

# **MODELO DE MOVILIDAD DE EMT: PLANIFICACIÓN A PARTIR DE FUENTES BIG DATA**

**Cristina Valdés Serrano**

Responsable del Modelo de Movilidad. Área de Diseño e Ingeniería de Transporte,  
Dirección de Servicios de Transporte. EMT Madrid

**Ana Pérez del Olmo**

Jefa de Área de Diseño e Ingeniería de Transporte, Dirección de Servicios de Transporte,  
EMT Madrid

## **RESUMEN**

El Modelo de Movilidad desarrollado por EMT es un modelo multimodal que incluye vehículo privado y transporte público. Si bien su ámbito de actuación es la ciudad de Madrid, el Modelo abarca también el entorno metropolitano.

La red es una malla completa, basada en la cartografía de HERE. La oferta de transporte público incluye todos los sistemas de transporte colectivo – METRO, Cercanías, autobuses interurbanos y autobuses de EMT- , sumando un total de 405 líneas.

El aspecto más innovador de esta herramienta es la metodología seguida para la obtención de las matrices origen-destino. Éstas se han obtenido a partir de la combinación de datos de telefonía (CDR's) y validaciones de la tarjeta de transporte público. Esta combinación de datos ha permitido obtener un amplio abanico de matrices modales para diferentes periodos del año, distintos tipos de día y diferentes intervalos horarios.

La zonificación utilizada – 1259 zonas - es la definida por el Consorcio Regional de Transportes de Madrid para la Encuesta Domiciliaria de Movilidad realizada en 2018 (EDM 18) – 605 zonas para la ciudad de Madrid y 654 para el área metropolitana.

La obtención de unas matrices origen-destino tan detalladas, combinado con el diferente grado de precisión geográfica de origen y destino del viaje de las dos fuentes de datos utilizadas (CDR's y validaciones de la tarjeta de transporte público), ha supuesto un reto importante.

## **1. OBJETIVOS Y ESTRUCTURA**

### **1.1. Objetivos y aspectos principales**

El Modelo de Movilidad de Madrid es una herramienta de planificación para EMT que incluye tanto transporte público como vehículo privado. El Proyecto de construcción de este Modelo se inició en abril de 2017, terminándose en octubre de 2019.

El Objetivo Principal de este Modelo de Movilidad es dotar a EMT, el operador de movilidad pública de la ciudad de Madrid (100% propiedad del Ayuntamiento), de una herramienta de planificación que le permita una mejor planificación, así como propuestas relacionadas con los servicios de transporte que gestiona.

El principal aspecto innovador de este Proyecto es la obtención de matrices origen-destino a partir de fuentes de datos como los CDR's obtenidos de la telefonía móvil y las tarjetas inteligentes de transporte.

## **1.2. Etapas del Proyecto**

El Proyecto se puede estructurar en 3 fases:

- **Desarrollo del Modelo de Oferta:** el Modelo, si bien se centra en la ciudad de Madrid, abarca el conjunto de la región, incluyendo tanto la red viaria como la oferta de servicios de transporte público (EMT, METRO, autobuses interurbanos y Cercanías). La descripción de esta red está en el capítulo 2.
- **Desarrollo del Modelo de Demanda:** en este caso se ha planteado un amplio abanico de matrices origen-destino tanto para vehículo privado como para transporte público, así como para diferentes periodos del año (invierno, verano y entretiempo), días de la semana (l-j, viernes, sábados y domingos&festivos) e intervalos horarios (punta AM, punta PM, punta MEDIODÍA y horas VALLE).

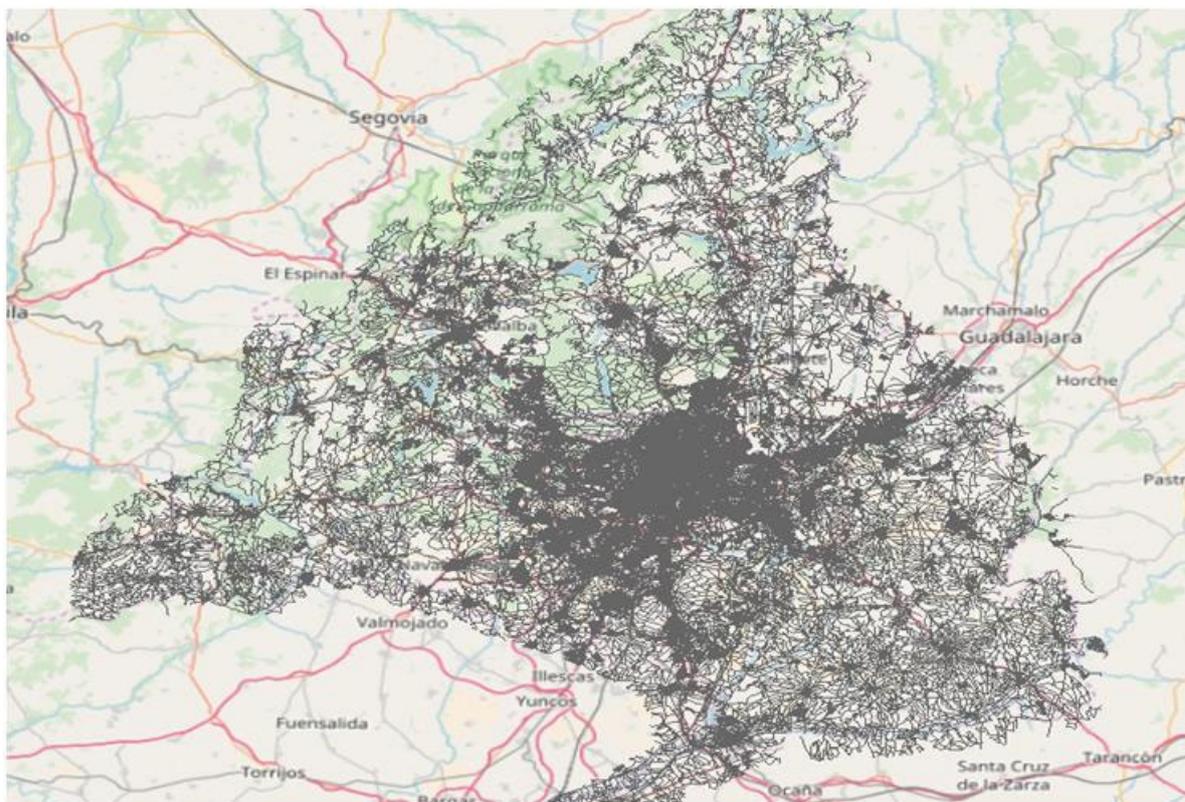
Para el cálculo de estas matrices, se puso en marcha una plataforma BIG DATA, basada en una Distribución HORTON HDP 2.6, Actualmente esta Plataforma se está utilizando para otros proyectos en curso en EMT, como la estimación de aforo en los autobuses.

- **Validación y Calibrado de resultados.**

Los resultados obtenidos al asignar dichas matrices se han contrastado con los datos de aforos de tráfico, así como subidos y bajados en transporte público. Primeramente han servido para afinar algunos parámetros del Modelo, y posteriormente para calibrar dichas matrices.

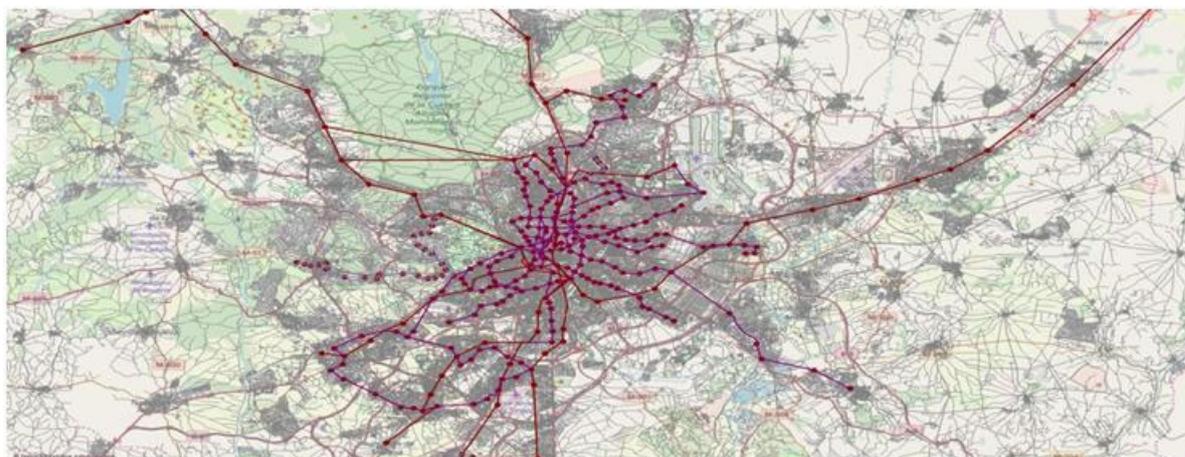
## **2. LA RED DE TRANSPORTE: DESCRIPCIÓN**

La red viaria se ha construido a partir de la cartografía HERE para el conjunto de la región – 2.000 km. de vías de alta capacidad, 5.500 km. de carreteras y más de 22.000 km. de red urbana.

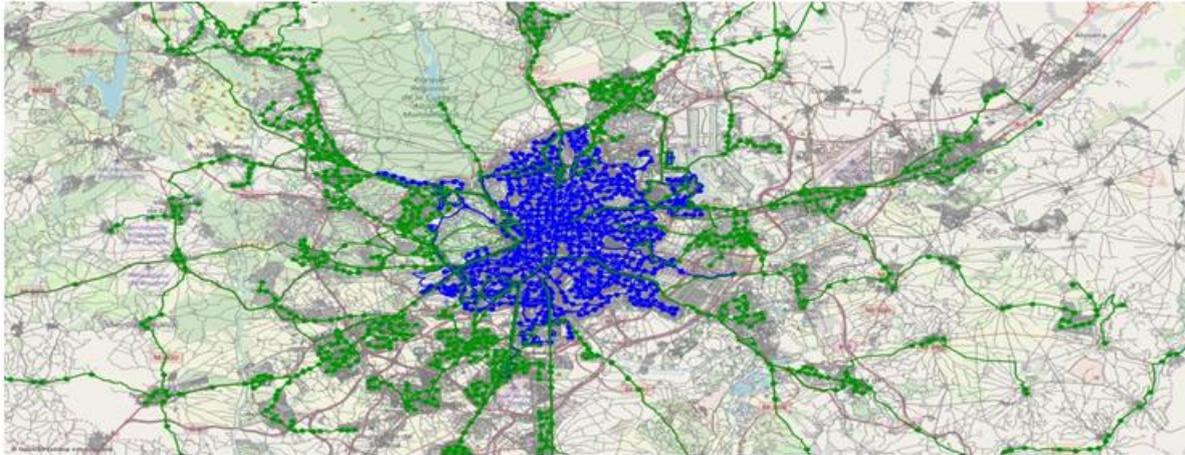


**Fig. 1. Red viaria**

También se han incluido las diferentes redes de (EMT, Metro, Cercanías, Metro ligero y autobuses interurbanos), junto con sus correspondientes horarios y/o frecuencias.



**Fig. 2. Red de transporte público: CERCANÍAS Y METRO**



**Fig. 3. Red de transporte público: Autobús urbano e interurbano**

Esta oferta supone, en total, 393 líneas, más de 9.000 paradas y 62.000 servicios diarios.

La zonificación considerada es la definida por el Consorcio de Transportes de Madrid para la encuesta de movilidad llevada a cabo en 2016, 1259 zonas, de las cuales 605 zonas corresponden a la ciudad de Madrid mientras las otras 654 pertenecen a los otros 199 municipios que integran el conjunto de la región.

### **3. EL MODELO DE DEMANDA**

#### **3.1. Metodología para la obtención de matrices origen – destino (OD)**

Los enfoques tradicionales de obtención de matrices O/D se basan en encuestas de hogares y / o recuentos de tráfico. Las encuestas origen / destino y las encuestas de hogares para obtención de orígenes y destinos implican una recopilación de datos costosa y, consecuentemente, tienen tamaños de muestra limitados y frecuencias de actualización más bajas de lo deseado, entre 5 y 10 años incluso en las ciudades más desarrolladas (Toole et al, 2015). De hecho, según Iqbal et al (2016), son propensas a sesgos de muestreo y errores de notificación.

Al mismo tiempo, solo proporcionan una instantánea del rendimiento del tráfico, ya que se realizan en un momento puntual y no de manera continua en el tiempo (White et al, 2002).

Frente a estos métodos clásicos de recopilación de matrices O-D, este trabajo se centra en la obtención de estas matrices de movilidad utilizando fuentes BIG DATA como los datos de telefonía o las tarjetas de transporte. La obtención de las matrices origen-destino mediante combinación de estas fuentes ha conllevado diferentes etapas, que se detallan a continuación.

El primer paso ha sido la obtención de matrices de movilidad general O-D para la Comunidad de Madrid a partir de telefonía móvil. Estos datos de telefonía proporcionan una información completa sobre viajes y distribución espacial con las siguientes ventajas en relación a los métodos convencionales: alto volumen de muestra, buena distribución,

volumen de datos a lo largo de todo el año, lo que nos ha permitido considerar diferentes períodos del año, semana o día sin incrementar demasiado el coste. Además, permite capturar información sobre la movilidad de los no residentes en la ciudad de forma menos sesgada.

El segundo paso ha consistido en la obtención de las matrices O/D de transporte público de la ciudad de Madrid a partir de los datos recogidos en las tarjetas inteligentes de transporte público. Para ello, hemos realizado un cruce de información de los datos almacenados (con anonimización previa) con el resto de datos almacenados en la plataforma Big Data.

La información almacenada en estas tarjetas permite la estimación de los viajes realizados en diferentes modos de transporte público, teniendo en cuenta diferentes matices que se explicarán más adelante con más detalle.

Por último, una vez obtenidas ambas matrices OD (movilidad general y transporte público) el siguiente paso ha sido diseñar un procedimiento para obtener matrices de vehículos privados a partir de las anteriores, restando según diferentes zonas la matriz OD pública a la matriz de movilidad general. Este último paso, muy sencillo en apariencia, ha conllevado importantes dificultades derivadas de la diferente aproximación espacial de unos y otros datos.

### **3.2. Obtención de matrices O/D de movilidad general a partir de datos de telefonía**

Las matrices de O / D de movilidad general para la Comunidad de Madrid se han obtenido a partir de datos de telefonía móvil, mediante registros de actividad de teléfonos móviles o CDR (Call Detail Records).

Los datos utilizados para la obtención de estas matrices han sido datos de telefonía celular (datos geolocalizados desde dispositivos móviles, Grupo Orange España), datos de población residente y datos de movimientos fronterizos (FRONTUR, INE), datos de movimientos fronterizos y datos de uso del suelo (Sistema de Información de Ocupación del Suelo).

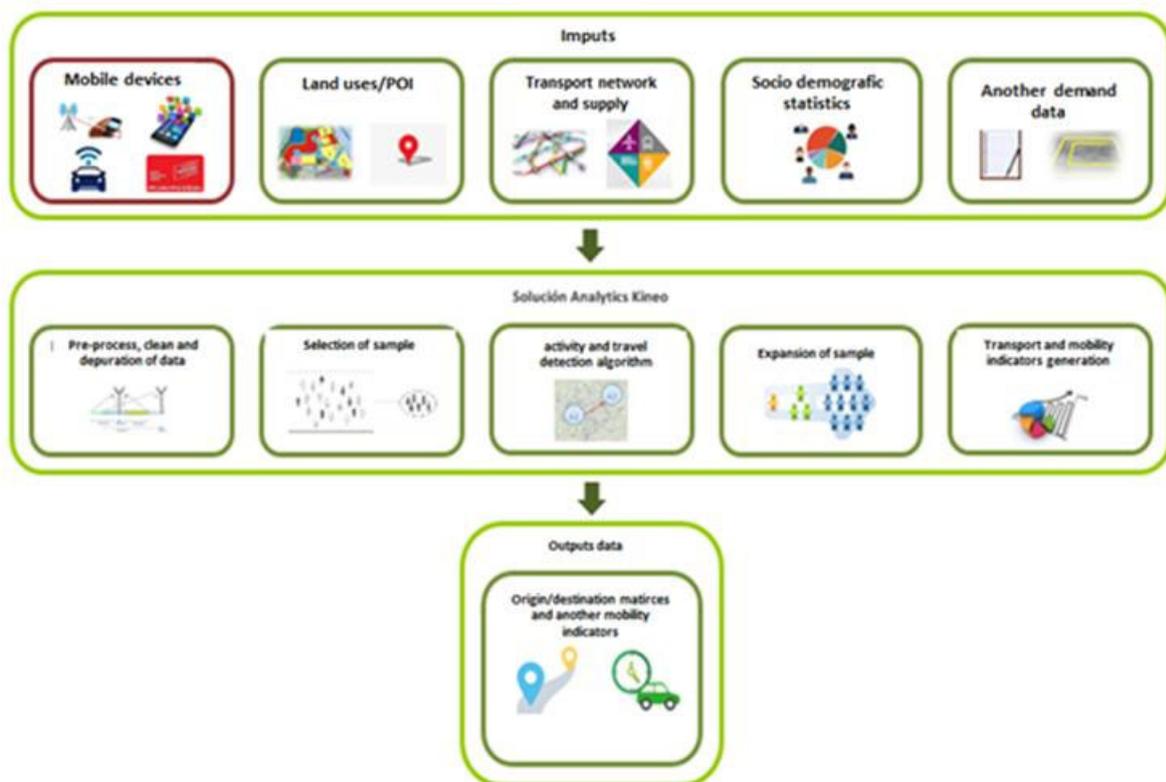
Al mismo tiempo, se ha tenido en cuenta:

- Población de estudio: residentes en España mayores de 10 años y personas residentes en el extranjero que visiten España.
- Zonificación: las matrices origen-destino estarán referidas a la zonificación establecida por el Consorcio de Transportes de Madrid para la Encuesta de Hogares de 2018, ZT1259.
- Viajes objeto de estudio: viajes con origen y destino en la Comunidad de Madrid con una distancia superior a los 1.000 metros. Este criterio de distancia permite filtrar la mayor parte de los viajes no mecanizados, que no están cubiertos por este estudio.

- Períodos de estudio: se define como período de estudio la combinación de mes o tipo de mes (invierno, verano y temporada media), día o tipo de día (día laborable, viernes, sábado y domingos y festivos) y franja horaria (pico de la mañana, pico de mediodía, pico de tarde y horas valle). Esta combinación resultó en 32 períodos de estudio. En cada uno de ellos, el nivel de desagregación de viajes es el siguiente:
  - Segmentación por género y edad (menores de 20 años, 20-44 años, 45-64 años, 65-79 años y mayores de 70 años)
  - Segmentación por lugar de residencia (municipio de Madrid, resto de la Comunidad de Madrid, resto de España y fuera de España)

La metodología ha seguido los siguientes pasos:

- Preprocesamiento y depuración de telefonía móvil (los datos base se han extraído de los meses de junio, julio, septiembre, octubre y noviembre de 2017)
- Construcción de la muestra potencial de usuarios de telefonía móvil
- Identificación del lugar de residencia de los usuarios residentes en España y características de los no residentes
- Abstracción de la actividad y diarios de viaje
- Aumento de la muestra a la población general
- Generación de las matrices origen-destino



**Fig. 4. Proceso de obtención de las matrices O/D generales**

### **3.3. Obtención de matrices O/D de transporte público mediante los registros de las tarjetas inteligentes de transporte**

Los datos de la red de telefonía nos dan la posibilidad de comprender mucho mejor los patrones de viaje dinámicos, lo cual tiene una gran cantidad de aplicaciones diferentes dentro de la gestión, el análisis y el apoyo a la toma de decisiones del tráfico y el transporte. Sin embargo, estas fuentes de datos tienen varias características clave que son diferentes de las fuentes de datos tradicionales y que deben manejarse con cuidado mientras se procesan los datos con fines de estimación y predicción. A diferencia de los sistemas de infraestructura fija para la recopilación de datos, los datos de señalización de teléfonos móviles no están limitados por ningún modo de transporte ni por ninguna región espacial específica. Esto no permite analizar la demanda de viajes y los tiempos de viaje según el modo de viaje (David Gundlegard et al., 2016). Esta es la razón por la que hemos utilizado la información de las tarjetas de transporte inteligentes

Hoy en día, la mayoría de los sistemas AFC (recolección automática de tarifas) en todo el mundo solo registran la información de embarque de los pasajeros y se pierden los datos de bajada, como la ruta del autobús, la parada de embarque y la hora de embarque. (Daming Li, Xinliang Zhao et al. 2011). La estimación de orígenes y destinos (O/D) de trayectos en transporte público es un producto importante del procesamiento de datos de tarjetas inteligentes.

En los últimos años, varios métodos de estimación O/D que utilizan el enfoque de encadenamiento de viajes han atraído mucha atención tanto de investigadores como de profesionales.

En este caso, para estimar la cadena modal (monoetapa o multietapa) para cada usuario, hemos considerado a los usuarios de la tarjeta de transporte público de Madrid con billetes mensuales, anuales y de 10 viajes, ya que para el billete simple no es posible inferir la salida del viajero debido al uso ocasional de los servicios de transporte público.

En cuanto a la metodología para la recolección de viajes a partir de las validaciones, el proceso seguido se basó en el enfoque de encadenamiento de etapas.

Para unir las validaciones de viaje utilizamos el programa que predice la parada de salida (si no la tenemos) en función de la distancia mínima de transferencia y el tiempo de viaje por modo, que se fijará en función del tipo de día y franja horaria (estos límites se determinarán en función de los datos de otras fuentes externas, como Simplycity, CDRs, etc. ).

Paralelamente, se desarrolló un algoritmo que sirve para encontrar la parada más cercana a la siguiente validación en los casos en que no existan distancias calculadas previamente para poder saber qué parada se utilizó para el traslado, teniendo en cuenta que el operador y la línea de salida deben ser los mismos que la línea de entrada.

Los escenarios adoptados han sido los siguientes:

- Cuando la diferencia horaria entre validaciones en una misma tarjeta es inferior a 45 minutos, las hemos considerado como etapas de un mismo viaje y las hemos encadenado en un mismo viaje.
- Las tarjetas de transporte se almacenan en el punto de validación (parte superior de acceso a la red de transporte público). Para ir más allá de este punto de acceso a la red hasta la zona de disparo de origen, hemos redistribuido los viajes entre las áreas vecinas a cada parada de la red. Las zonas de influencia se generan con dos criterios diferentes en cuanto al radio de acción:
  - En las áreas de metro y ferrocarriles (es decir, en estos operadores) se necesitan 600 metros.
  - Se toman 500 metros al resto de operadores.
  - Para la redistribución se utiliza el producto cartesiano individual para cada zona de origen y destino.
  - La redistribución se realiza a nivel de viajes.

Para cada uno de los periodos de estudio, se ha tenido en cuenta la distribución espacial de viajes entre áreas de la matriz integral provenientes de la matriz de telefonía móvil.

Dado que no ha sido posible disponer de información suficiente de los viajes en transporte público con origen o destino fuera de la ciudad de Madrid, hemos optado por considerar solo aquellos viajes con origen / destino dentro de la ciudad de Madrid, excluyendo el análisis de los viajes al resto de áreas en la Comunidad. Para éstos, se ha estimado la misma distribución de viajes que había en la última Encuesta Domiciliaria.

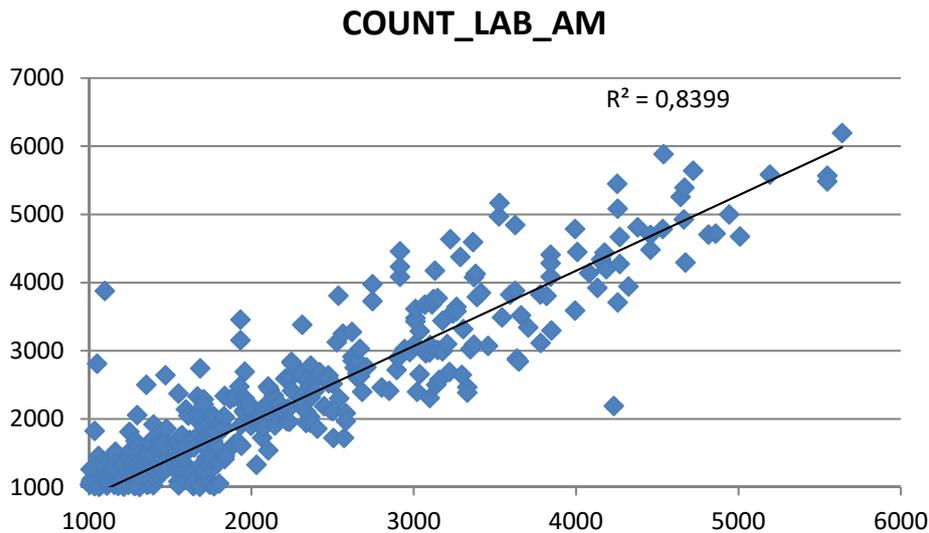
### **3.4. Obtención de las matrices O/D de vehículo privado**

El último paso ha consistido en la recogida de matrices de transporte privado. En este punto ya conocíamos las matrices integrales y las matrices de transporte público, por lo que procedimos de la siguiente manera:

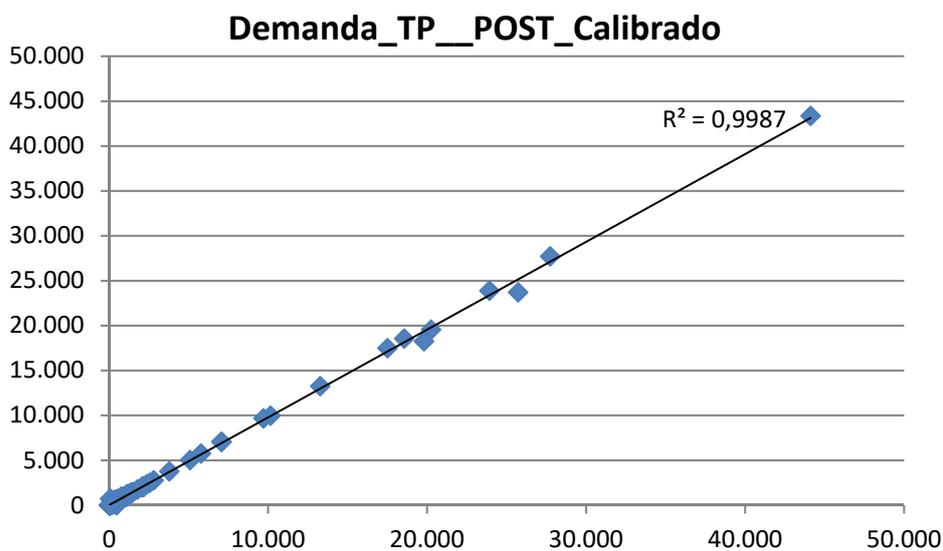
- De las matrices de transporte público O D se han eliminado aquellos viajes cuya distancia sea inferior a 1.000 metros, ya que estos viajes no estaban incluidos en la matriz integral O/D obtenida de la telefonía móvil.
- A cada matriz de O/D integral obtenida de la telefonía móvil (considerando solo los viajes entre áreas de la ciudad de Madrid), se ha eliminado su contraparte de la matriz de transporte público, dando como resultado la matriz de viajes en vehículo privado.
- Cuando lo hacemos, en algunas combinaciones, se obtiene un resultado negativo restando información entre CDR y O/D público.

Para minimizar la densidad de valores negativos que no cubre la versión original del algoritmo, es necesario modificar el algoritmo entre la creación del “buffer” y la asociación de zonas candidatas (que se realizará a nivel desagregado).

Por último, se han calibrado estas matrices usando los datos de tráfico y de viajeros subidos por línea. Las figuras 5 y 6 muestran los resultados para el periodo punta de la mañana de un día laborable de invierno.



**Fig. 5. Resultados calibrado matriz vehículo privado: periodo AM día laborable de invierno**

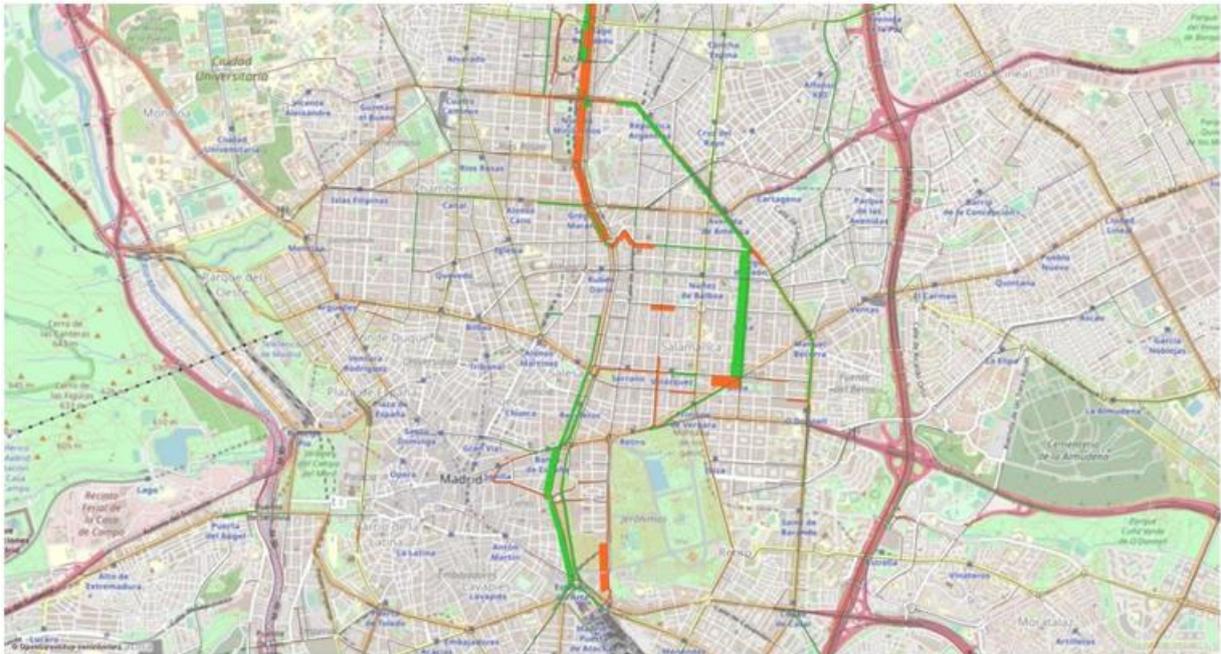


**Fig. 6. Resultados calibrado matriz transporte público: periodo AM día laborable de invierno**

#### 4. UTILIDADES Y CASOS DE ESTUDIO

El Modelo de Movilidad se ha utilizado y se está utilizando actualmente para diferentes escenarios de análisis, tanto desde la necesidad operativa (1 y 2) como para la investigación (3), tales como:

- 1) Cierre del túnel de Recoletos de la red de Cercanías: variación de la movilidad y necesidades de refuerzo mediante servicio alternativo de EMT.



**Fig. 7. Variación del tráfico con el cierre del túnel de Recoletos de Cercanías**

- 2) Definición y selección de itinerario de las nuevas líneas de autobús 00 (cero coste, cero emisiones): 001 Atocha Renfe – Moncloa y 002 Argüelles – Puerta de Toledo
- 3) Proyecto europeo MOMENTUM: este Proyecto, perteneciente al Programa H2020, tiene entre sus objetivos caracterizar los cambios en las pautas de movilidad derivados de las nuevas opciones emergentes de movilidad, así como incorporar en el Modelo de Movilidad opciones para incorporar el impacto de estas nuevas opciones.

#### AGRADECIMIENTOS

Queremos expresar nuestro agradecimiento a las empresas VIRTUALDESK, TARYET y NOMMON por el gran trabajo realizado tanto en el desarrollo y puesta en marcha de la plataforma como en el trabajo de integración y analítica para la obtención de las matrices origen-destino.

---

**REFERENCIAS**

DAMING, L., YONGJIE, L., XINLIANG, Z., HONGJUN, S. and NAN, Z. (2011). Estimating a Transit Passenger Trip Origin-Destination Matrix Using Automatic Fare Collection System. DASFAA Workshops, LNCS 6637, pp. 502–513.

GUNDLEGARD, D., RYDERGREN, C., BREYER, N. and RAJNA, B. (2016). Travel demand estimation and network assignment based on cellular network data. *Computer Communications* 95, pp 29–42.

IQBAL, M.S., CHOUDHURY, C.F., WANG, P. and GONZALEZ, M.(2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C*, 40, pp. 63–74

TOOLE, J.L., COLAK, S., STURT, B., ALEXANDER, B.S., EVSUKOFFC, A. and GONZALEZ, M. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C* 58, pp. 162–177

WHITE, J. and WELLS, I. (2002). Extracting origin destination information from mobile phone data. *Eleventh International Conference on Road Transport Information and Control*, 2002, pp. 30 – 34