

# **ESTIMACIÓN DIRECTA DE LA DEMANDA DE TRANSPORTE A NIVEL DE ESTACIÓN MEDIANTE EL USO DE LA REGRESIÓN GEOGRÁFICAMENTE PONDERADA**

Javier Gutiérrez Puebla, Universidad Complutense de Madrid, [javiergutierrez@ghis.ucm.es](mailto:javiergutierrez@ghis.ucm.es)  
Osvaldo Daniel Cardozo, Universidad Nacional del Nordeste, [odcardozo@hum.unne.edu.ar](mailto:odcardozo@hum.unne.edu.ar)  
Juan Carlos García Palomares, Universidad Complutense de Madrid, [jcgarcia@ghis.ucm.es](mailto:jcgarcia@ghis.ucm.es)

## **RESUMEN**

En los últimos años se han desarrollado modelos de predicción directa de la demanda a nivel de estación basados en SIG y análisis de regresión múltiple. Estos modelos suelen utilizar el modelo regresión múltiple, asumiendo la estabilidad paramétrica de los resultados. El presente estudio propone un modelo basado en la regresión geográficamente ponderada para estimar la demanda del metro de Madrid. Los resultados indican que este modelo no sólo obtiene un mejor ajuste que el tradicional. Además la información que suministra sobre la variación espacial de las elasticidades de los predictores y su significación estadística permite alcanzar resultados más realistas.

*Palabras claves: demanda de transporte público, modelos de predicción directa, regresión ponderada geográficamente.*

## **ABSTRACT**

In recent years direct forecasting models at the station level have been developed based on GIS and multiple regression analysis. These models typically use the multiple regression model, assuming parametric stability of results. This study proposes a model based on geographically weighted regression to estimate metro boardings in Madrid. Results show that the model not only get a better fit than the traditional one. In addition, the information supplied regarding the spatial variation of the predictors elasticities and their statistical significance provides more realistic results.

*Keywords: transit ridership, direct forecasting models, geographically weighted regression.*

## **1 INTRODUCCIÓN**

Modelizar y predecir la demanda de viajes es una de la cuestión crucial en la planificación del transporte. En el marco generalizado de austeridad en el que se encuentran las administraciones públicas, los estudios de demanda adquieren una importancia todavía mayor (Blainey, 2010). En el caso de las estaciones de transporte público esta cuestión tiene un fuerte trasfondo geográfico, aunque no siempre se considere ese trasfondo de forma provechosa. Es un hecho probado que la demanda de las estaciones puede incrementarse considerablemente creando proximidad en la distribución espacial de la oferta -paradas o estaciones- y los factores que explican su utilización, como las altas densidades de población, la concentración del empleo, la mezcla de usos del suelo, etc. En este sentido Kobayashi and Lane (2007) señalan la necesidad de analizar con mayor detalle espacial las características del entorno de las estaciones para comprender mejor la conexión entre inversión y uso del servicio.

Los expertos en planificación del transporte han dedicado mucho tiempo y esfuerzo al estudio de la demanda de transporte. Para ello utilizan generalmente el denominado modelo de cuatro etapas (McNally, 2007; Ortúzar y Willumsen, 2008), un modelo de eficacia probada, pero muy sofisticado y de alto coste, que sin embargo no contempla adecuadamente las características del entorno de las estaciones, al trabajar con zonas de transporte relativamente extensas (Marshall y Grady, 2006; Cervero, 2006). Recientemente han surgido los denominados modelos de predicción directa de la demanda a nivel de estaciones, que tratan de superar esa deficiencia formulando modelos de regresión múltiple -Ordinary Least Squares (OLS)- alimentados con variables sobre las características de las estaciones y sus entornos, calculadas éstas últimas mediante herramientas SIG (Walters y Cervero, 2003; Kuby et. al., 2004; Cervero, 2006). Estos modelos podrían aumentar su capacidad explicativa si utilizaran técnicas de estadística espacial del tipo de la regresión geográficamente ponderada -Geographically Weighted Regression (GWR)-. Una de las grandes ventajas de la modelización espacial es la capacidad de medir la inestabilidad paramétrica -y en consecuencia también la heterogeneidad espacial- a partir de la magnitud que presentan los coeficientes a través del territorio (Clark, 2007). Esto no es posible en la regresión tradicional (OLS) porque asume la hipótesis de estabilidad paramétrica, lo cual equivale a pensar que los coeficientes calculados no presentan diferencias significativas en el espacio.

El objetivo principal del presente artículo es demostrar las ventajas de los modelos de regresión geográficamente ponderada en la estimación de la demanda del transporte público a nivel de estación. Se utiliza como caso de estudio la red de metro de Madrid. Se contrastan los resultados obtenidos mediante los modelos OLS y GWR.

## **2 LOS MODELOS DE PREDICCIÓN DIRECTA DE DEMANDA A NIVEL DE ESTACIÓN Y LA REGRESIÓN GEOGRÁFICAMENTE PONDERADA**

La historia de la modelización del transporte ha estado dominada por el modelo de cuatro etapas (generación, distribución, asignación y reparto modal) (McNally, 2007). Se trata de una familia de modelos interrelacionados, de gran complejidad y con un elevado coste de implantación y mantenimiento (Marshall and Grady, 2006). Dado que su objetivo inicial en los años cincuenta

era la predicción del tráfico y preferentemente a escalas regionales, no es de extrañar que su eficacia para la previsión de viajes disminuya en otros modos de transporte (público) y en escalas de detalle (nivel de estación). A la vista de estas circunstancias han aparecido los denominados modelos de predicción directa (Cervero, 2006), de respuesta rápida y bajo coste. Estos modelos se basan en la aplicación del análisis de regresión múltiple (OLS) para estimar directamente la demanda de las estaciones a partir de un conjunto de predictores relativos a las características de la estación y de su área de influencia. Estas últimas se calculan con la ayuda de Sistemas de Información Geográfica, lo que permite un tratamiento detallado de las mismas, frente al excesivo nivel de agregación espacial que caracteriza a las zonas de transporte utilizadas en el modelo de cuatro etapas.

Como es sabido, el modelo de cuatro etapas utiliza modelos de regresión múltiple o análisis de categorías para la generación y atracción de viajes, mientras que en la etapa de distribución se suele recurrir a modelos de entropía y en la de reparto modal normalmente se utilizan modelos de elección discreta. En cambio los modelos de predicción directa estiman la demanda a nivel de estación directamente a través del análisis de regresión múltiple (Walters y Cervero, 2003; Chu, et al. 2004; Kuby et al., 2004; Gutiérrez et. al, 2011). Sus unidades de análisis son los entornos de las estaciones, no las zonas de transporte. Esto supone que el número de observaciones suele ser relativamente pequeño, lo que reduce los grados de libertad y el número de variables que pueden ser introducidas en el modelo (Cervero, 2006). En opinión de Cervero (2006), estos modelos pueden aportar resultados aproximados (predecir órdenes de magnitud), pero no pueden sustituir al modelo de cuatro etapas. Sin embargo, los modelos de predicción directa permiten analizar con precisión las variables de las áreas de influencia de las estaciones y, a partir de las elasticidades, conocer las repercusiones de posibles actuaciones urbanísticas (nuevos desarrollos, operaciones de densificación, etc.) sobre la demanda de viajeros, lo que resulta de particular interés en la planificación de los TOD (Transit Oriented Developments) (Cervero, 2006). Además, la combinación de bajo coste y sencillez les convierten en una alternativa razonable para municipios de tamaño pequeño o de recursos limitados, que necesitan evaluar de forma rápida y precisa la demanda prevista en sus planes de inversión en infraestructuras de transporte. Hay que hacer notar en este sentido que el modelo de cuatro etapas es alimentado con costosas encuestas de movilidad, mientras que los datos de movilidad necesarios para los modelos de predicción directa son simplemente los aforos de las estaciones.

El empleo del OLS en los modelos de estimación directa tiene el defecto de enmascarar las variaciones geográficas de las relaciones entre las variables (Lloyd and Shuttleworth, 2005) al adoptar el supuesto de relaciones estacionarias entre variables, ignorando la posibilidad de que existan variaciones locales a causa de la heterogeneidad propia del espacio (Páez, 2006). Sin embargo es bien conocido que los datos geográficos normalmente no tienen un comportamiento estacionario (Haining, 2010). Cuando en los modelos de regresión se asume que el poder explicativo (medido a través del coeficiente de determinación) es igual para todo el conjunto de observaciones, la idea de un comportamiento uniforme y constante del ajuste a través del espacio geográfico resulta por lo menos sospechosa, y estaríamos frente al problema conocido en la literatura como inestabilidad paramétrica (Arbia, 2006; Páez et al. 2010; Haining, 2010). Aceptar como verdadero aquel supuesto es negar la posibilidad de conocer las interacciones locales que se producen dentro del área de estudio.

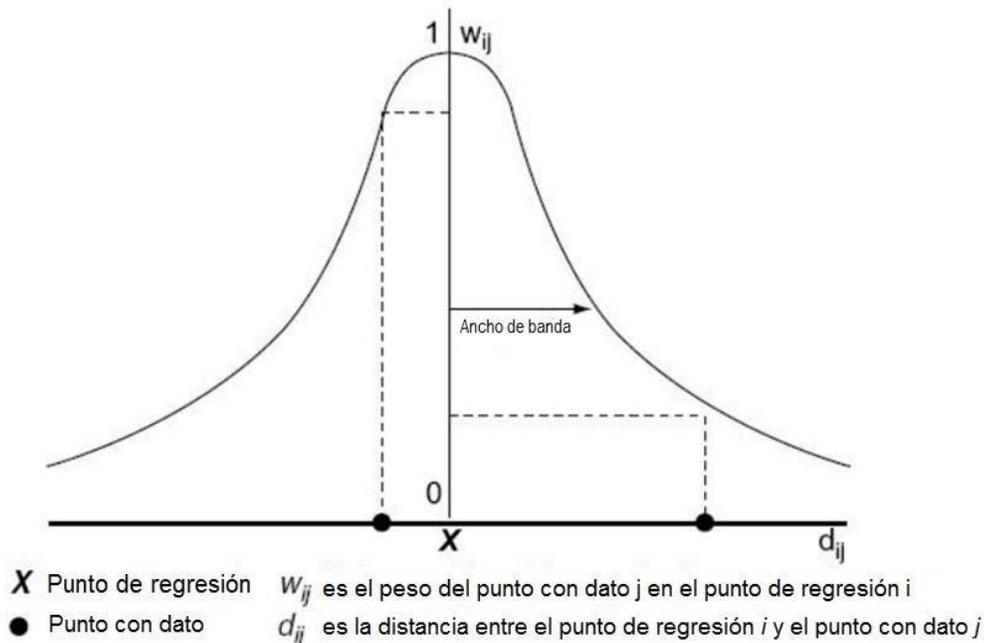
En el modelo OLS cada observación es considerada de forma independiente. Pero lo cierto es que los datos espaciales no cumplen la hipótesis de independencia, debido a que normalmente están autocorrelacionados, por lo que la fuerza de la relación entre las variables del modelo no será la misma en toda el área de estudio. En este sentido Lloyd and Shuttleworth (2005) destacan la necesidad de incluir en los modelos especificaciones más apropiadas, que consideren la naturaleza intrínseca de los datos espaciales, que normalmente están autocorrelacionados. La autocorrelación espacial se produce cuando el valor observado de una variable en un lugar determinado depende de los valores de la misma variable observados en lugares vecinos, por ello con frecuencia también es llamada dependencia espacial.

Los modelos estadísticos tradicionales pueden mostrar problemas de especificación cuando no consideran la presencia de autocorrelación espacial en los datos. Una consecuencia importante es el riesgo de estimar coeficientes ineficientes para representar la magnitud de la relación entre las variables, por lo que las pruebas de significación estadística sobre los mismos serán cuestionables debido a la inflación en los errores estándar. Por esta razón, antes de cualquier interpretación de los resultados de un análisis regresión, es necesario someter los residuos a test específicos (como el el I de Moran) para demostrar estadísticamente su aleatoriedad espacial. En caso contrario, es decir, en presencia de autocorrelación espacial, es recomendable evaluar el empleo de alguna técnica diseñada específicamente para hacer frente a este tipo de situaciones, como la regresión geográficamente ponderada (para más información sobre modelos de regresión espaciales (para más información ver por ejemplo Anselin, 1988 o Chasco, 2003).

Brunsdon et al., (1996) introdujeron el término regresión geográficamente ponderada (GWR) para aludir a una familia de modelos de regresión “ajustados al espacio” donde es posible observar las variaciones espaciales de los parámetros estimados y con ello saber exactamente dónde y cuánto es el efecto de una variable explicativa sobre la dependiente. Se trata de ajustar tantas regresiones como observaciones (unidades espaciales) se consideren en el análisis, en base al concepto de distance decay (se da más peso a las observaciones más próximas y menos a las más lejanas), operacionalizado por medio de una función kernel que simule el efecto de caída con la distancia (Figura 1). En consecuencia se pueden realizar estimaciones ajustadas a cada observación, aplicando su correspondiente ecuación.

Aunque la GWR es una técnica todavía muy joven, en el ámbito del transporte se ha utilizado para explorar las variaciones espaciales que presentan variables como los accidentes de tráfico (Hadayeghi et al., 2010), la accesibilidad (Mountain et al., 2007), la tenencia de coche (Clark, 2007), el tráfico medio anual de las carreteras (Zhao and Park, 2004), las distancias medias recorridas por los commuters (Lloyd and Shuttleworth, 2005), el uso del transporte público (Chow et al., 2006 y 2010; Kobayashi y Lane, 2007; Blainey y Preston, 2010) o la influencia de las infraestructuras de transporte en el uso del suelo (Páez, 2006). Sin embargo la GWR apenas ha sido incorporada a los modelos de predicción directa de la demanda a nivel de estación.

Figura 1: Esquema de la función kernel y ancho de banda en la regresión geográficamente ponderada.



Fuente: Fotheringham, et al., 2002

### 3 DATOS Y METODOLOGÍA

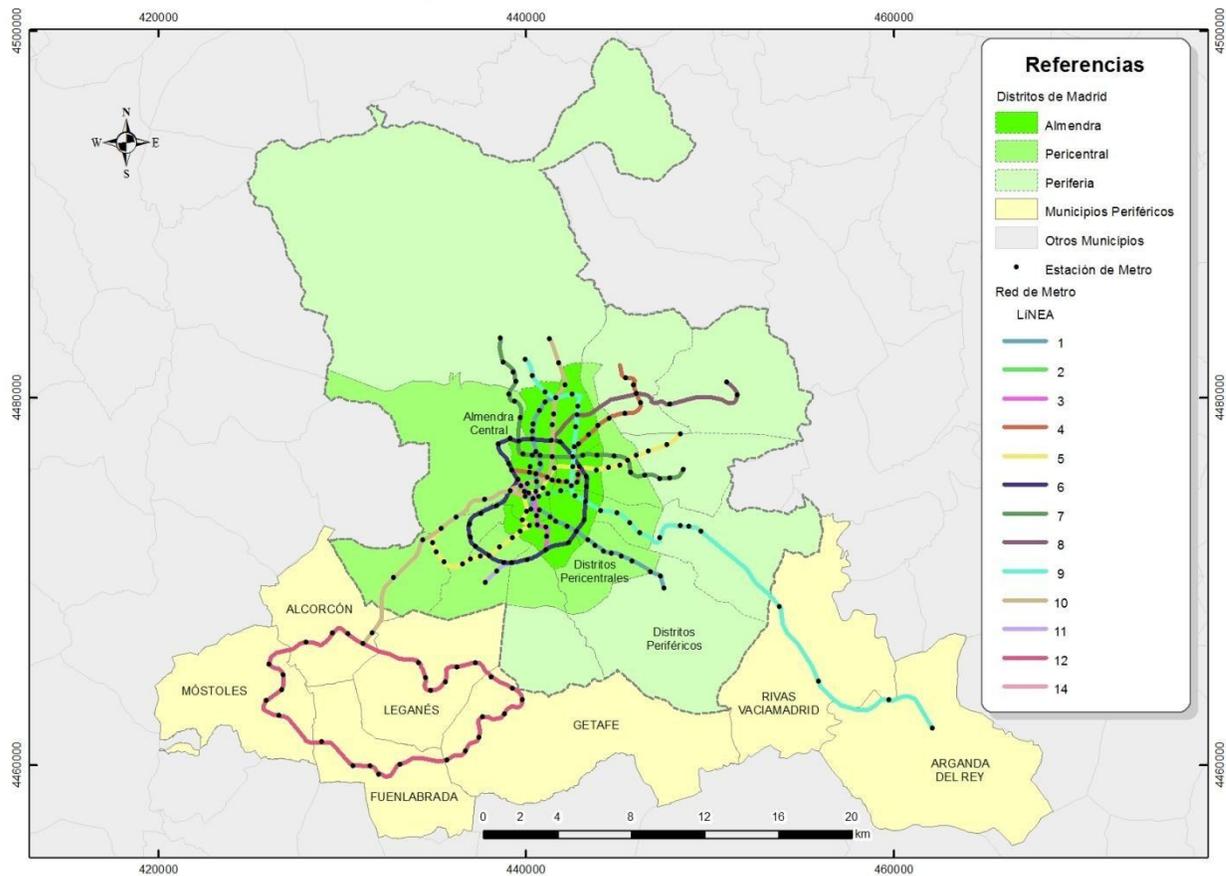
#### 3.1 El área de estudio y los datos

El estudio se centra en la red de Metro de Madrid, una de las más largas de Europa. Cuenta con un total de 12 líneas y 190 estaciones (Figura 2). El año de referencia es 2004, ya que para esa fecha se dispone de un aforo de viajeros en todas las estaciones y una encuesta de movilidad.

Para alimentar los modelos de regresión se dispuso de las siguientes capas de información en formato shape (ArcGIS):

- Estaciones de la red de metro de Madrid (Consortio Regional de Transportes de Madrid).- Esta capa contiene información relativa a las características de las estaciones (por ejemplo, número de líneas de metro que pasan por la estación) y al número de viajeros que entraron en cada una de las 190 estaciones de la red en el mes de noviembre de 2004 (variable dependiente en los modelos de regresión).
- Red viaria.- Esta capa fue utilizada para delimitar las áreas de influencia de las estaciones por medio de distancias a través del viario -no en línea recta-.
- Zonas de transporte.- Contiene datos como población, empleos, ocupados, etc., referidas también al año 2004, que permiten caracterizar los entornos de las estaciones.
- Estaciones de las redes de autobuses urbanos e interurbanos.- Estas capas fueron utilizadas para calcular el número de líneas alimentadoras en el entorno de las estaciones de metro.

Figura 2: Red de Metro de Madrid



Fuente: Elaboración propia

Para calcular las variables del entorno de las estaciones (población, empleo, usos del suelo, etc.), la mayor parte de los estudios anteriores delimitan áreas de influencia a partir de distancias euclidianas (por ejemplo, Walters y Cervero, 2003; Chu, et al. 2004). En este trabajo las distancias se calculan a través de la red con herramientas SIG (Gutiérrez y García-Palomares, 2008) hasta un umbral de distancia de 800 metros en torno a las estaciones, umbral en el que se obtuvo el máximo ajuste para la mayor parte de los modelos calibrados (Gutiérrez et al., 2010).

### 3.2 Metodología

Inicialmente se consideraron diez variables candidatas como predictores. De ellas siete están calculadas sobre el área de influencia de 800 metros con respecto a cada estación: población total, población ocupada, estudiantes, hogares sin coche, mezcla de usos del suelo, densidad viaria y número de empleos. Dos están calculadas sobre un área de influencia de 200 metros de cada estación: número de líneas de autobuses urbanos y suburbanos con parada en ese área de influencia (líneas alimentadoras). Finalmente una se refiere al atractivo de la estación para el usuario (número de líneas que pasan por la estación). Estas variables fueron identificadas a partir de varios trabajos que discutían los factores que afectan al número de usuarios que entran en las estaciones (Blainey y Preston, 2010; Chu, 2004; Chow et al., 2010; Gutiérrez et al., 2011; Kuby

et al., 2004; Walters y Cervero, 2003). Sin embargo algunas de estas variables tenían en nuestro caso un bajo poder explicativo sobre el número de entradas y además varias de ellas presentaban problemas de colinearidad. Se construyeron múltiples modelos OLS combinando distintas variables independientes. Finalmente se eligió el que se incluye en este trabajo por tratarse de un modelo simple, con un relativamente alto poder explicativo, que incluye cuatro variables claramente relevantes desde el punto de vista de la planificación del transporte, y que no presentaban problemas de colinearidad.

Dado que los residuales del modelo OLS elegido presentaban autocorrelación espacial, se recurrió a la regresión geográficamente ponderada. La característica más importante de la GWR es que considera explícitamente la componente espacial de los datos, incorporando en su ecuación el valor de las coordenadas geográficas de las observaciones, ya sea un punto, centroide de polígono o celda. Frente a la regresión múltiple clásica, la particularidad de la regresión geográficamente ponderada es que los coeficientes  $\beta_j$  ( $j = 0, 1, \dots, p$ ) de los  $j$  predictores  $x_j$  ( $j = 1, \dots, p$ ) varían para cada localización, es decir, que para cada localización definida por sus coordenadas  $(u_i, v_i)$ , el valor de la variable dependiente  $y_i$  es estimado según:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i) x_1 + \beta_2(u_i, v_i) x_2 + \dots + \beta_p(u_i, v_i) x_p \quad (1)$$

El análisis de regresión tradicional emplea un método que minimiza las diferencias al cuadrado respecto a la línea de ajuste y obtiene un valor para el conjunto de observaciones, por lo tanto adopta una perspectiva global; este procedimiento no es idóneo cuando el valor de las relaciones del modelo cambia a través del área de estudio (Rosensheim, 2008). En cambio la regresión geográficamente ponderada es un modelo local. El modelo ponderado geográficamente ajusta una regresión para cada localización en la que los vecinos tienen más peso en función a su proximidad. Las estimaciones de los parámetros de regresión en cada localización permiten evaluar cómo y cuánto dichos parámetros varían en el espacio geográfico, superando de esta forma la grave limitación que presentan los modelos globales.

Algunas de las ventajas más importantes que justifican el empleo de la GWR son:

- Permite moverse desde una perspectiva global a un análisis local del problema, obteniendo un mayor grado de detalle y precisión.
- La posibilidad de estimar coeficientes de determinación locales para cada unidad espacial a partir de los valores de un conjunto de observaciones vecinas, permite conocer la forma en que se combinan localmente las variables de la regresión para obtener el “ajuste específico” en una localización.
- Los coeficientes de cada uno de los predictores (elasticidades) varían de una unidad espacial a otra (inestabilidad espacial).
- La desagregación del coeficiente de determinación ( $R^2$ ) global en coeficientes locales y el análisis de su distribución geográfica permiten reconocer dónde las variables independientes tienen un mayor o peor poder explicativo.
- En la gran mayoría de los casos, esta clase de regresión anula o reduce el problema de la autocorrelación espacial de los residuos.

- La implementación de esta técnica en los Sistemas de Información Geográfica (SIG) facilita la elaboración de una amplia variedad de mapas con los resultados generados: variables dependientes e independientes,  $R^2$  locales, coeficientes locales de los predictores (elasticidades), valores  $t$ , residuos estandarizados...
- Es posible generar superficies interpoladas para conocer la distribución espacial continua de los parámetros y aplicar los principios de la “predicción espacial” para hallar los valores de las observaciones que faltan.
- Al proveer resultados específicos para cada localización, éstos pueden ser usados como evidencias para apoyar políticas o tomas de decisiones locales; por eso con frecuencia estas técnicas son llamadas “basadas en el lugar”.

#### **4 PRINCIPALES RESULTADOS**

La selección del modelo se apoyó en los métodos de los paquetes estadísticos convencionales y en la lógica de los modelos de respuesta rápida, buscando conciliar un alto poder explicativo con un bajo número de variables, de fácil obtención. El modelo final incorpora cuatro variables independientes: tres relativas al área de influencia de la estación (cantidad de ocupados, cantidad de empleos, número de líneas de autobuses interurbanos) y una relativa a las características de las estaciones (número de líneas que pasan por la estación).

El ajuste del modelo global (OLS) ofreció unos  $R^2$  y  $R^2$  ajustado de 0,56 y 0,57, respectivamente, lo que significa que con solo cuatro variables se explica un poco más de la mitad de la variabilidad en el número de viajeros que entran en las estaciones de la red de Metro (Tabla 1). Todas las variables independientes son significativas a un nivel de 0,05 y presentan los coeficientes esperados. Los valores de VIF (todos por debajo del umbral de 7,5) indican que no hay problemas de multicolinealidad entre las variables explicativas. El elevado valor del estadístico F (60,6) y su bajo p-value asociado (0,000000) denotan la elevada significación estadística del modelo. Sin embargo, los resultados de los test (I de Moran global y local) aplicados, indican presencia de autocorrelación espacial en los residuos (Tabla 3 y Figura 3). Por lo tanto existen evidencias suficientes para recurrir a la regresión geográfica.

Manteniendo constante el número de observaciones y variables, se empleó ArcGIS versión 10 para estimar una GWR con un kernel gaussiano y ancho de banda adaptativo por minimización de Akaike. Se eligió un ancho de banda variable debido a la irregular distribución de las estaciones, de forma que el número de estaciones considerado en cada ajuste es el mismo para todas las observaciones. Los  $R^2$  y  $R^2$  ajustado obtenidos son de 0,73 y 0,70, respectivamente, lo cual supone una mejora muy importante con respecto al modelo OLS (Tabla 2). También resulta un valor más reducido del AIC, así como de otros parámetros (Sigma y Desvío Standard) referidos al error en el modelo (Tabla 2). Así mismo, el análisis de los residuos también muestra mejores resultados en la GWR que en la OLS. Esta mejora se comprueba estadísticamente con los datos de la Tabla 3 relativos al índice I de Moran, que permiten descartar la existencia de autocorrelación espacial en los residuos. Más claros aún son los mapas de clusters y significación (Figura 3) donde los residuos del modelo tradicional presentan clusters de valores significativamente altos, tanto positivos como negativos, mientras que en el modelo espacial se descarta algún tipo de agrupamiento.

Tabla 1: Modelo OLS

Resumen modelo OLS					
Variable	Coefficient	StdError	t-Statistic	Probability	VIF
Intercept	-233283,245	38117,3564	-6,120132	0,000000*	-----
NUM_LIN	306115,745	26118,7554	11,720151	0,000000*	1,172695
OCU	14,65796	5,288407	2,771715	0,006145*	1,598478
LINT	1295,41846	497,645297	2,603096	0,009982*	1,036861
EMP	7,073422	3,109245	2,274964	0,024044*	1,83634
Diagnóstico modelo OLS					
Number of Observ.	190				
Number of Variables	5				
Adjusted R <sup>2</sup>	0,557				
R <sup>2</sup>	0,567				
AIC	5157,6892				
Degrees of Freedom:	185				
F-Statistic	60,637	Prob(>F), (4,185) degrees of freedom:		0,000000*	
Wald Statistic	132,434	Prob(>chi-squared), (4) degrees of freedom:		0,000000*	
Koenker (BP) Statistic	16,534	Prob(>chi-squared), (4) degrees of freedom:		0,002380*	
Jarque-Bera Statistic	1403,820	Prob(>chi-squared), (2) degrees of freedom:		0,000000*	

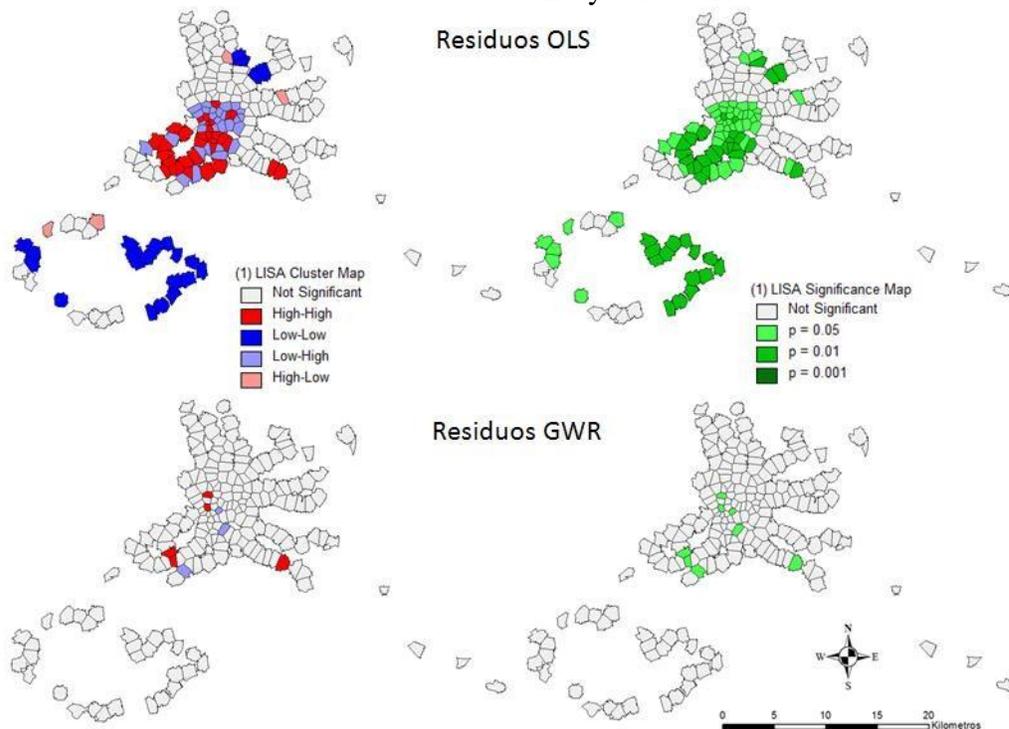
Tabla 2: Parámetros Estimados y Estadísticos de Diagnóstico en ambos modelos

Variables	Modelo OLS	Modelo GWR
Intercepto	-233283,245	-252229,598 / -81629,735
NUM_LIN	306115,745	188054,547 / 317444,429
OCU	14,657960	-3,604863 / 30,026943
EMP_TOT	7,073422	2,740591 / 13,333257
LINT	1295,418463	-833,075699 / 11598,573841
Diagnóstico		
R <sup>2</sup>	0,57	0,73
Adjusted R <sup>2</sup>	0,56	0,70
AICc	5157,68	5099,78
Numero de parámetros	5	24,49
Sigma ( )	35114930078,30	155446,28
Desvío Standard en residuos	184907,89	145031,96

Tabla 3: Test de Moran en los Residuos

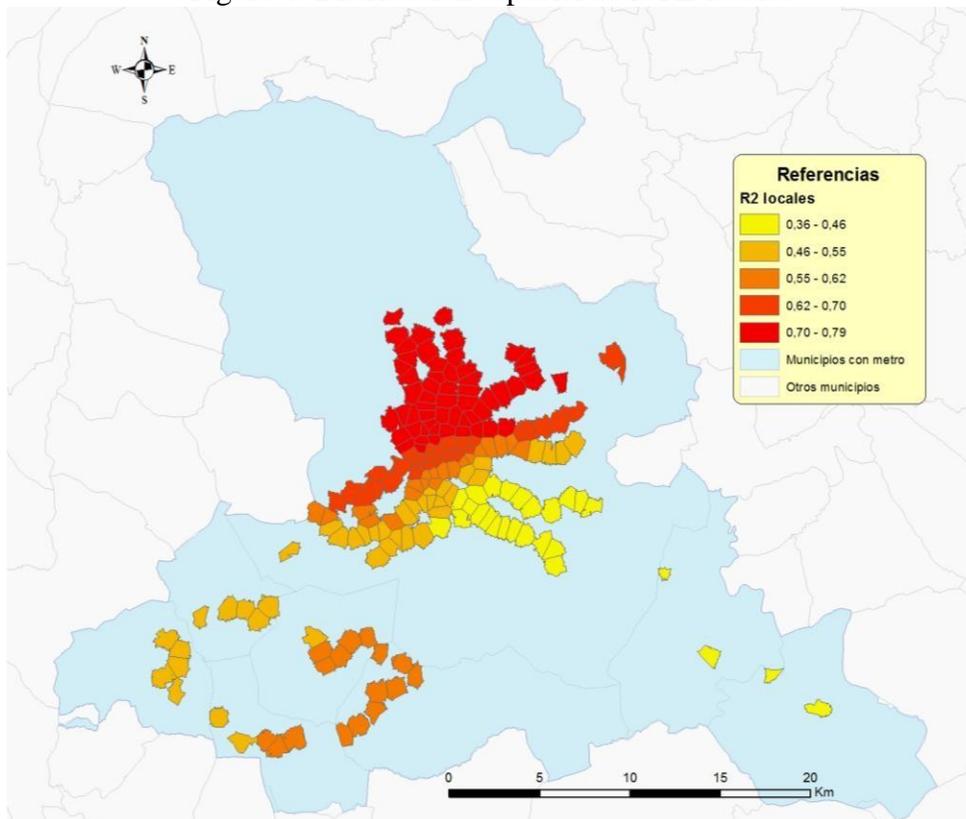
	Método OLS	Método GWR
I calculado	0,022755	-0,020655
I esperado	-0,005291	-0,005291
Varianza	0,000153	0,000149
z-score	2,264975	-1,257582
p-value	0,023514	0,208543

Figura 3: Variación de los clusters espaciales y significación en los errores del modelo tradicional OLS y del GWR



La distribución espacial de los ajustes locales producidos con la GWR, permite conocer la variación espacial del poder explicativo del modelo. La figura 4 muestra que el modelo tiene mejor capacidad predictiva hacia el norte y centro de la red principalmente, mientras que disminuye su precisión en el sur, particularmente en las líneas que salen fuera del municipio de Madrid y dan servicio a municipios vecinos.

Figura 4: Distribución espacial de la R2 locales

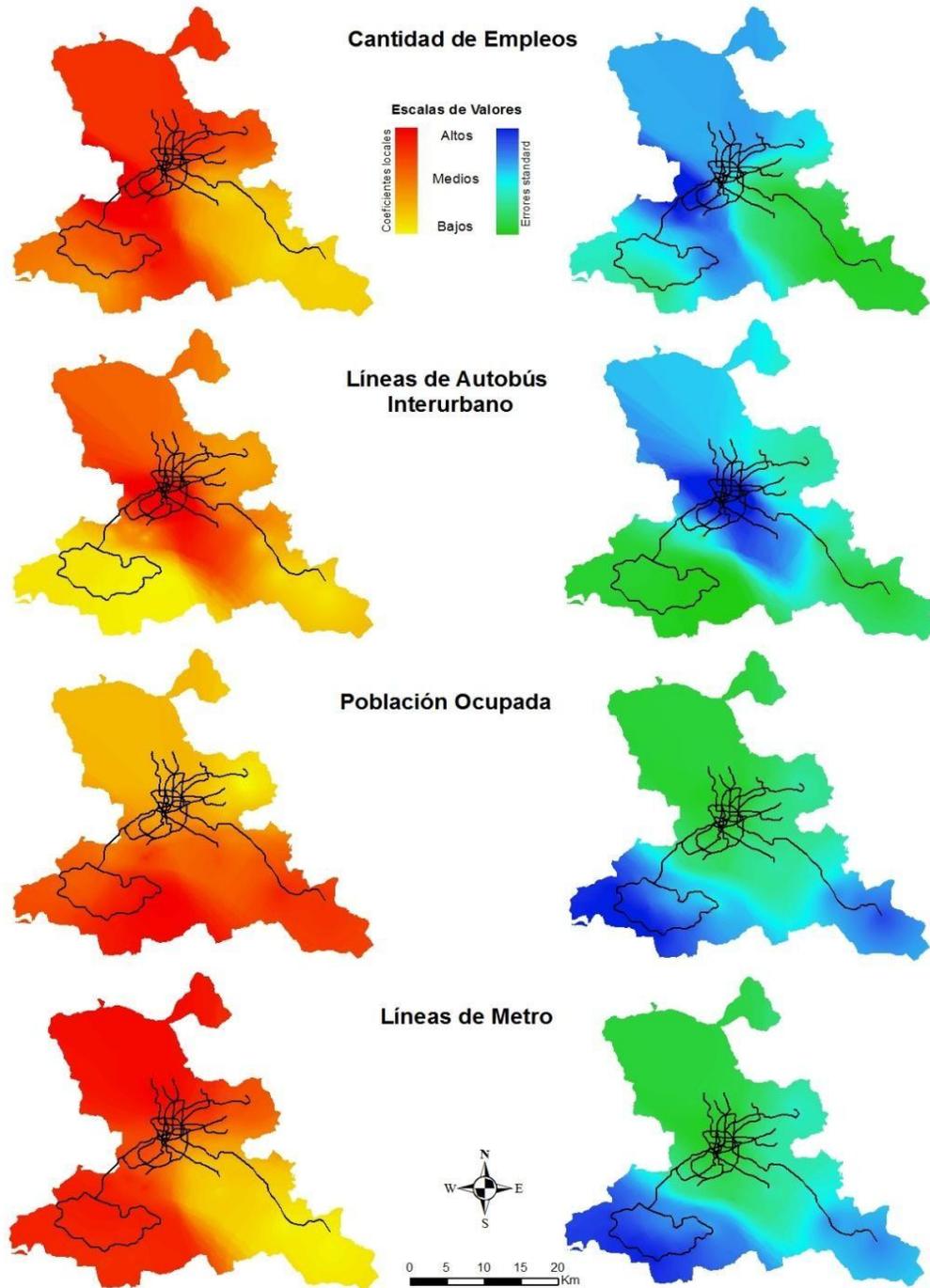


Por último, otra posibilidad que brindan estas regresiones es el análisis de la variabilidad espacial de los coeficientes locales de las variables independientes (elasticidades). Estos valores se han interpolado mediante el método IDW para obtener superficies continuas que ayudan a comprender el patrón espacial de la fuerza de cada variable en el modelo (figura 5). El modelo de regresión múltiple OLS daba una elasticidad de 14,2 para la variable población ocupada. Esto significa que por cada nuevo trabajador residente en el entorno de una estación se esperan 14,2 viajes más en metro en esa estación. Este es un valor constante para toda el área de estudio, ya que el modelo OLS ajusta una única regresión. En cambio con el modelo GWR se puede ver cómo la elasticidad de esa variable varía en el espacio, siendo más alta (más viajes por ocupado) en el sur (estratos sociales bajo y medio-bajo, más dependientes del transporte público) que en el norte (estratos sociales alto y medio-alto, más proclives al uso del transporte privado). Así mismo los coeficientes de la variable empleo son muy elevados en el centro y centro-oeste, donde se concentran los empleos comerciales, administrativos y educativos. Este tipo de empleos presenta ratios de viajes atraídos por empleo más elevados que otros tipos de empleo, como por ejemplo los industriales, lo cual explica que en las áreas donde se concentran el valor de las elasticidades sea más elevado.

Las bajas elasticidades de la variable número de líneas de autobuses interurbanos en las estaciones más periféricas (suroeste y sureste) se explica por el hecho de que en esos espacios los autobuses interurbanos no se comportan como alimentadores de las estaciones de metro (sino más bien como competidores), con lo que el hecho de tener un elevado número de paradas de autobuses urbanos se refleja poco en la demanda de la estación. Además en ambos espacios el

nivel de significación de esta variable es bajo. Finalmente la baja elasticidad de la variable número de líneas de metro en la estación en el sureste se puede explicar por el hecho de que ese área está servida por una sola línea y por lo tanto esta variable se comporta como una constante en ese espacio.

Figura 5: Distribución espacial de los coeficientes locales



## 5 CONSIDERACIONES FINALES

El modelo de cuatro etapas ocupa indudablemente un lugar predominante en la planificación del transporte. Pero los denominados modelos de predicción directa de la demanda de transporte público a nivel de estación presentan algunas ventajas en relación al tradicional modelo de cuatro etapas, como son la respuesta rápida, simplicidad, fácil comprensión de los resultados, menores necesidades de información y bajo coste. Es cierto que carecen de la perspectiva regional del modelo de cuatro etapas y que su capacidad explicativa es menor (Gutiérrez et al., 2011), pero permiten profundizar en el análisis del entorno de las estaciones y evaluar la influencia que los cambios en esos entornos (por ejemplo, nuevos desarrollos residenciales o áreas de actividad económica) tendrán en la demanda del transporte público. Ello gracias a que utilizan los SIG como herramienta para el cálculo de esas variables, mediante operaciones de desagregación espacial que permiten conocer con detalle los entornos de las estaciones (una aproximación de grano fino, en palabras de Cervero, 2006).

La incorporación de técnicas de estadística espacial –medidas de autocorrelación espacial y GWR- en los modelos de predicción directa supone claras ventajas sobre el empleo de métodos estadísticos tradicionales (OLS). En presencia de autocorrelación espacial (lo que es habitual en los datos geográficos) el modelo GWR se comporta de forma más eficiente porque considera las variaciones espaciales de las relaciones entre las variables y obtiene parámetros locales que reflejan más exactamente su influencia en el uso del transporte público. El modelo GWR no sólo suele tener un mayor poder explicativo que el OLS, sino que además permite profundizar en el análisis local, ya que se dispone de una ecuación de regresión para cada observación, ajustada localmente. Así es posible conocer dónde el modelo tiene un mayor o menor ajuste ( $R^2$ ) y cómo cambia la relación entre las variables en el espacio (coeficientes de los predictores) y con qué significación estadística. El conocimiento de la variación de cada coeficiente en el espacio (elasticidades) permite extraer conclusiones más realistas, basadas en el análisis local: operaciones urbanísticas en el entorno de las estaciones podrán producir un mayor o menor número de viajes en función del lugar donde se realicen. Sin duda esto resulta de gran ayuda para los decisores políticos, al poder disponer de predicciones más realistas. Estos modelos no solamente son útiles para conocer los efectos de las políticas de uso del suelo en la demanda de las estaciones, sino que también pueden ser una alternativa razonable al modelo de cuatro etapas para predecir la demanda de nuevas estaciones.

Sin embargo todavía existen muy pocos modelos de predicción directa de demanda a nivel de estación que incorporen la GWR en lugar del método OLS. Este trabajo aporta evidencia de las ventajas de considerar el análisis local en este tipo de modelos. El modelo resultante es sencillo, de respuesta rápida y aporta resultados relevantes. El coeficiente de determinación aumentó de forma dramática con respecto al obtenido con el modelo OLS y el análisis de su distribución espacial y la de las elasticidades aportó información complementaria de gran interés. Estos resultados permiten demostrar algunas de las ventajas del modelo propuesto y aportan argumentos para su utilización en la planificación del transporte.

## Agradecimientos

Agradecemos el financiamiento de la Fundación Carolina (para la realización de la tesis doctoral de Osvaldo Daniel Cardozo) y del proyecto SPILLTRANS (TRA2011-27095, Ministerio de Economía y Competitividad).

## Referencias

- Anselin L. (1988) **Spatial Econometrics: Methods and Models**. Dordrecht, Kluwer Academic Publishers.
- Arbia, G. y Baltagi, B.H. (2009) **Spatial econometrics. Methods and applications**. Heidelberg, Physica-Verlag.
- Blainey, S.P. (2010) Trip end models of local rail demand in England and Wales, **Journal of Transport Geography**, 18(1), 153-165.
- Brunsdon, C.; Fotheringham, A.S. y Charlton, M.E. (1996) Geographically weighted regression: a method for exploring spatial nonstationarity, **Geographical Analysis**, 28(4), 281-298.
- Cervero, R. (2006) Alternative approaches to modeling the travel-demand impacts of smart growth, **Journal of the American Planning Association**, 72(3), 285-295.
- Chasco, C. (2003) **Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales**. Madrid, Consejería de Economía e Innovación.
- Chow, L.F.; Zhao, F.; Min-Tang, L.; Liu, X. y Ubaka, I. (2006) Transit ridership model based on geographically weighted regression, **Journal of the Transportation Research Board**, 1972, 105-114.
- Chu, X. (2004) **Ridership models at the stop level**, National Center for Transit Research, Florida Department of Transportation. Report No. BC137-31
- Clark, S.D. (2007) Estimating local car ownership models, **Journal of Transport Geography**, 15, 184-197.
- Fotheringham, A.S.; Brunsdon, C. y Charlton, M.E. (2002): **Geographically weighted regression: the analysis of spatially varying relationships**. Chichester, UK, Wiley.
- Gutiérrez Puebla, J. y García Palomares, J.C. (2008) Distance-measure impacts on the calculation of transport service areas using GIS, **Environment and Planning B**, 35, 3, 480-503.
- Gutiérrez, J., Cardozo, O. y García-Palomares, J.C. (2011) Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. **Journal of Transport Geography**. 19, 1081–1092.

Hadayeghi, A.; Shalaby, A.S. y Persaud, B.N. (2010) Development of planning level transportation safety tools using geographically weighted poisson regression, **Accident Analysis and Prevention**, 42, 676-688.

Haining, R.P. (2010) The nature of georeferenced data, en Fischer, M.M. and Getis, A. (Ed.): **Handbook of applied spatial analysis. Software tools, methods and applications**. Berlin, Germany, Springer, 197-218.

Kobayashi, T. y Lane, B. (2007) Spatial heterogeneity and transit use, **11th World Conference on Transportation Research**. Berkeley, USA.

Kuby, M.; Barranda, A. y Upchurch, C. (2004) Factors influencing light-rail station boardings in the United States, **Transportation Research A**, 38(3), 223-247.

Lloyd, C.D. y Shuttleworth, I. (2005) Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning, **Environment and Planning A**, 37(1), 81-103.

Marshall, N. y Grady, B. (2006) Sketch Transit Modeling Based on 2000 Census Data, **Journal of the Transportation Research Board**, 1986, 182-189.

McNally, M.G. (2007) The four step model, in Hensher, D.A. y Button, J.K. (Ed.): **Handbook of transport modeling**. 2nd edition. Oxford, UK, Pergamon, 35-52 (816).

Mountain, D.M.; Tsui, J.L.Y y Raper, J.F. (2007) Modelling accessibility via transportation networks based upon previous experience: a geographically weighted regression approach, **Geocomputation 2007**, National University of Ireland, Maynooth, Ireland.

Ortúzar, J.D. y Willumsen, L.G. (2008) **Modelos de Transporte**. Santander, España, PubliCan Ediciones.

Páez, A. (2006) Exploring contextual variations in land use and transport analysis using a probit model with geographical weights, **Journal of Transport Geography**, 14(3), 167-176.

Páez, A.; Le Gallo, J.; Buliung, R.N. y Dall'erba, S. (2010). **Progress in spatial analysis. Methods and applications**. Heidelberg, Germany, Springer.

Rosenshein, L. (2008) Análisis de regresión con datos espaciales, **XV° Conferencia Latinoamericana de Usuarios de ESRI**. Santiago de Chile.

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

Zhao, F. y Park, N. (2004) Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic, **Journal of the Transportation Research Board**, 1879, TRB, 99-107.