

# Modelos de demanda potencial de viajeros en redes de transporte público: aplicaciones en el Metro de Madrid. <sup>1</sup>

## Models of potential demand of passengers on public transport networks: applications in the Metro de Madrid.

*Javier Gutiérrez Puebla<sup>2\*</sup>, Osvaldo Daniel Cardozo<sup>3\*\*</sup>, Juan Carlos García Palomares<sup>\*4</sup>*

\* Departamento de Geografía Humana de la Universidad Complutense de Madrid (UCM)

\*\* Instituto de Geografía de la Universidad Nacional del Nordeste (UNNE)

### Resumen

Este trabajo propone una metodología basada en Sistemas de Información Geográfica para estimar la demanda de viajes en estaciones de redes de transporte público, tomando como ejemplo la red de metro de Madrid. Primero se emplea una serie de datos descriptivos para caracterizar la red, clasificar las estaciones y obtener una tipología de las mismas. Luego, con el objetivo de explicar y predecir los viajes (entradas a la red) se generan dos modelos: uno sencillo a partir de las tasas de penetración de uso del metro en función de la distancia (*distance decay*), y otro más complejo basado en un modelo de regresión lineal múltiple (MRLM) que incorpora variables relativas a la estación y su entorno (densidad, mezcla de usos, diseño urbano, presencia de modos competidores). Su aplicación muestra resultados alentadores, y se plantea como una alternativa a los clásicos modelos de cuatro etapas, más complejos y con un mayor coste económico.

**Palabras Claves:** Modelos de Demanda / Red de Metro / Funciones de Uso / Regresión Lineal Múltiple / Madrid.

### Abstract

This paper proposes a methodology based on geographic information systems for estimating the demand for travel stations in public transport networks, such as taking the subway in Madrid. First employs a number of descriptive data to characterize the network, sorting stations and obtain a typology of the same. Then, in order to explain and predict travel (tickets to the network) are generated two models: one simple from the penetration rates of use subway depending on the distance (*distance decay*), and other more complex based in a multiple linear regression model (MRLM) that incorporates variables related to the station and its surroundings (density, mixed use, urban design, presence of competitors modes). Its implementation shows encouraging results, and arose as an alternative to traditional models of four stages, more complex and more cost effective manner.

**Keywords:** Demand Model / Metro Network / Functions of Use / Multiple Linear Regression / Madrid.

<sup>1</sup> Trabajo presentado en el "III Seminario Internacional de Ordenamiento Territorial – La Interdisciplina en el Ordenamiento Territorial". Encuentro internacional interdisciplinario, organizado por el Instituto CIFOT, Facultad de Filosofía y Letras, Universidad Nacional de Cuyo. Setiembre de 2005. Mendoza. Argentina.

Fecha de aprobación: 22/07/2008 – Fecha de publicación: 31/07/2008.

<sup>2</sup> Doctor en Geografía. Profesor Titular del Departamento de Geografía Humana. Experto en Sistemas de Información Geográfica y Geografía del Transporte.

<sup>3</sup> Becario del Plan de Formación de Personal Investigador de la Comunidad de Madrid. Realiza su tesis doctoral sobre transporte en el área metropolitana de Madrid, utilizando Sistemas de Información Geográfica.

<sup>4</sup> Profesor de Geografía. Departamento e Instituto de Geografía. Universidad del Nordeste.

## Modelos de demanda potencial de viajeros en redes de transporte público: aplicaciones en el Metro de Madrid

### 1. Introducción

La manera en que se distribuye la red de un sistema de transporte sobre el territorio, pone en evidencia su carácter estructurante y organizador, particularmente en las áreas metropolitanas donde determina en gran medida el modelo funcional que permitirá las interacciones, tanto en el interior como en su espacio adyacente. Por ello, la planificación del transporte requiere una visión global del sistema, para de esta manera conocer de la mejor forma, su funcionamiento interno y la interacción con el territorio sobre el que se emplaza.

En esa labor, una de las tareas más relevantes consiste en determinar la *demanda* del servicio. En ese sentido, la accesibilidad a las redes de transporte público, entendida como la facilidad de alcanzar una estación o parada desde el lugar de origen del viaje, es una cuestión a la que cada vez se da más relevancia. Desde una perspectiva espacial, la accesibilidad normalmente es percibida en términos de proximidad. La importancia de la facilidad de acceso al transporte público es obvia: cuanta más población resida, trabaje o realice actividades cerca de las estaciones o paradas de transporte público mayor es la probabilidad de que lo utilicen (MURRAY ET AL., 1998). De hecho, se reconoce que la accesibilidad a las redes de transporte público tiene una importancia similar a la velocidad y la puntualidad de los servicios desde el punto de vista de la captación de viajeros. Evidentemente este tema es relevante para planificadores y operadores, pero también lo es desde la perspectiva ambiental (KEIJER AND RIETVELD, 2000). Así la accesibilidad al transporte público es una cuestión crítica tanto desde el punto de vista de la movilidad sostenible (BLACK, 1996) como de la accesibilidad sostenible (WEBER, 2006).

Sin embargo en la planificación del transporte público se ha prestado mucha más atención al viaje en transporte público que al recorrido hasta las paradas y estaciones. Se ha dedicado mucho esfuerzo a construir complejos modelos de demanda de cuatro etapas<sup>5</sup> (que asumen una perspectiva "regional" de la movilidad), pero poco a indagar sobre la movilidad desde los orígenes de los viajes hasta paradas y estaciones (lo cual significa adoptar una perspectiva más local, analizando con detalle el entorno de paradas

<sup>5</sup> *Modelos secuenciales*, denominados así por sus etapas sucesivas y que consisten en: I) Generación de viajes. II) Distribución zonal. III) Elección modal. IV) Asignación de ruta.

y estaciones).

Dada la connotación espacial en la composición de la demanda, los Sistemas de Información Geográfica (SIG) constituyen una herramienta esencial para analizar la movilidad hasta las estaciones y los factores que influyen en la misma. Así, los SIG han sido utilizados frecuentemente en el análisis de la accesibilidad a las redes de transporte y en su representación cartográfica (GUTIÉRREZ Y GARCÍA, en prensa; MORENO Y PRIETO, 2003; KIMPEL Y OTROS, 2006). La mayor parte de los usuarios del Metro acceden a las estaciones a pie. Está demostrado que la distancia a pie a las estaciones es un factor crítico en la elección de este modo. Apoyándose en SIG se han realizado análisis clásicos de cobertura, que utilizan radios distintos (300, 600, 900 metros) para reflejar el gradiente de la distancia. Sin embargo, poco se ha hecho por contrastar la adecuación de estos radios a la movilidad real de la población en su acceso a las estaciones.

En este marco, nuestra comunicación intenta analizar la relación existente entre la distancia a estaciones y el número de viajeros entrados a pie en la red de transporte para, a partir de esa relación, y en función de un conjunto de variables relativas a la estación y su entorno, proponer un modelo capaz de estimar la demanda de la red. Se parte de una pequeña aproximación a los modelos de predicción de la demanda a nivel de estación y las variables relevantes utilizadas en otros trabajos, para presentar más tarde un modelo, basado en herramientas SIG, que permita predecir el número de viajeros entrados en futuras estaciones. Se propone un modelo que profundiza en el análisis local y resulta de fácil aplicación. La propuesta se ha realizado en la ciudad de Madrid y con el Metro como modo de transporte, tratando de estimar las entradas en cada una de las estaciones de la red.

## **2. Modelos de predicción a nivel de estación y variables relevantes**

En la literatura existente sobre modelos de predicción de la demanda del transporte público, queda en evidencia que la mayoría se basan en el clásico modelo de cuatro etapas, surgido en los años 50'. En teoría, la robustez de estos modelos les convierte en la mejor herramienta para análisis y predicción de la demanda, pero en la práctica, muestran varios inconvenientes (MARSHALL Y GARDY, 2006; FLYVBJERG-HOLM-BUHL, 2006):

- ✓ *Precisión del modelo:* los resultados fiables sólo se logran a nivel regional siendo poco precisos a nivel de corredor, ruta, segmento o estación, debido a que basan sus análisis en unidades espaciales como zonas de transporte o distritos censales, con tamaños variables y no siempre de composición homogénea. FLYVBJERG, HOLM Y BUHL (2006) han estudiado las desviaciones entre las predicciones de viajes y los valores reales en una amplia muestra de proyectos de transporte, encontrando una sobreestimación media del 106% en proyectos de ferrocarriles y un error de  $\pm 20\%$  en proyectos de carreteras y por ello demandan una mayor transparencia así como nuevos métodos de predicción. Aumentar su precisión supone un gran esfuerzo en términos de tiempo, costes, y la necesidad de modificar cada vez el modelo (ver KAPLAN Y OTROS, 2003).
- ✓ *Sensibilidad al uso del suelo:* existe un consenso generalizado entre investigadores y planificadores de que las 3Ds (densidad, diversidad y diseño) son fundamentales para explicar el acceso al transporte público (CERVERO Y KOCHELMAN, 1997). Sin embargo estos modelos son poco sensibles a detectar la influencia del diseño urbano y los usos del suelo.
- ✓ *Barreras institucionales y coste excesivo:* los modelos más sofisticados son costosos y requieren de personal con alta preparación técnica.
- ✓ *Sesgo conceptual:* desde la concepción de su idea original, están orientados a la eficiencia de todo un sistema de transporte, y en especial el desplazamiento en vehículo particular. Recordemos el modelo que fue pensado inicialmente en los años 50' para EE.UU, donde las facilidades de compra de un coche y las autopistas comenzaban su explosión.

A los problemas mencionados, en nuestra opinión deben añadirse al menos otro, relacionado con los anteriores:

- ✓ La incapacidad que muestran los actuales modelos para tratar, con una precisión satisfactoria, el efecto del gradiente de la distancia en la variación del volumen de la demanda. Generalmente, en los modelos de cuatro etapas se asume que todos los viajes de una zona parten desde su centro (centroide), cuando está comprobado que existe una fuerte caída en el número de viajes con el aumento de la distancia a las estaciones o paradas de transporte público. Esta cuestión, conocida como el gradiente de la distancia (*distance decay*) en los modelos de demanda, representa el efecto disuasivo de la distancia sobre los desplazamientos, de manera que se forman gradientes concéntricos de intensidad

decreciente.

A la vista de estos problemas, en la bibliografía reciente se ha planteado la necesidad de utilizar modelos más simples y de respuesta inmediata, y que sin embargo puedan recoger con mayor precisión el entorno de las estaciones. WALTERS Y CERVERO (2003) proponen un modelo directo de estimación de la demanda a nivel de estación. Se trata de un modelo de regresión múltiple, que considera de forma detallada el entorno de las estaciones. En realidad el uso de los modelos de regresión múltiple en la previsión de la demanda de transporte público ha sido muy frecuente, si bien las escalas de trabajo han sido muy diferentes. Algunos utilizan como unidades de análisis distintas ciudades o distintos barrios de la misma ciudad; otros intentan realizar predicciones a nivel de zona de transporte; otros para líneas o tramos de líneas; pero muy pocos ofrecen resultados a nivel de parada o estación.

Estos modelos de regresión múltiple incluyen una multiplicidad de variables independientes, unas externas (sociodemográficas, económicas, modos complementarios o competidores, disponibilidad de aparcamiento en la estación, espaciales, etc.) y otras internas (oferta de servicio, tarifas, frecuencia de servicio, puntualidad, etc.) (ver TAYLOR Y FINK, 2003; CERVERO, 2001; KIMPEL, 2001; CHU, 2004). En general las variables externas tienen más poder explicativo que las internas y dentro de éstas el nivel de servicio es más relevante que las tarifas (TAYLOR Y FINK, 2003). Esta división entre variables internas y externas tiene interés no sólo desde la perspectiva de generar modelos explicativos, sino también desde la perspectiva de las estrategias a seguir para aumentar el uso del transporte público. Así, la European Commission on Transportation Research habla de estrategias directas como las relativas a los factores internos (tarifas, calidad y cantidad del servicio, etc.) y de estrategias externas basadas en factores externos como la planificación del uso del suelo, las restricciones al uso del automóvil, etc...

Recientemente se ha prestado un interés creciente a las variables de tipo espacial. Se destaca la importancia de variables relativas al uso del suelo en la demanda del transporte público. Básicamente se trata de las tres Ds: densidad, diversidad y diseño. La relevancia de las variables del entorno en el uso del transporte público ha llevado a planificadores norteamericanos a promover "transit oriented development" (TOD), desarrollos nuevos o transformaciones en desarrollos existentes, situados en el entorno de las estaciones que deben frenar el sprawl y la dependencia del coche (NELSON Y

NILES, 1999; CERVERO, 2004). Se trata de desarrollos compactos, con la mezcla de usos y transitabilidad peatonal, que favorecen el uso del transporte público. Estudios llevados a cabo en TODs situados en el entorno de las estaciones de ferrocarril en California demostraron que la densidad es el principal predictor del uso del ferrocarril. Una vez que existen altas densidades, la transitabilidad peatonal y la mezcla de usos adquieren una importancia menor (CERVERO, 2004), aunque fuera de los TODs sí se ha demostrado la importancia de la mezcla de usos.

La importancia de la distancia en el uso del transporte público ha quedado demostrada en multitud de trabajos. Sin embargo muchos de los modelos de regresión no tratan la distancia de forma adecuada. Sólo lo hacen en tanto que consideran la población o el empleo, por ejemplo, en un determinado radio de cobertura. Pero de esta forma no pueden recoger la mayor propensión a utilizar el transporte público cuando la distancia es menor dentro del radio de cobertura elegido. Sólo algunos modelos incluyen de alguna forma la distancia, normalmente porque incorporan como variables independientes no la población o el empleo en toda el área de cobertura, sino en distintas franjas de cobertura (WALTERS Y CERVERO, 2003). Cuando se ha utilizado de una forma más precisa la distancia se ha tratado no ya de modelos de regresión sino de funciones de disminución de la población servida con la distancia, lo que se conoce como *distance-decay*.

La forma más utilizada de integrar el *distance-decay* en la estimación de la demanda es mediante el empleo de *curvas de penetración* (LEVINSON Y BROWN-WEST, 1987; ZHAO et al., 2003). En la propuesta de ZHAO et al. (2003), a partir de los datos de viajeros entrados en la red de autobuses según distancia a la parada, calculan una función de caída la población servida con la distancia. Parten de datos reales sobre el número de viajes entrados en la red a pie desde distintas distancias, y después para cada intervalo de distancia normalizan esos viajes por la población. La función obtenida indica la tasa con la que el uso del autobús decrece con la distancia. El modelo es muy simple y de fácil y rápida aplicación. Pero precisamente por su simplicidad resulta poco fiable, especialmente en espacios heterogéneos, donde hay cambios relevantes en algunas variables explicativas, como la densidad de población o de empleo, la transitabilidad peatonal o la competencia con otros transportes públicos.

### 3. Propuestas para la estimación de la demanda a la red de Metro de Madrid

Partiendo de la revisión anterior a los modelos de estimación de la demanda propuestos

en la literatura reciente, en este apartado se propone un modelo de regresión lineal múltiple, que trata de integrar la influencia de la caída de la demanda con la distancia, a partir del cálculo inicial de *curvas de penetración*. Las curvas son utilizadas en el modelo de regresión como un factor de ponderación de las variables que hacen referencia al entorno de las estaciones. El modelo propuesto se ha realizado para la estimación de las entradas en las estaciones de la red de Metro de la ciudad de Madrid, lo que hace necesario presentar inicialmente las características básicas de dicha red.

### **3.1 El Metro en el sistema de transporte público de Madrid: caracterización de la red y demanda de viajes**

En funcionamiento desde hace casi 90 años (se inauguró en 1919), la red de Metro de Madrid ha sufrido en las últimas décadas un intenso proceso de modernización y ampliación. En los últimos 8 años se han construido 111 km de líneas nuevas, con 63 estaciones, para contar en la actualidad (enero de 2007) con un total de 190 estaciones y una longitud de 215 km.

De las 190 estaciones, 152 estaciones (el 80%) son simples, por ellas pasa una sola línea, pero el resto de las estaciones son compartidas (de conexión). 28 estaciones (15%) poseen conexión con dos líneas, 9 estaciones (5%) con tres, y sólo una cuenta con una conexión cuádruple. Por otra parte, el Metro se conecta con otros modos de transporte, como los autobuses interurbanos (en 20 estaciones), los ferrocarriles de cercanías (21 estaciones), o el aeropuerto de Barajas (2 estaciones). Cabe señalar que 6 estaciones poseen conexión intermodal mixta, con las redes de autobuses interurbanos y cercanías.

La demanda de transporte público en Madrid muestra un crecimiento sostenido en los últimos años, con un aumento superior a los 100 millones de viajes entre 2001 y 2006. El Metro es el sistema de transporte público con mayor demanda. En 2006 alcanzó la cifra de 650 millones de viajes, que representan un 40.5% del total de los viajes en transporte público. Esta situación preponderante se afianzó en los últimos años, con un crecimiento del 19,7% entre 2001 y 2006, claramente por encima del correspondiente a cercanías, mientras que interurbanos mantiene sus cifras estables y EMT pierde viajeros. Estos datos refuerzan el interés por los estudios referidos a la demanda de Metro.

Si nos fijamos en las entradas al Metro que nos proporciona el aforo diario realizado en el

mes de noviembre de 2004 por el Consorcio Regional de Transportes<sup>6</sup>, la demanda real fue de casi 56 millones de viajes en toda la red, lo cual, entre las 190 estaciones en funcionamiento arroja una media de 294.000 viajes por estación en ese mes. Pero el número de viajeros entrados en cada una de las estaciones encierra fuertes disparidades. Así, por ejemplo, en la estación Manuela Malasaña de la línea 12, sólo entraron algo más de 123.000 pasajeros, frente a los 1.760.000 viajeros entrados en la estación Moncloa, lo que representa más de 143 veces el volumen de Manuela Malasaña. Si agrupamos las estaciones según intervalos de pasajeros entrados (tabla 1) vemos una marcada distribución asimétrica, con predominio de estaciones (el 83 %) debajo del límite de los 400.000 viajeros entrados, de las que 18 de ellas (9,4 %) no superaron los 100.000 viajeros. Sólo 6 estaciones (3 %) rebasaron el millón, todas ellas -excepto Sol-, son estaciones intermodales.

**Tabla 1: Pasajeros entrados según estaciones**

Entradas	Número de estaciones	%
menos de 100.000	18	9,4
100.000 – 200.001	60	31,3
200.000 – 300.001	57	29,7
300.000 – 400.001	25	13,0
400.000 – 500.001	9	4,7
500.000 – 600.001	5	2,6
600.000 – 700.001	7	3,6
700.000 – 800.001	2	1,0
800.000 – 900.001	0	0,0
901.000 – 1.000.001	3	1,6
1.000.000 – 1.430.000	6	3,1

Fuente: Aforo noviembre de 2004. Consorcio de Regional de Transportes

De estas entradas, los datos de la encuesta domiciliaria de movilidad de 2004 (EDM2004) muestran cómo entre el total de los desplazamientos en los que se usa el Metro en alguna de sus etapas la mayoría acceden andando a la estación, el 79,0% (tabla 2), lo cual confirma que la clave a la hora de estimar las entradas es modelizar adecuadamente los viajes que acceden peatonalmente a la red de Metro. Las entradas desde EMT,

<sup>6</sup> La mayor parte de las variables utilizadas para estimar la demanda en el modelo de regresión propuesto posteriormente parten de la Encuesta Domiciliaria de Movilidad, realizada por el Consorcio Regional de Transportes en noviembre de 2004, lo que justifica la utilización del aforo en esa fecha (demanda conocida).



interurbanos y cercanías se concentran en las estaciones que tienen el carácter de intercambiadores, pero mucho menos en las demás (terminales, pasantes, de conexión), en las que el acceso peatonal es generalizado.

**Tabla 2: Modo de acceso a Metro**

Modo	Viajeros	%
Andando	1,876,188	79.67
Ferrocarril	122,161	5.18
EMT	138,687	5.89
Interurbanos	146,570	6.22
Otros	71,470	3.03
TOTAL	2,355,076	100

Fuente: EDM2004

### 3.2 Fuentes de información y metodología en el SIG

Debido a las características de los datos necesarios, las fuentes de información resultan diversas. A continuación se detallan las principales y los atributos más relevantes de los datos obtenidos.

- ∨ *Redes de transporte público.* Proporcionada por el Consorcio en forma de cartografía (soporte digital de tipo vectorial). Se ha podido contar con las redes y paradas/estaciones de los cuatro sistemas de transporte público: Cercanías, EMT, Autobuses Interurbanos y Metro. En el caso del Metro se dispuso de información referida al número de accesos, sus coordenadas geográficas, número líneas, nombre de estación, longitud de vías, etc.
- ∨ *Aforo de accesos a la red de Metro.* El aforo fue realizado durante todo el mes noviembre 2004 por el Consorcio Regional de Transportes. Contiene datos de accesos diarios, por línea de Metro, por estación, el día y la suma total del mes.
- ∨ *Red viaria.* El callejero del Instituto de Estadística de la Comunidad de Madrid, dotado de plena conectividad, representa todo el viario de la ciudad. Esta capa es fundamental para calcular las distancias a estaciones (a través del viario y no en línea recta). También es necesaria para calcular índices de diseño urbano (densidad viaria) y poder así considerar este tipo de variables en los modelos de regresión múltiple.

- ∨ *Orígenes de los viajes en primera etapa en Metro.* De la EDM2004, se obtuvo una base de datos de unos 17.000 registros, con información relativa al origen de los viajes que utilizan Metro en primera etapa: coordenadas x e y, viajes expandidos y nombre de la estación donde accedieron al Metro, datos esenciales para el cálculo de las funciones de uso (figura 1).
- ∨ *Variables socioeconómicas.* Los datos correspondientes a las variables socioeconómicas del área de estudio fueron obtenidos del SIG de la movilidad del Consorcio de Transportes a nivel de zonas de transporte. Las mismas contienen datos del tipo de empleo, sector de actividad, número de ocupados, estudiantes, extranjeros, mujeres, población total y discriminada por grupos de edad, número de hogares y personas residentes, cantidad de vehículos por hogar. Estas variables tienen una influencia clara sobre la movilidad y serán utilizadas en los modelos de regresión múltiple.

La metodología propuesta trata de formular modelos de regresión lineal múltiple (*MRLM*), que permitan una respuesta inmediata, a partir de un reducido número de variables y de las que se pueda tener información futura (lo que asegure su utilización en las tareas de planificación). No obstante, entre las variables introducidas tienen que estar presentes tanto las características internas de la red de Metro como las externas, ya tengan que ver con la población o con el entorno urbano (*3Ds*). En estos modelos se integrará la curva de caída de la demanda calculada en un paso previo a dichos modelos.

Por lo tanto, fue necesario conocer la función de caída de la demanda con la distancia de las estaciones. Para ello se ha trabajado inicialmente con los datos desagregados de la EDM2004, para poder construir curvas de penetración del uso del Metro y, a partir de éstas, calcular las variables del modelo de regresión, dando más peso a los valores de las franjas más cercanas. Se plantea así una integración de modelos: el de curvas de penetración tiene interés por sí mismo (como modelo de respuesta inmediata, pero elevado margen de error), pero también en tanto que genera los pesos según la distancia que serán utilizados en el modelo de regresión múltiple

Para obtener las funciones de uso del Metro a partir de información de las estaciones y su entorno, una primera tarea fue establecer las franjas que componen las áreas de influencia de las distintas estaciones. Dichas franjas deben contener información relativa a las principales variables explicativas del entorno de la estación, que se han obtenido mediante operaciones de superposición de capas en un SIG.

Las franjas de distancia fueron calculadas a través del viario. Se calcularon cada 100 metros de distancia con respecto a las estaciones de Metro, hasta un límite máximo de 1.500 metros. Para tener en cuenta la competencia entre estaciones, se calcularon polígonos de Thiessen, que fueron superpuestos a las franjas obtenidas anteriormente. El paso siguiente consistió en la superposición de los polígonos de Thiessen y las Zonas de Transporte (ZT), que contienen información correspondiente a la población empadronada para el año 2004, según el Instituto de Estadística de la Comunidad de Madrid. Es frecuente que varias zonas de transporte queden parcialmente incluidas en un polígono de Thiessen, situación resuelta con la asignación proporcional del dato (población por ejemplo) a la superficie contenida en el área del Thiessen. Idéntico procedimiento fue seguido en relación a los polígonos que representan las franjas de distancia. De esta forma se puede disponer de información socioeconómica referida a franjas de distancia pertenecientes a cada una de las estaciones.

Por otro lado, a partir de las coordenadas de los puntos que representa los orígenes de los viajes que usan Metro en primera etapa se obtuvo una capa de puntos en el SIG. Dado que para cada viaje se conoce la estación de entrada, se pudo calcular la distancia a través del callejero entre cada punto de origen y la correspondiente estación. Con estos datos y la población o empleo según franjas se calcularon las funciones de uso de las curvas de penetración.

Otra aplicación importante en el SIG fue obtener algunos datos en torno a las estaciones, como determinar la densidad vial, las paradas de autobuses y estaciones de Cercanías por franjas de distancias. El procedimiento básicamente consistió en superponer las capas de puntos de interés (paradas, estaciones) a las franjas de distancia y luego cuantificar los elementos intersectados.

Dado que la atención se ha puesto sobre las entradas de población que llega andando a las estaciones, los modelos propuestos se han realizado únicamente para las estaciones terminales, pasantes y de conexión, mientras que en los intercambiadores se no han sido considerados. Finalmente, los modelos propuestos fueron validados en función del ajuste de los valores estimados con respecto a los valores observados (entradas registradas en las estaciones de la red de Metro).

### 3.4. Curvas de uso del Metro

Como vimos, las curvas de penetración del uso del Metro están basadas en la relación funcional existente entre ratios de viajes y la distancia a la estación. Para ello se han considerado sólo los viajes en los que se accede a pie al Metro, que como vimos son mayoritarios. El objetivo de las curvas es obtener factores de ponderación para las variables según franjas que serán utilizados en los modelos de regresión.

A continuación se presentan dos curvas de penetración, en función de la procedencia del usuario de Metro, ya acceda desde el hogar o desde un lugar diferente al hogar. Para los viajes que acceden desde el hogar, los ratios de viajes se han calculado dividiendo dichos viajes entre la población residente en la correspondiente franja de distancia. Para los viajes que acceden desde un origen distintos al hogar, los ratios se han calculado dividiendo dichos viajes entre el empleo localizado en la correspondiente franja de distancia.

#### A. Viajes con origen en el hogar

Los resultados muestran que el número de viajes en Metro en los que se accede peatonalmente desde casa está fuertemente relacionado con la distancia a la estación: aumenta con la distancia hasta la franja de 200-300 metros y a partir de allí disminuye de forma paulatina (tabla 3). Por una parte existe una menor propensión a utilizar el Metro desde localizaciones más lejanas (distance decay), pero por otro las franjas de distancia ocupan una superficie mayor cuanto más lejanas están con respecto a la estación.

La distribución de la población residente en el entorno de las estaciones es parecida a la de los viajeros (tabla 3), pero menos asimétrica. Las estaciones de Metro se sitúan lógicamente en espacios con altas densidades. De hecho las densidades de población son decrecientes con la distancia a las mismas. El hecho de que las franjas más próximas a las estaciones contengan menos población se debe a su menor tamaño, ya que se trata de franjas (irregulares) concéntricas.

**Tabla 3: Viajes con origen en el hogar, población y ratios según franjas de distancia.**

Franja de distancia	Viajes	Población	Ratio viajes/población
0-100	34791	108730	0.32
100-200	101644	334581	0.30
200-300	135530	470749	0.29
300-400	122514	512871	0.24
400-500	112880	511110	0.22
500-600	90269	338522	0.27
600-700	58738	289841	0.20
700-800	38786	214671	0.18
800-900	26140	175873	0.15
900-1000	17440	156013	0.11
1000-1100	9914	91814	0.11
1100-1200	7827	85091	0.09
1200-1300	4469	68911	0.06
1300-1400	2846	55879	0.05
1400-1500	1870	51450	0.04
Total	765657	3466106	0.22

Fuente: EDM04.

Si los viajeros entrados en Metro se relacionan con la cantidad de población residente en las franjas de distancia (tabla 3) se observa claramente cómo los ratios tienden a decrecer con la distancia (figura 1), desde más de 0,3 viajes por habitante en las dos primeras franjas hasta menos de 0,15 viajes a partir de la franja de los 900 metros. La regularidad con la que los ratios decrecen con la distancia queda de manifiesto cuando se considera que se ha obtenido un coeficiente de determinación muy alto (0,97).

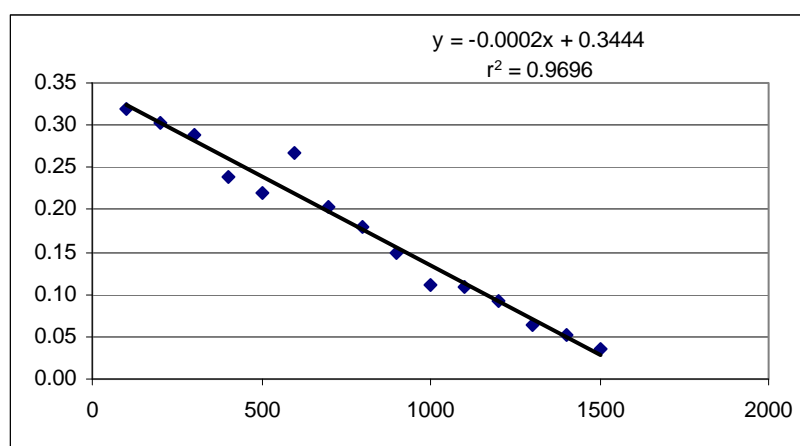


Figura 1: Curva de uso del Metro: ratios de viajes/población según franjas de distancia.

#### B. Viajes sin origen en el hogar

Los viajes en los que se accede sin origen en el hogar presentan una distribución según franjas que difiere sustancialmente a la obtenida para los viajes con origen en el hogar: ahora la distribución es más asimétrica y la mayor parte de los viajeros proceden de franjas más próximas a la estación (tabla 4). La moda se sitúa en las franjas segunda y tercera, mientras antes lo hacía en las franjas tercera y cuarta. O dicho de otra forma: por término medio se anda más para acceder al Metro desde casa que desde los lugares de actividad. Así, la distancia media para acceder al Metro desde el hogar es de 433 metros, que se reduce a sólo 363 metros en el caso de los viajes que no proceden del hogar.

Este hecho se puede explicar básicamente por las distintas distribuciones de población y empleo. El empleo se concentra fuertemente en el entorno de las estaciones, lo que en parte se explica por su fuerte concentración en los barrios más céntricos, donde la densidad de estaciones es alta y por lo tanto la áreas de influencia pequeñas al entrar en "competencia" entre ellas. Así, si se considera exclusivamente la población y el empleo residente en un radio de 1.500 metros con respecto a la estación de Metro más próxima, resulta una distancia media de 533 m para la primera y 441 m para el segundo, que confirma la mayor concentración del empleo en torno a las estaciones.

**Tabla 4: Viajes sin origen en el hogar, empleo y ratios según franjas de distancia.**

Franja de distancia	Viajes	Empleo	Ratio viajes/empleo
0-100	64276	73889	0.87
100-200	86378	224318	0.39
200-300	84099	290266	0.29
300-400	72792	270242	0.27
400-500	59592	211242	0.28
500-600	36672	120616	0.30
600-700	30807	91619	0.34
700-800	13397	60484	0.22
800-900	11562	48184	0.24
900-1000	10202	46125	0.22
1000-1100	3418	24279	0.14
1100-1200	3788	22115	0.17
1200-1300	1868	19623	0.10
1300-1400	2527	17892	0.14
1400-1500	1504	15376	0.10
Total	482882	1536270	0.31

Fuente: EDM04.

Para el conjunto del área considerada (1.500 metros a las estaciones) el ratio de viajes (sin origen en el hogar) por empleo (0,31) es algo más alto que el obtenido para los viajes (con origen en el hogar) por población (0,22). Los ratios son siempre más altos en el caso de viajes sin origen en el hogar, pero especialmente en la primera franja, posiblemente debido a una gran concentración de empleos comerciales que atraen una gran cantidad de viajes. Esta concentración de viajes en las primeras franjas supone una fuerte caída de estos con la distancia, ya no lineal sino potencial (figura 2).

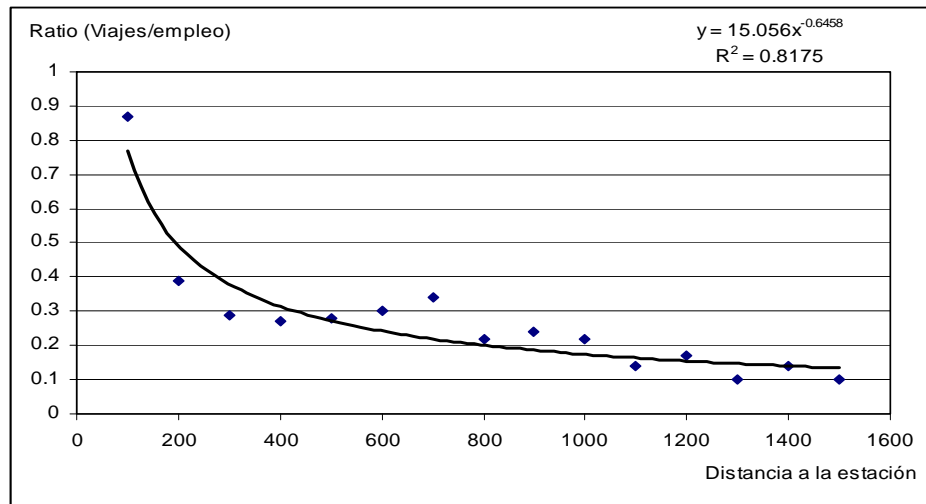


Figura 2: Curva de uso del Metro: ratios de viajes/empleo según franjas de distancia.

Existe, pues, una mayor propensión a utilizar el Metro cuando el origen del viaje se sitúa cerca de las estaciones, tanto para los hogares como en mayor medida para los viajes que parten desde fuera del hogar. No parece lógico que en los modelos de regresión se consideren como variable independiente el valor total de cada variable (por ejemplo, la población) en el área de influencia de la estación, ya que la población más lejana utilizará menos el Metro que la más cercana. Así, estas dos curvas de penetración se han usado para ponderar los datos socioeconómicos del entorno en las estaciones en cada una de las franjas en función de su ratio de viajes/población o viajes/empleo. Se busca dar más peso a los datos en las franjas próximas a la estación, al quedar demostrado que es donde más se utiliza el Metro.

### 3.5 Modelo de regresión lineal múltiple

Por medio del análisis de regresión es posible expresar las complejas relaciones entre variables de manera sencilla. En este apartado se realiza un *Modelo de Regresión Lineal Múltiple* (MRLM) que tiene por objeto explicar el comportamiento de una variable dependiente designada  $Y$ , utilizando información suministrada por los valores de un conjunto de variables explicativas  $X_1, X_2, \dots, X_n$ ; su expresión matemática es conocida:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + u$$

Las entradas utilizadas como variable dependiente fueron las medidas en las estaciones de la red durante el mes de noviembre de 2004. Como variables independientes se han introducido las comentadas en el apartado anterior (tabla 5), que luego serán reducidas



en base a su significación para el modelo. Para ejecutar este análisis se ha utilizado el software SPSS v.14.

**Tabla 5: Variables independientes.**

<i>Características de la población</i>	<i>Características económicas</i>	<i>Características urbanísticas</i>	<i>Características de las estaciones</i>
1. POB_TOT: Población Total	9. EMP_TOT: Total de los empleos	15. DIVER_UsoSUELO : Un índice de mezcla de usos del suelo	17. TIPO_EST: Tipos de estaciones (pasante, conexión, intermodal, terminal)
2. POB_OCU: Población Ocupada	10. EMP_IND: Empleos en la industria	16. DEN_VIAS: Densidad del viario	18. CORONAS: Localización de la estación en las coronas metropolitanas
3. POB_EXTR: Población Extranjera	11. EMP_COM: Empleos en comercio		19. DIST_EUCLID: Distancia euclidiana medida al centro de la Red
4. POB_MUJ: Población Mujer	12. EMP_SAN: Empleos en la sanidad		20. ORD_EST: Orden de la estación (según Línea)
5. POB_VIE: Población mayor de 65 años	13. EMP_EDU: Empleos en la educación		21. NUM_BOCA: La cantidad de accesos que dispone la estación
6. POB_JOV: Población menos de 19 años	14. EMP_ADM: Empleos en la administración		22. NUM_LIN: La cantidad de líneas que pasan por la estación
7. NUM_ESTUD: N° de Estudiantes			23. Velocidad de Explotación (según Línea) *
8. HO_SINCO: porcentaje de de hogares sin vehículo particular			24. NUM_TREN: La cantidad de trenes disponibles (según Línea) *
			25. CAP_LIN: Capacidad de la línea *
			26. CAP_TREN: Capacidad de los trenes (según Línea)*

\*Están referidos a datos de las líneas de Metro asignadas a sus correspondientes estaciones, y en el caso de que por una de ellas pase más de una línea, se le otorga el mejor valor.

Inicialmente se ha realizado un análisis de correlación bivariada entre cada una de las variables independientes con la variable dependiente, que permite obtener una primera aproximación al poder explicativo de cada variable incorporada al modelo. Los análisis se realizaron agrupando las variables explicativas en diferentes distancias: 400, 500, 600, 900, 1200 y 1500 metros, siempre usando las curvas de penetración como elemento de ponderación en esa agrupación.

Las mayores correlaciones se dan con una agrupación de la información en un radio de 500 metros. La proximidad entre estaciones en el centro de la ciudad hace necesario acortar los radios de distancia. A partir de 500 metros, a medida que se incrementan los radios de distancia las correlaciones se reducen, siendo menores en un radio intermedio (900 metros) y más bajas en radios muy grandes (1500 metros), lo que parece mostrar como en estas distancias entran en juego con más fuerza otras variables en la elección modal o el efecto de competencia entre estaciones, que perjudican los modelos finales. En consecuencia en los modelos de regresión lineal múltiple se han introducido las variables para un radio de distancia de 500 metros.

Igualmente, con objeto de conocer las relaciones entre las propias variables independientes y la posible incidencia en los modelos finales de factores de auto-correlación entre variables independientes se calculó una matriz con los coeficientes de correlación de Pearson que se establecen entre ellas. Las relaciones más elevadas (coeficientes por encima de 0.5) se presentan entre actividades dirigidas a la población, como el empleo en educación, sanidad y en comercio, y la distribución de dicha población, ya sea total u ocupados. Existe además una relación clara entre la propia población total y los ocupados, los ocupados y el empleo industrial, y la población total y el número de estudiantes y sobre todo entre estos últimos y el empleo en educación. Entre las variables utilizadas para la caracterización de la población aparece correlación entre el porcentaje de extranjeros y la proporción de hogares sin coche. Mientras que entre las variables de diseño y modelo territorial, la densidad viaria tiene una correlación alta con la distribución de la población, e igualmente aparece correlación (negativa) entre la distancia al centro y los hogares sin coche o la distancia al centro y la densidad viaria.

Puede establecerse un primer modelo sencillo donde se introducen las principales variables agregadas, intentando que queden representadas las 3D. Las variables incorporadas son: población total, empleo total, densidad de vivienda multifamiliar,

densidad viaria, diversidad de usos y, como variables de servicio de la red, el número de líneas y número de bocas de la estación. En la tabla 5 se presentan los modelos resultantes. Los cinco modelos permiten un ajuste de la estimación del 79-78%. El más sencillo (modelo 5), con apenas cinco variables permite alcanzar una explicación un Rcuadrado de 0.78%.

**Tabla 5. Resumen de los modelos**

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,888(a)	,789	,776	67728,991
2	,888(b)	,789	,777	67499,833
3	,887(c)	,787	,777	67489,182
4	,886(d)	,785	,776	67643,921
5	,885(e)	,783	,776	67733,509

a Variables predictoras: (Constante), NUM\_BOCA, POBEXT, DENVIA, NUM\_LIN, Diver\_UsoSuelo, HOSCOCH, POBTOT, EMPTOT, D\_Vivmultifam

b Variables predictoras: (Constante), NUM\_BOCA, POBEXT, DENVIA, NUM\_LIN, HOSCOCH, POBTOT, EMPTOT, D\_Vivmultifam

c Variables predictoras: (Constante), NUM\_BOCA, POBEXT, NUM\_LIN, HOSCOCH, POBTOT, EMPTOT, D\_Vivmultifam

d Variables predictoras: (Constante), NUM\_BOCA, POBEXT, NUM\_LIN, POBTOT, EMPTOT, D\_Vivmultifam

e Variables predictoras: (Constante), NUM\_BOCA, POBEXT, NUM\_LIN, POBTOT, EMPTOT

Los coeficientes Beta estandarizados mostraron que las variables con mayor poder explicativo en los modelos son aquellas relacionadas con el nivel de servicio: número de líneas y número de bocas (en este caso a través de áreas de cobertura mayores). Por detrás en poder explicativo se sitúan las variables de empleo y población total. A continuación se presenta la ecuación del modelo más sencillo, para el que en el siguiente apartado se han testado los errores.

$$\text{ENTRADAS} = -82477,290 + (0.563 * \text{EMPTOT}) + (1.059 * \text{POBTOT}) + (1745,988 * \text{POBEXT}) + (24518,508 * \text{NUM\_BOCA}) + (150854,88 * \text{NUM\_LIN})$$

A pesar de haber realizado numerosas pruebas, otros modelos donde se han introducido un conjunto mayor de variables (desagregando la población y el empleo según tipos o incorporando la información referida al nivel de servicio de las estaciones) no han incrementado en gran medida los valores Rcuadrado. Parece que los factores de autocorrelación entre las propias variables explicativas pueden influir en los modelos.

### 3.5 Estimación de los errores del modelo

Con el fin de testar el modelo propuesto se han obtenido las entradas estimadas a partir de los mismos en cada una de las estaciones y comparado con las entradas según el aforo. Acorde con la capacidad explicativa del modelo, los datos estimados presentan un error medio según estación del  $\pm 24.8\%$ , con una cierta tendencia a la subestimación de las entradas.

Los resultados no parecen presentar una lógica espacial en las sobreestimaciones ni subestimaciones, de manera que éstas aparecen indistintamente tanto en estaciones del centro de la ciudad como en los lugares más periféricos. Igualmente, tampoco aparece un patrón en los errores según tipos de estaciones, y en estaciones con conexión o terminales presentan sobre o subestimaciones indistintamente.

En la tabla 6 los datos son agregados según líneas. La aleatoriedad de sobreestimaciones y subestimaciones queda nuevamente patente. El ajuste es muy bueno en las líneas 5 y 8. Sin considerar las líneas 11 y 14, donde el número de estaciones es muy reducido, en las líneas 1, 3, 6 y 10 el modelo subestima las entradas, mientras en las restantes las estimaciones son sobredimensionadas. Las compensaciones entre sobre y subestimaciones de las estaciones reduce los errores según líneas, de manera que únicamente en la línea 2 presenta un error por encima del 25%, mientras en las líneas 10 y 12 los valores alcanzan el 20% de error.

Tabla 16: Estimación de los errores según línea.

Línea	Estaciones consideradas	Entradas observadas	Entradas estimadas	Diferencia	Diferencias relativas
1	22	6702243	6241817	-460426	-6.87
2	9	1931514	2483514	552000	28.58
3	7	3026240	2421845	-604395	-19.97
4	15	4086851	4547214	460363	11.26
5	14	3198750	3169617	-29133	-0.91
6	15	5075123	4460182	-614941	-12.12
7	18	3835249	4083068	247819	6.46
8	4	1144196	1138016	-6180	-0.54
9	19	3367346	3697664	330318	9.81
10	8	2059344	1591411	-467933	-22.72
11	2	167860	265008	97148	57.87
12	21	2280063	2794630	514567	22.57
14	1	136334	155235	18901	13.86

#### 4. Consideraciones finales

Frente a la insostenibilidad de un modelo metropolitano basado en el transporte individual, una de las tendencias con mayor aceptación en los organismos de gestión-planificación es potenciar el uso del transporte público, y por esta razón resulta de gran importancia sumar estudios referidos al análisis de la demanda. La mayor parte de los trabajos para estimar la demanda del transporte público parten del clásico modelo de cuatro etapas, que a pesar de su complejidad metodológica, produce resultados con una considerable desviación con respecto a los valores reales (FLYVBJERG, HOLM Y BUHL, 2005 y 2006). Uno de los mayores problemas de este modelo estriba en la escala de análisis (regional) y en no considerar la caída de la demanda con la distancia (*distance decay*).

En esta comunicación se ha presentado un modelo de estimación de la demanda, aplicado a las entradas en las estaciones del Metro en Madrid, donde se obtienen errores similares (y a veces menores) al de cuatro etapas, pero con la ventaja de que su aplicación es rápida, sencilla, y además considera a la distancia como un elemento primordial para explicar el uso del transporte público. En este sentido cabe destacar el avance en la integración de los modelos de estimación de la demanda: las curvas de *distance decay* en el análisis de regresión lineal múltiple, como factor de ponderación en las franjas de distancia.

## 5. Bibliografía

BLACK, W.R. (1996), Sustainable transportation: a US perspective. *Journal of Transport Geography*, 4. 151-159.

CERVERO, R. (2001). Walk-and-Ride: factors influencing pedestrian access to transit. *Journal of Public Transportation*. Vol. 3. Nº 4. Pp: 1-23.

CERVERO, R. (2004). Transit Oriented Development in America: contemporary practices, impacts, and policy directions. *International Planning Symposium on Incentives, Regulations and Plans*. National Center for Smart Growth Research and Education, University of Maryland.

CERVERO, R; KOCKELMAN, K. (1997) Travel Demand and the 3Ds: Density, Diversity, and Design. *Transportation Research-Part D 2*: 199-219.

CHU, X. (2004). *Ridership Models at the Stop Level*. National Center of Transit Research. University of South Florida. 69 p.

FLYVBJERG, B., HOLM, M.K.S. Y BUHL, S.L. (2005): How (In)accurate Are Demand Forecasts in Public Works Projects? The Case of Transportation. *Journal of the American Planning Association*, 71, 2, 131-146

FLYVBJERG, B.; HOLM, M. K. S. Y BUHL, S. L. (2006). Inaccuracy in Traffic Forecasts. *En: Transport Reviews*. Vol. 26, Nº 1, January. Pp: 1-24.

GUTIÉRREZ, J. Y GARCÍA PALOMARES, J.C. (en prensa): Distance measure impacts of public transport service areas. *Environment and Planning B, Planning and Design*.

KAPLAN, B; ENGLISHER, L. Y WARNER, M. (2003). *Actual Versus Forecast Ridership on MetroLink in St. Clair County, Illinois*. *En: Transportation Research Circular E-C058: 9th National Light Rail Transit Conference*.

KEIJER, MJN. y RIETVELD, R. (2000), How do people get to the railway station? The Dutch experience. *Transportation Planning and Technology*, 3-23, 215-135.

KIMPEL, T.J.; DUEKER, K.J. y EL-GENEIDY, A.M. (2006), Using GIS to measure the effect of overlapping service areas on passengers boardings at bus stops". URISA (<http://urisa.org/kimpel>).

KIMPEL, T. (2001). *Time Point-Level Analysis of Transit Service Reliability and Passenger Demand*. Portland State University.

LEVINSON, HS. Y BROWN-WEST, O. (1984). Estimating bus Ridership. *Transportation Research Record*, 965. 8-12.

MARSHALL, N. Y GRADY, B. (2006): Sketch Transit Modeling Based on 2000 Census Data. *TRB 2006 Annual Meeting CD-ROM*.

MORENO, A. Y ESCOLANO, S. (1992). *Los Servicios y el Territorio*. Colec. Espacios y Sociedades. Serie General N° 19. Editorial Síntesis. Madrid. 240 p.

MURRAY, A.T., DAVIS R., STIMSON, R.J., y FERREIRA, L. (1998). Public transport access. *Transportation Research D*, 3-5, 319-328.

NELSON, D. Y NILES, J. (1999). *Essentials for Transit-Oriented Development Planning: analysis of non-work activity patterns and a method for predicting success*. En: *Proceeding of the Seventh National Conference of the Transportation Research on the Application of Transportation Planning Methods*. Boston.

TAYLOR, B. Y FINK, C. (2003). *The Factors Influencing Transit Ridership: a review and analysis of the ridership literature*. Working Papers. Department of Urban Planning. UCLA, Los Angeles.

WALTERS, G. Y CERVERO, R. (2003). *Forecasting Transit Demand in a Fast Growing Corridor: the direct-ridership model approach*. Fehrs and Peers Associates.

WEBER, J. (2006), Reflections on the future of accessibility. *Journal of Transport Geography*, 14. 296-398.

ZHAO, F.; CHOW, L.F.; LI, M.T.; GAN, A. y UBAKA, I. (2003), Forecasting transit walk accessibility: a regression model alternative to the buffer method. *Transportation*

*Research Board Annual Meeting* ([http://www.ltrc.lsu.edu/TRB\\_82/TRB2003-001007.pdf](http://www.ltrc.lsu.edu/TRB_82/TRB2003-001007.pdf)).

### **Agradecimientos**

Esta comunicación recoge parte de los resultados obtenidos en un trabajo de investigación más amplio realizado por los autores para el Área de Movilidad y Accesibilidad del Consorcio Regional de Transportes de Madrid. Igualmente la comunicación se enmarca dentro del proyecto TRA2005-06619/Modal, del Plan Nacional de I+D+I del Ministerio de Educación y Ciencia.