esquina inferior izquierda (en
EPSG:4326)'),
        type = QgsProcessingParameterNumber.Double,
        minValue=-90,
        maxValue=90
    )
)

self.addParameter(
    QgsProcessingParameterNumber(
        'INPUT_coordX_max',
        self.tr('Xmax: Longitud de la esquina superior derecha (en
EPSG:4326)'),
        type = QgsProcessingParameterNumber.Double,
        minValue=-180,
        maxValue=180,
    )
)

self.addParameter(
    QgsProcessingParameterNumber(
        'INPUT_coordY_max',
        self.tr('Ymax: Latitud de la esquina superior derecha (en
EPSG:4326)'),
        type = QgsProcessingParameterNumber.Double,
        minValue=-90,
        maxValue=90
    )
)

self.addParameter(
    QgsProcessingParameterBoolean(
        'INPUT_place',
        self.tr('Añadir tweets con geolocalización solo por campo
"place" (coordenadas 0, 0)'),
    )
)

self.addParameter(
    QgsProcessingParameterBoolean(
        'INPUT_lite',
        self.tr('Descargar solo campos de localización, id tweet, id
usuario, nombre usuario, día y hora de tweet, y texto'),
    )
)
```



```
        defaultValue = True
    )
)

self.addParameter(
    QgsProcessingParameterFileDestination(
        'OUTPUT_file',
        self.tr('Fichero JSON de resultados'),
    )
)

def processAlgorithm(self, parameters, context, feedback):
    #Captura de valores de la interfaz grafica
    ckey = self.parameterAsString(
        parameters,
        'INPUT_API_key',
        context
    )

    csecret = self.parameterAsString(
        parameters,
        'INPUT_API_secret_key',
        context
    )

    atoken = self.parameterAsString(
        parameters,
        'INPUT_Access_token',
        context
    )

    asecret = self.parameterAsString(
        parameters,
        'INPUT_Access_token_secret',
        context
    )

    coordX_min = self.parameterAsDouble(
        parameters,
        'INPUT_coordX_min',
        context
    )

    coordX_max = self.parameterAsDouble(
        parameters,
        'INPUT_coordX_max',
        context
    )

    coordY_min = self.parameterAsDouble(
        parameters,
        'INPUT_coordY_min',
        context
    )

    coordY_max = self.parameterAsDouble(
        parameters,
        'INPUT_coordY_max',
        context
    )

    place = self.parameterAsBool(
```

```
        parameters,
        'INPUT_place',
        context
    )

    lite = self.parameterAsBool(
        parameters,
        'INPUT_lite',
        context
    )

    result_file = self.parameterAsString(
        parameters,
        'OUTPUT_file',
        context
    )

    #Esta parte del código se puede ejecutar en cualquier consola de Python, indicando correctamente los parámetros anteriores
    import datetime, dateutil.parser, datetime, time, json, threading
    from html.parser import HTMLParser
    from tweepy import Stream
    from tweepy import OAuthHandler
    from tweepy.streaming import StreamListener

    def prepare_tweet(data_in, fields_in):
        json_new = {"type": "Feature", "geometry": data_in["coordinates"],
                    "properties": {}}

        data_in.pop("coordinates", None)

        if fields_in == []:
            json_new["properties"] = data_in
        else:
            for i in fields_in:
                j = i.split(".")
                try:
                    if len(j) == 1:
                        json_new["properties"][i] = data_in[i]
                    elif len(j) == 2:
                        json_new["properties"][i] = data_in[j[0]][j[1]]
                except:
                    pass

        json_new = json.dumps(json_new)

        return json_new

    class listener(StreamListener):
        def __init__(self, time_limit=61):
            self.start_time = time.time()
            self.limit = time_limit

        def write_tweet(self, data, fields):
            try:
                file_results = open(result_file, "a")

            except Exception:
                feedback.pushInfo("No escribir el fichero de resultados")

            else:
```

```

        to_file = prepare_tweet(data, fields)

        file_results.write("{}\n".format(to_file))
        file_results.close()

    def on_data(self, data):
        if (time.time() - self.start_time) < self.limit: #Controlador
de tiempo para parada de Streaming en QGIS
            tweet = json.loads(HTMLParser().unescape(data))

            if isinstance(tweet["coordinates"],dict) or
(place and isinstance(tweet["place"],dict) and not
isinstance(tweet["coordinates"],dict)):
                if place and isinstance(tweet["place"],dict) and not
isinstance(tweet["coordinates"],dict): #Si se aceptan tweets solo con place,
añadir coordenada 0,0
                    tweet["coordinates"] ={"type": "Point", "coordinates":
[0, 0]}

                new_date = dateutil.parser.parse(unicode(tweet["created_
at"]))
                tweet["created_at"] = new_date.strftime("%Y-%m-%d
%H:%M:%S")

                if lite:
                    t = threading.Thread(target=self.write_tweet,
args=(tweet, ["id_str", "user.id_str", "user.screen_name", "created_at",
"full_text", "place.name", "place.country"]))
                else:
                    t = threading.Thread(target=self.write_tweet,
args=(tweet, []))

                t.start() #Activación de hilos para evitar bloqueo
por saturación
                t.join(5) #El hilo solo puede estar activo 5 segundos,
sino se elimina

                if t.is_alive():
                    t.kill()
                else:
                    twitterStream.disconnect() #Desconectar Streaming. Se
volverá a conectar si no hay orden de parada en QGIS

class twitter_connection(object):
    def get_ath(self):
        try:
            auth = OAuthHandler(ckey, csecret)
            auth.set_access_token(accessToken, asecret)
            return auth
        except Exception:
            feedback.pushInfo("No se ha podido realizar la conexión
con Twitter. Revise sus credenciales")
            raise SystemError

#Código principal
goon = True

if coordX_min < coordX_max and coordY_min < coordY_max:
    feedback.pushInfo("{} INICIANDO LA ESCUCHA Y GUARDADO DE TWEETS\n".
format(datetime.datetime.now()))

    while goon: #Cada 60 segundos se desconecta de Twitter para ver

```

```
si ha recibido orden de parada de QGIS. Solo se pueden hacer 15 conexiones en
periodos de 15 minutos
    try:
        conT = twitter_connection()
        auth = conT.get_ath()
    except:
        goon = False
    else:
        feedback.pushInfo("\tConexión a Twitter a las {}".
format(datetime.datetime.now()))
        twitterStream = Stream(auth, listener())
        twitterStream.filter(locations = (coordX_min, coordY_
min, coordX_max, coordY_max))

        if feedback.isCanceled(): #SOLO EN QGIS, Abortar proceso
            goon = False
    else:
        feedback.pushInfo("Las coordenadas no cumplen los criterios.
Revíselo")
    return {}
```

## C. Requisitos

Python 3.7.0 (versión instalada en QGIS 3.10 A Coruña)

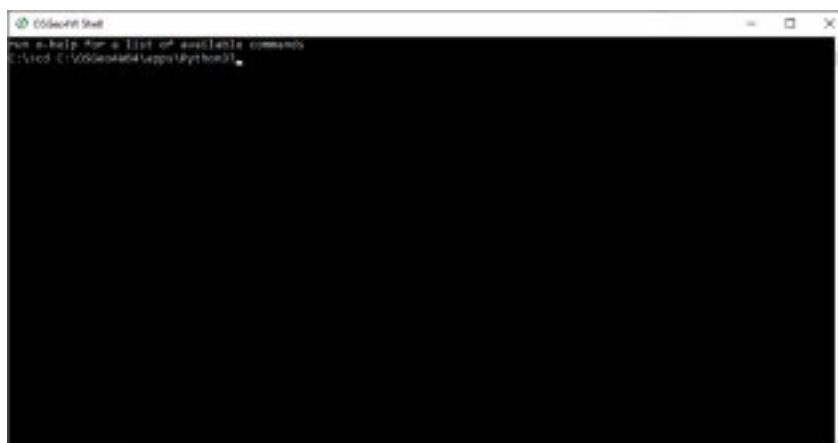
Librerías, a instalar en este orden: tweepy, setuptools, html.parser

### NOTA DE INSTALACIÓN LIBRERÍAS EN PYTHON DE QGIS EN WINDOWS

Para instalar las librerías requeridas en la versión de Python que usa QGIS de OSGeo4W, se deben realizar los siguientes pasos:

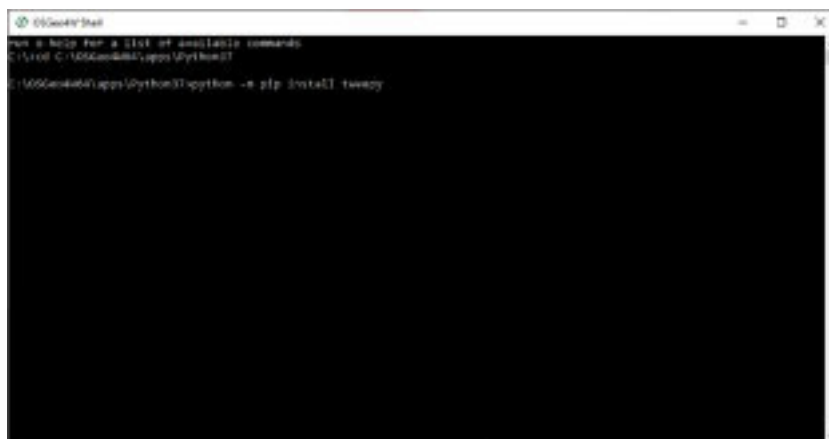
- 1. Copiar la carpeta "pip" de otra versión de Python 3.7.x, cuya ruta suele ser "C:\Python37\Lib\site-packages" y pegarla en la carpeta equivalente de OSGeo4W, que suele ser "C:\OSGeo4W64\apps\Python37\lib\site-packages"
- 2. Abrir la aplicación "OSGeo4W Shell"
- 3. Dirigir la consola a la ruta de OSGeo4W que contenga Python. Si la ruta es "C:\OSGeo4W64\apps\Python37" escribid "cd C:\OSGeo4W64\apps\Python37", tal y como se muestra en la Figura X.90.





**Figura X.90:** Dirigir consola a ruta de Python de OSGeo4W.

- 4. Instalar las librerías usando el comando “python -m pip install <nombre de la librería>”. Donde <nombre de la librería> debe ser tweepy, setuptools o html.parser. En la Figura X.91 se muestra para el caso de tweepy.



**Figura X.91:** Ejemplo de instalación de librería tweepy con pip



