

El *big data* y la inteligencia artificial revolucionan la forma de conocer la movilidad de las personas

An aerial view of a busy city street at dusk, with a network overlay of white nodes and lines connecting various points across the scene. The street is filled with pedestrians and cars, and the sky is a mix of purple and orange. The text is overlaid on the center of the image.

Mitma
analiza la
movilidad
con *big data*
e **inteligencia**
artificial



El Ministerio de Transportes, Movilidad y Agenda Urbana (Mitma) utiliza desde 2018 la tecnología *big data* aplicada a los registros anonimizados de telefonía móvil para analizar la movilidad de las personas. Conocer y caracterizar la demanda de forma precisa y actualizada resulta fundamental para mejorar nuestro sistema actual de transporte, adaptando la oferta de infraestructuras y servicios a la demanda existente y también a la demanda prevista según nos indican los modelos de transporte.

- **Texto:** Tania Gullón Muñoz-Repiso, Unidad de Análisis de Movilidad e Infraestructuras; Subdirección General de Planificación, Red Transeuropea y Logística. Miguel Picornell, Nommon.

Nuestra huella digital

La digitalización y el acceso cotidiano a nuevas tecnologías de movilidad está cambiando de manera radical la forma en la que viajamos. Hoy en día, nos parece extraño viajar en coche sin encender el **navegador de conducción con GPS** para consultar el camino más rápido a nuestro destino. Circulamos por carreteras desde las que multitud de **cámaras** nos observan y varias **espiras magnéticas** registran el paso de nuestras ruedas. Muchos de **nuestros coches** están registrando la meteorología o la rugosidad del pavimento por el que circulan y están generando y enviando datos a medida que sube el cuentakilómetros.

Cuando cogemos el metro, rara vez nos planteamos que la **tarjeta inteligente de transporte** que validamos a la entrada, y en algunas ocasiones también a la salida,

deja un registro digital de nuestro paso por el sistema de transporte público. Tampoco concebimos salir de casa sin llevar encima nuestro **teléfono móvil** y varias veces al día lo utilizamos para realizar llamadas, mandar mensajes, buscar en Internet o conectarnos a las redes sociales.

Dispositivos como los mencionados generan una **huella digital** que proyecta un reflejo muy fiel de nuestra actividad y movilidad cotidiana. Actualmente, gracias a las enormes posibilidades que ofrecen las nuevas tecnologías de **big data** e inteligencia artificial, siempre respetando el anonimato y la privacidad de los usuarios, se pueden llevar a cabo análisis más precisos y actualizados sobre la movilidad de la población, con una caracterización detallada de los patrones de actividad y movilidad, lo que nos ayuda a diseñar **mejores**

políticas públicas, más eficientes y sostenibles, para dar un mejor servicio a la ciudadanía.

Estudios de demanda: piedra angular de Mitma

Los estudios de demanda son una de las piedras angulares de Mitma para la monitorización, evaluación y diseño de mejoras en el Sistema Nacional de Transportes en favor de una movilidad sostenible. Disponer de información acerca de la movilidad de las personas ha sido siempre de vital importancia para la planificación, seguimiento y gestión de sus infraestructuras y servicios de transporte. Tradicionalmente, para conocer los hábitos de viaje de las personas a nivel nacional, Mitma realizaba las **encuestas** conocidas como "Movilia", las cuales suponían un importante esfuerzo económico, técnico y hu-



mano. La última encuesta realizada por Mitma se hizo en 2007.

En 2018 Mitma se embarcó en este viaje de **innovación de la mano del big data**, buscando nuevas formas de obtener información de la demanda de transporte a nivel nacional, con el objetivo de superar las limitaciones que las metodologías tradicionales planteaban.

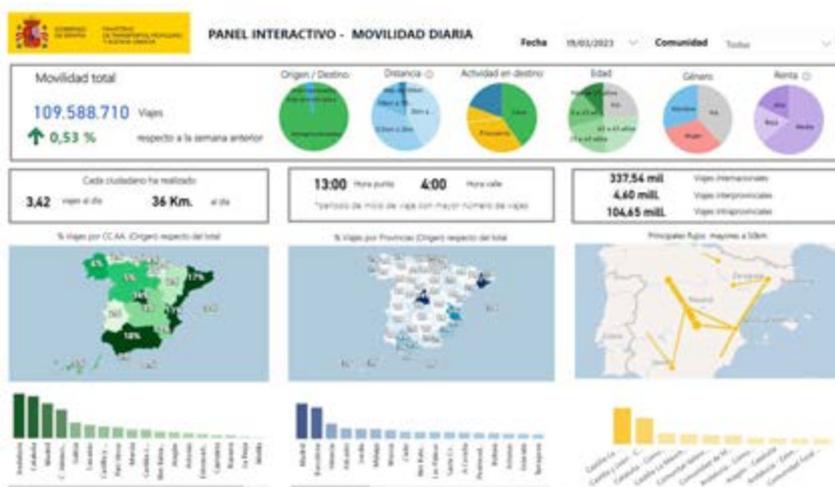
La llegada del big data

La fuente de datos seleccionada para llevar a cabo este primer estudio de movilidad con tecnologías **big data** fue la información procedente de **teléfonos móviles**, la tecnología que por aquel entonces (y también actualmente) ofrecía el mayor potencial para caracterizar la movilidad de la población a gran escala. El estudio tomó el nombre de **“Estudio piloto de movilidad interprovincial”** y en él se analizaron dos meses tipo, uno de invierno y otro de verano, generando matrices origen-destino de viajes y etapas a nivel provincial para aquellos viajes superiores a 50 km. La técnica utilizada permitía identificar los modos de transporte, diferenciando entre carretera, aéreo, ferroviario y marítimo.

Los resultados de aquel estudio alimentaron el primer **Modelo Nacional de Transportes** y sirvieron como primera toma de contacto con esta nueva tecnología que abría paso a un mundo desconocido para nosotros.

Herramienta contra la COVID-19

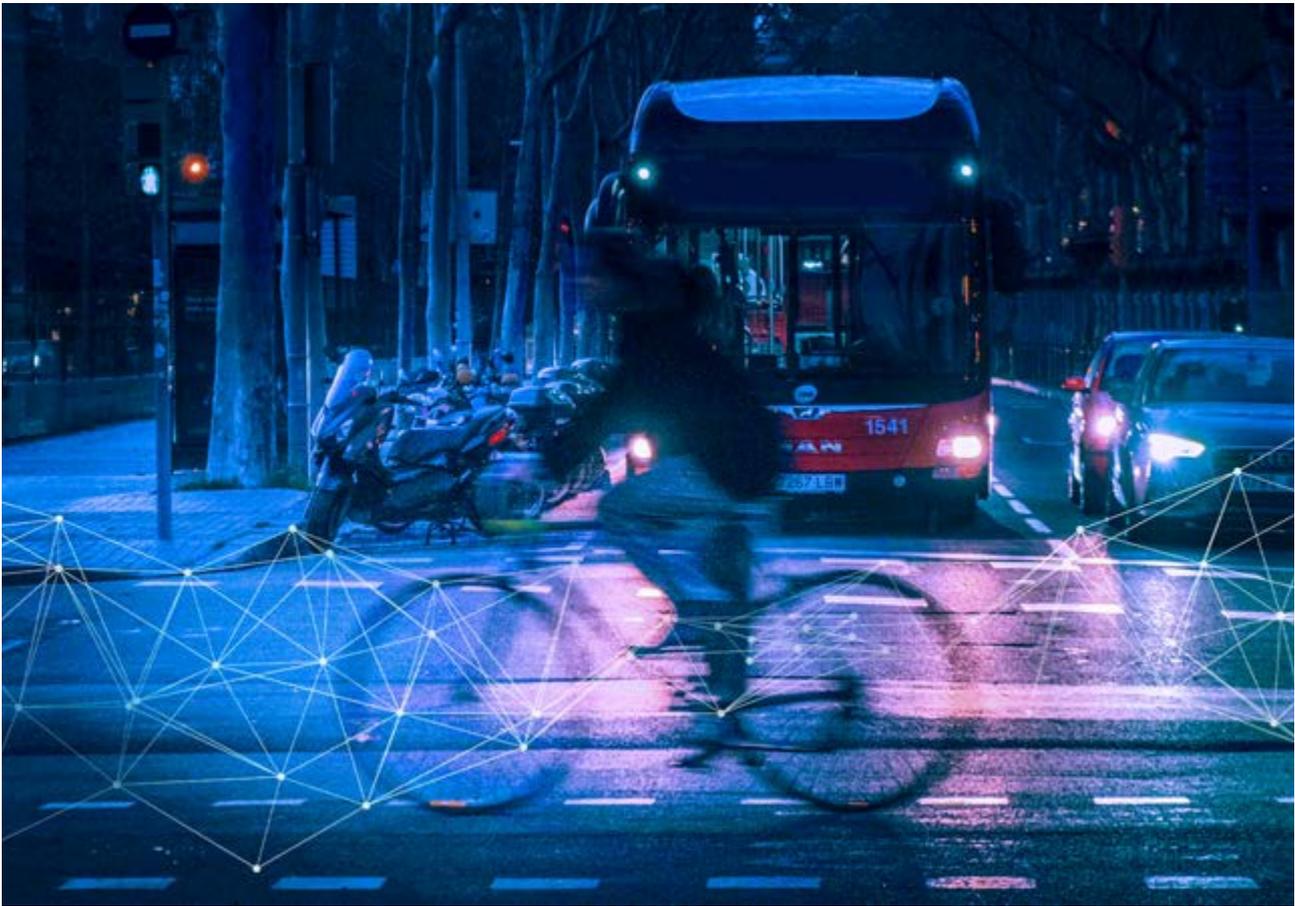
En 2020 llegó la pandemia. El virus COVID-19 sacudió los hogares de todo el mundo e hizo que las autoridades tuvieran que tomar medidas de restricción de la movilidad con



Disponer de información acerca de la movilidad de las personas ha sido siempre de vital importancia para la planificación, seguimiento y gestión de las infraestructuras y servicios de transporte.

el único objetivo de **luchar contra la pandemia** que amenazaba las vidas de los ciudadanos. En ese contexto mundial de emergencia sanitaria, en España se declaró el

Estado de Alarma el 14 de marzo de 2020. Pocos días después, el Gobierno encargó a Mitma **medir la movilidad diaria** en el territorio nacional de la forma más rápida y



precisa posible, así que, basándose en la experiencia previa adquirida en 2018, se puso en marcha un nuevo estudio con tecnología *big data* a partir de los registros de telefonía móvil: *Estudio de movilidad con big data durante la pandemia*.

La maquinaria se puso en marcha enseguida. Se creó un **grupo de trabajo** formado por seis expertos en tecnologías geoespaciales, movilidad e informática de Mitma y por varios expertos en *big data* e inteligencia artificial de la empresa tecnológica Nommon, encargada del estudio de 2018 y seleccionada para contribuir a este nuevo reto. Al igual que en 2018, se emplearon datos procedentes del operador Orange España y se aplicaron técnicas de elevación muestral para estimar la movilidad del total poblacional.

El objetivo principal fue la actualización de información de movilidad a nivel nacional de forma diaria (con un decalaje de tres días para

la recogida del dato y el cálculo de los indicadores), algo nunca acontecido y que suponía un **gran reto** y un importante **avance tecnológico**.

Se trabajó de forma intensa y se **tomaron decisiones** conjuntas con el objetivo de que se pudieran publicar los datos en abierto, **proporcionando herramientas de visualización** que permitieran facilitar la explotación e interpretación de los resultados. Se facilitó información de movilidad por **hora del día y a nivel de comunidad autónoma, provincia y municipio**, llegando en algunos casos a niveles de distritos censales, siempre cumpliendo la legislación en materia de protección de datos personales.

Los resultados del estudio fueron utilizados como soporte a la toma de decisiones por parte de las autoridades y para evaluar las difíciles medidas de restricción de la movilidad que se iban adoptando. Este estudio de movilidad fue

de vital importancia para alimentar los modelos epidemiológicos y los diferentes estudios e investigaciones que organizaciones como el Instituto Carlos III o el Barcelona Supercomputing Center llevaron y siguen llevando a cabo, con objeto de luchar contra la COVID-19. Afortunadamente, el virus fue perdiendo terreno, el Estado de Alarma finalizó el 9 de mayo de 2021 y con él el segundo estudio de movilidad.

Un nuevo estudio mucho más preciso que los anteriores

A partir de 2020 Mitma no se planteó dejar de medir la movilidad haciendo uso de la tecnología *big data*; esta nueva forma de analizar viajes había venido para quedarse. Así pues, tras este forzoso pero valioso período de aprendizaje, en mayo de 2022 Mitma comenzó un nuevo estudio, *Estudio de movilidad de viajeros de ámbito nacional*

En los nuevos estudios, no solo se ha ampliado el alcance, sino que también se ha evolucionado la tecnología empleada.

aplicando la tecnología big data, mucho más preciso y completo que los anteriores, con el objetivo inicial de dar continuidad a esta tipología de estudio hasta 2025.

El proyecto actual presenta un estudio enmarcado en tres grandes ámbitos: estudios básicos, estudios completos y estudio de rutas por carretera. Los estudios básicos pretenden dar continuidad a la monitorización continua de la movilidad llevada a cabo durante la COVID-19, proporcionando información similar pero con mayor resolución espacial y con nuevos indicadores, como información sociodemográfica de los viajeros o viajes fuera del territorio nacional. Los estudios completos y el estudio de rutas tienen como objetivo dar continuidad a los estudios realizados en 2018. El objetivo principal de estos dos estudios es analizar la movilidad (viajes, etapas y tours) por modo de transporte y proporcionando información de ruta. Al igual que los estudios básicos, estos estudios presentan mejoras con respecto a su proyecto antecesor, ampliando, entre otros aspectos, su resolución espacial y su alcance temporal. Los productos generados dentro del proyecto serían los siguientes:

- Estudios básicos
 - Matrices de viajes básicas (sin modo de transporte)
 - Matrices de viajes por persona
 - Matrices de pernoctaciones
- Estudios completos
 - Matrices de viajes completas (con modo de transporte)

- Matrices de etapas
- Matrices de tours
- Matrices del área de influencia de terminales de transporte
- Matrices de la movilidad transfronteriza
- Matrices de frecuencia de viaje
- Rutas de carretera
 - Matriz de rutas por carretera

Para más información del alcance del proyecto actual, puede consultarse la siguiente página web:

<https://www.mitma.gob.es/ministerio/proyectos-singulares/estudio-de-movilidad-con-big-data>

En estos nuevos estudios no solo se ha ampliado el alcance, sino que también se ha evolucionado en relación con la tecnología empleada. En proyectos anteriores se empleaban principalmente Call Detail Records (CDRs) para caracterizar la movilidad. Estos datos recogen información activa del dispositivo móvil (llamadas, SMS, sesiones de datos, etc.), proporcionando información sobre el instante preciso de interacción con la red y la ubicación del dispositivo a nivel de antena de telefonía móvil. En el proyecto actual se están empleando, además, datos procedentes de sondas de red (eventos pasivos), que generan un registro cada vez que el dispositivo móvil se conecta a una nueva antena, siendo el volumen de datos empleado en este

proyecto cuatro veces superior al empleado en el proyecto anterior, alcanzando los 6000 millones de registros al día. La mejora del dato permite aumentar la cantidad de muestra útil (aquella muestra con un número de registros suficiente para su análisis) y caracterizar mejor la cadena de actividades y viajes de la población.

Profundizando en el estudio de la movilidad

El análisis de la movilidad que se está haciendo desde mayo de 2022, por un lado, tiene mayor alcance que el estudio realizado durante la pandemia de la COVID-19 y, por otro lado, es mucho más completo. Respecto al alcance, durante la pandemia solo se estudiaron los viajes realizados por los residentes en España en territorio nacional, ahora, en cambio, se analizan los viajes tanto de residentes nacionales como de extranjeros y no solo se tienen en cuenta viajes dentro de España, sino también cualquier viaje internacional con origen o destino en España.

Además, resulta más completo porque incluye productos nuevos, como por ejemplo, el estudio de las rutas por carretera o el estudio de movilidad obligada o recurrente.

¿Cómo se procesan los registros de telefonía móvil?

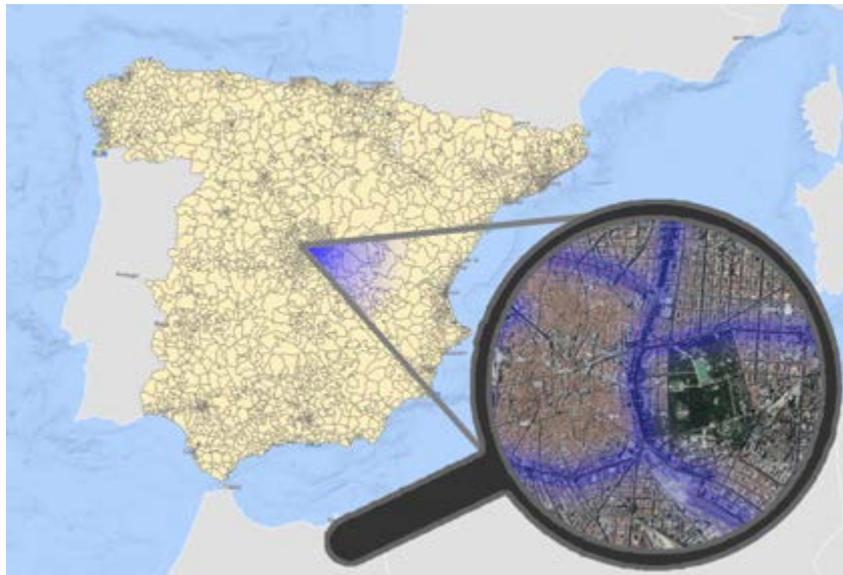
El objetivo principal del proceso de análisis de datos de telefonía móvil consiste en la generación de los denominados diarios de actividades y viajes de la población que, como su nombre indica, reflejan las distintas actividades y viajes realizados por la población a lo largo del día. Para detectar las actividades se emplean técnicas de clusterización espacio-temporal que permiten detectar la estancia de un dispositivo

anónimo en un lugar determinado durante un cierto tiempo. A partir de la **detección de estancias**, se emplean distintos algoritmos basados en la duración de la estancia, ubicación, frecuencia de aparición, relación con estancias anteriores y posteriores, etc., para caracterizar si la estancia hace referencia a una actividad o a una parada intermedia asociada a un viaje más largo (por ejemplo, una parada en un aeropuerto antes de embarcarse en un avión). A partir de esta clasificación, es posible identificar posteriormente los viajes como **desplazamientos entre actividades**.

La información contenida en el diario de actividades y viajes se complementa a través de distintos procesos de análisis de los datos, entre los que destacan: la **caracterización socio-demográfica** de los usuarios empleando como fuente de datos principal los datos de cartera de cliente del operador; la determinación del **modo de transporte** a partir de técnicas de **map-matching**; y la caracterización de las actividades y los viajes a través del **análisis longitudinal** de los datos (por ejemplo, para la detección del lugar de residencia o para la detección de viajes realizados por conductores profesionales).

Asegurando y midiendo la calidad

A pesar de tener una fuente de datos más precisa y de utilizar una metodología suficientemente depurada y testeada, es necesario, como en cualquier proceso de análisis, evaluar sistemáticamente la calidad de los resultados para depurar la metodología y **detectar de forma temprana posibles anomalías** en los datos de origen de los registros de telefonía. Además, este proceso es especialmente relevante porque ofrece un servicio



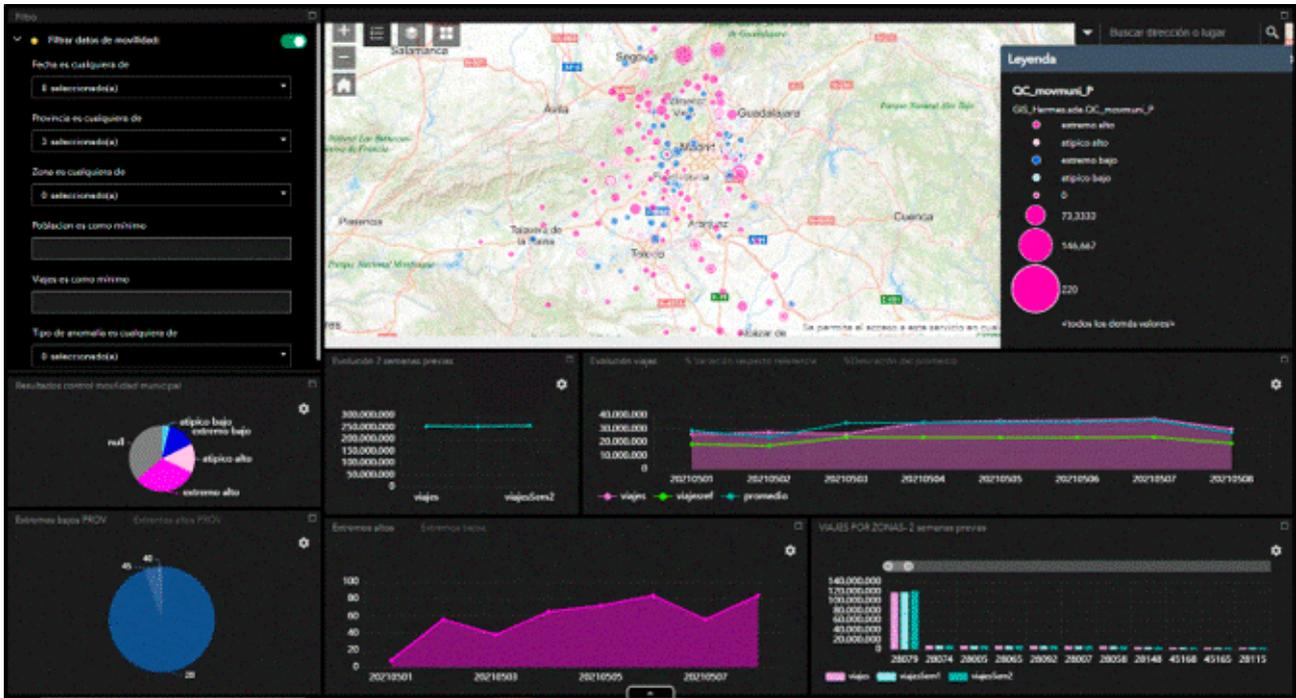
La tarea de validación de los resultados ha ido avanzando con cada proyecto, añadiendo nuevos controles, ajustando parámetros y desarrollando un ojo cada vez más crítico para identificar la procedencia de las posibles anomalías en los datos.

público y produce unos datos que son la base de numerosos estudios y modelos, por lo que se debe garantizar la calidad requerida y conocer el grado de exactitud de los datos realizando las medidas comparativas oportunas.

En 2018, cuando se empezó a realizar este tipo de análisis *big data*, no se sospechaban los exhaustivos controles de calidad que se llegarían a desarrollar años después gracias al conocimiento adquirido cada día. En esta tarea de validación de los resultados se ha ido avanzando con el proyecto, aprendiendo con cada tropiezo, añadiendo nuevos controles y

ajustando parámetros. Se ha desarrollado un ojo cada vez más crítico para identificar la procedencia de las anomalías detectadas, a veces justificadamente, por algún festivo local o eventos que movilizan a la población, y a veces por retrasos en los datos fuente o diferentes casuísticas que no hubieran sido contempladas de antemano en la metodología.

En el primer estudio, así como en la primera fase del estudio que se realizó durante la pandemia, debido a la premura y urgencia en la obtención y publicación de resultados, se revisó **visualmente mediante gráficos** todas las matrices a diferentes



El estudio que realiza actualmente Mitma permite diferenciar el modo de transporte empleado en los desplazamientos, así como la actividad en destino o la residencia del viajero.

niveles de agregación, lo cual era una tarea tediosa que requería algunas horas cada día. Poco a poco se fueron introduciendo controles en los datos crudos que proporcionaba la operadora de telefonía (a veces no llegaban a tiempo y/o se proporcionaban incompletos) y en los volúmenes totales de viajes, pero cuando todo parecía ir bien, y la metodología estaba suficientemente probada, aparecían eventos que precisaban revisar los datos y seguir mejorando los controles. Por ejemplo, un día, al inicio de la fase de desescalada, se descubrió que el número de viajes entre islas superaba con creces los datos facilitados por las navieras en ese

momento. En este caso se aprendió que no todos los desplazamientos entre antenas (compatibles espacio-temporalmente con viajes plausibles) pueden ser interpretados como viajes reales y que es necesario introducir **constricciones adicionales** (por ejemplo, criterios basados en la repetición de los viajes en el día) y/o un contraste con datos de oferta o demanda, procedente de los operadores de servicios de transporte, que permita calibrar los modelos para diferenciar entre un viaje a la isla vecina y una conexión de un teléfono móvil a la isla vecina sin desplazamiento del dispositivo. Este es solo un ejemplo de los varios retos

que se han ido encontrando en el camino, mientras se implementaban los controles de calidad diarios que nos han permitido aprender, mejorar todo el proceso y depurar los algoritmos.

Controles automáticos con ETL y visualizadores GIS

En verano de 2020, mientras se generaban las matrices origen-destino diarias y varias personas se dedicaban diariamente a la revisión, se comenzó a desarrollar un control automático mediante herramientas ETL (Extract Transform and Download). Los resultados de este control automático se redirigen dinámicamente a un **visualizador GIS web** que permite descubrir patrones anómalos agrupados espacialmente, que, de otra forma, sin esta **visión integral geoespacial**, serían imposibles apreciar.

Estos controles automáticos se basan en la comparación de los viajes de cada día respecto a los viajes del mismo día en las **semanas previas** y respecto a los viajes de una **semana de referencia** elegida a conciencia, además de comprobar aspectos más obvios

02/03/2023 23/03/2023

Provincia origen

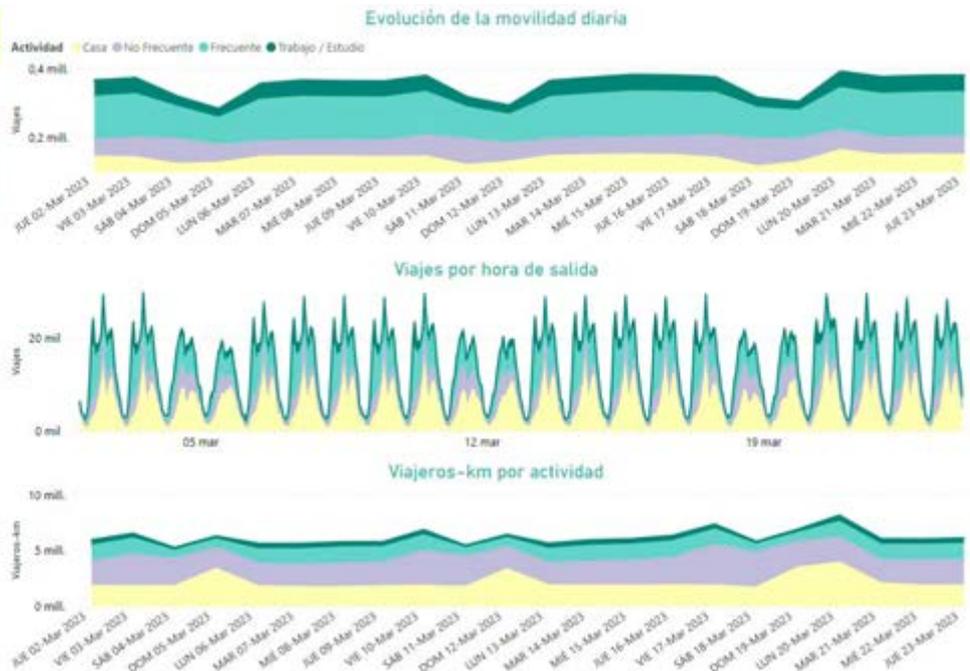
Cuenca

Provincia destino

Todas

Actividad

- Casa
- No Frecuente
- Frecuente
- Trabajo / Estudio



como el volumen diario de viajes a nivel nacional, provincial y local. La idea es transformar, en fórmulas matemáticas basadas en operadores estadísticos, el criterio que impondría un experto en movilidad al revisar los datos diarios para discernir entre lo que pudiera ser un valor normal y lo que, presumiblemente, pudieran ser valores atípicos, extremos o excepcionales.

El objetivo es que salten las alarmas si se detecta un número de viajes anómalo en algún par origen-destino o en alguna zona. No obstante, que salten las alarmas no significa que sea un error en la fuente de datos o en la metodología, sino que se deben revisar las causas que han producido esa anomalía para discriminar los errores de lo que no lo son (por ejemplo, podría darse el caso de estar ante fechas festivas, partidos de fútbol, conciertos, ferias o congresos importantes). Diariamente se van detectando dichas anomalías y se van justificando, o corrigiendo la incidencia si es el caso, para garantizar la calidad de los resultados obtenidos.

La movilidad de estos últimos años es cada vez menos predecible. Nada tiene que ver de un mes al siguiente si ocurre algún evento relevante para los hábitos de movilidad de las personas.

Ajustando los parámetros del control de calidad

¿Es normal que un martes, en una zona determinada, haya un 16 % más de viajes que el martes anterior? Establecer los umbrales y el ajuste de los parámetros del control de calidad no es una tarea sencilla y ha requerido de muchos meses de entrenamiento y aprendizaje. Se trata de un ajuste variable, ya que los parámetros que eran lógicos, por ejemplo, en época de pandemia, no resultan válidos a día de hoy. La movilidad de estos últimos años es cada vez **menos predecible** y poco o nada tiene que

ver un mes al siguiente si ocurre algún evento relevante para los hábitos de movilidad de las personas. Este aprendizaje nos ha hecho ser conscientes, por una parte, de que hay que seguir **ajustando estos parámetros** de control de calidad en base a las circunstancias de cada momento y, por otra, de la gran relevancia que tiene el uso de estas tecnologías que nos permiten seguir el ritmo a estos cambios acelerados de los patrones de movilidad. Finalmente, el poder ejecutar estos controles automáticos de forma diaria y liberarnos, en gran parte, de las revisiones visuales ha supuesto un **ahorro considerable en el tiem-**

po de revisión y una objetividad que, sin duda, hace que se pueda garantizar una calidad homogénea en los resultados.

Utilidad para el Ministerio

Mitma utiliza y utilizará la información de movilidad generada para alcanzar multitud de objetivos. El primer objetivo es, sin duda, el propio conocimiento de la demanda que tiene en cada momento el **Sistema Nacional de Transportes (SNT)**. Cuantificar la movilidad y detectar patrones, asimetrías estacionales, cuota modal, diferencias entre territorios, etc. es la base sobre la que trabajar para mejorar este SNT hacia una movilidad más sostenible y equitativa en todo el territorio nacional.

Otro objetivo fundamental, derivado del anterior, es el análisis del impacto sobre el SNT que tendrían las diferentes alternativas de mejoras que se están planteando sobre las infraestructuras y los servicios de transporte, mediante la **simulación de diferentes escenarios futuros** y la predicción de la demanda y del comportamiento del SNT en su conjunto. Esto es posible gracias a la inclusión de estas matrices origen-destino obtenidas con tecnologías *big data* en el **Modelo Nacional de Transporte**, una herramienta fundamental para el apoyo a la toma de decisiones en el ámbito de la planificación.

Aparte de estas dos utilidades, Mitma aprovecha todo el potencial de estos datos de movilidad para, entre otras cosas, la detección temprana de necesidades de ampliación de la capacidad de las carreteras del Estado, la gestión del tráfico en determinados tramos de la red, la priorización de las actuaciones, la conservación de la red de carreteras y la red de vértices

El uso de estas tecnologías nos permite seguir el ritmo a los cambios acelerados de los patrones de movilidad.

geodésicos, o la identificación del déficit de servicios de movilidad.

Utilidad para la sociedad

Conocer y caracterizar la movilidad con este grado de detalle es de gran utilidad tanto para Mitma como para el resto de administraciones públicas, así como para empresas, instituciones, universidades y otros usuarios, que ya están empleando los datos de movilidad (descargables de forma gratuita desde la web de Mitma) en multitud de proyectos.

Además, el uso de estas tecnologías facilita el diseño de iniciativas públicas basadas en datos, lo que se traduce en servicios adaptados a las necesidades reales de los ciudadanos, con medidas más eficientes respecto al uso de recursos públicos y más sostenibles, ya que permiten analizar mejor las distintas realidades sociales y los cambios que éstas sufren.

Qué nos depara el futuro

El futuro, desde luego, irá de la mano de la fusión de diversas fuentes de datos para mejorar la caracterización de la movilidad. Es probable que en los próximos años aparezcan nuevas fuentes de datos que puedan sustituir y/o complementar a las fuentes de datos con las que se trabaja actualmente, por lo que es importante estar siempre pendiente de las nuevas tendencias. Por ejemplo, los datos procedentes de **apps móviles** podrían ser una alterna-

tiva en el futuro si se mejora su tamaño muestral o su granularidad temporal. Por otro lado, los datos de **coches conectados y autónomos** pueden ser una fuente de datos muy relevante para analizar la movilidad por carretera. La aparición de nuevas fuentes de datos también supone un reto desde el punto de vista metodológico, que hará necesario repensar la aproximación empleada para obtener información de movilidad.

Muchas personas podrían plantearse si dejarán de ser necesarias fuentes clásicas, como **las encuestas** o los datos de demanda de viajeros obtenidos a partir de los datos de **billetaje**. Es posible que, durante los próximos años, se sigan utilizando e incluso continúen siendo una fuente complementaria y necesaria para el ajuste y calibración de los modelos de demanda empleados en los estudios *big data*. Seguramente se mejore la forma en la que esta información es recogida para aprovechar las sinergias con las nuevas fuentes de datos procedentes del *big data*, permitiendo hacer un uso más eficiente de los recursos (por ejemplo, realizando encuestas más reducidas para recoger la información que no puede extraerse de manera sencilla a través de otras fuentes de datos). También se espera que la compartición de datos de movilidad permita evolucionar los modelos existentes y que se puedan seguir mejorando los análisis, integrando las mejoras que las nuevas tecnologías nos depararán en el futuro. ■