



UNIVERSIDAD DE CANTABRIA



*Escuela Técnica Superior de Ingenieros de Caminos,
Canales y Puertos.*

Tesis Doctoral

**Modelos de Generación de viajes Considerando
Relaciones Espaciales**

Autor:

Alexandre Ayayi Amavi Dossavi

Directores de Tesis:

Dr. Borja Alonso Oreña

Santander, 2016

Amédédjisso Thérèse (Mi querida abuela)

RESUMEN

En la presente Tesis Doctoral se especifican modelos de Generación de viajes considerando relaciones espaciales. En primer lugar se ha especifica un modelo de regresión lineal múltiple (MLR), a partir de datos zonales. Estos modelos se comparan viendo las hipótesis de cada uno y los ajustes que reportan. A su vez, se plantean modelos de Generación teniendo en cuenta la correlación espacial, analizándose la mejora de estos nuevos modelos frente a los modelos que no consideran correlación espacial.

En la actualidad, el mundo es consciente de la importancia del transporte en nuestra vida cotidiana. Desde ese punto de vista y centrando nuestro objetivo en la ciudad de Santander, localizada en la provincia de Cantabria y perteneciente ésta a España, se propone se ha propuesto en esta tesis un trabajo consistente, primero, en la elaboración de unos modelos de regresión lineal múltiple de dos tipologías, modelos de atracción/producción para distintas categorías de viajes (totales, basados en el hogar motivo trabajo, basados en el hogar motivo otros y no basados en el hogar) utilizándose datos zonales y otros modelos de producción para dos categorías de viajes (basados en el hogar motivo trabajo y otros) mediante el uso de datos por hogares. También se realiza un completo análisis individualizado y agregado de los datos obtenidos mediante la representación de los mismos en unos planos donde aparecen las distintas zonas pertenecientes al área de estudio, así como la comparación de los resultados obtenidos en ciertas categorías de viajes, utilizando en este caso además de planos, el concepto de porcentaje de error. Las conclusiones obtenidas son muy extensas y están desarrolladas a lo largo de este trabajo, sin embargo se puede destacar como los modelos por zonas obtienen mejores estimaciones que los elaborados con datos por hogares, sobre todo en el caso de viajes producidos basados en el hogar motivos otros, o que todos los modelos están del lado de la seguridad en la estimación de viajes ya que tienden más a la sobre estimación que a la sub estimación.

Se especifica un modelo de correlación espacial global que realiza un examen conjunto de todas las unidades que componen la muestra para determinar si las unidades espaciales se encuentran distribuidas aleatoriamente o si, por el contrario, lo hacen conforme a un patrón determinado. Para considerar situaciones de aglomeración (clusters) en un área determinada se especifica un modelo de correlación espacial local que permite medir la auto-correlación espacial para la ubicación de cada observación. Se

compara la bondad de ajuste de los distintos modelos y se contrastan una serie de hipótesis metodológicas sobre el rendimiento de las especificaciones que consideran la existencia de relaciones espaciales entre observaciones.

Los modelos son aplicados al área metropolitana de Santander (España) con el objetivo de obtener modelos avanzados de Generación de viajes en dicha ciudad. Para el caso de Santander, los modelos que consideran la dependencia espacial entre observaciones obtienen una mejora sustancial en el ajuste frente a los modelos MLR.

ABSTRACT

This doctoral thesis specifies trip generation models considering spatial relationships. Firstly, a multiple linear regression (MLR) model is specified using zonal data. These models are compared evaluating the hypotheses of each one along with their fits. Trip generation models considering spatial correlation are also proposed and the improvements found using these new models are analysed and compared to those which do not consider spatial correlation.

The world is currently aware of the importance played by transport in everyday life. From this point of view and concentrating on the city of Santander, in the province of Cantabria, Spain, this thesis proposes the design of two types of multiple linear regression models, attraction/production models for different trip categories (overall, household based work related, household based, other reasons and not household based) using zonal data and other trip production models for two trip categories (household based for reasons of work and others) using household data. A thorough individualized and aggregated analysis is also carried out on the data obtained through their representation on plans showing the different zones located in the study area, as well as the comparison of the results obtained in certain trip categories, in this case using the concept of error percentages as well as plans. The conclusions drawn are extensive and developed throughout the work, however, it can be noted that zonal models obtain better estimations than those elaborated using household data, above all in the case of households based trips for other reasons, or that all the models are on the safe side in their trip estimations as they tend more on the side of over estimation than under estimation.

An overall spatial correlation model is specified for a group examination of all the units making up the sample in order to determine if the spatial units are randomly distributed or, on the contrary, follow a determined pattern. In order to consider cluster situations in a determined area a local spatial correlation model is estimated to measure the spatial auto correlation for the location of each observation. The goodness of fit of different models is compared and a series of methodological hypotheses are checked based on the performance of the specifications which consider the existence of spatial relationships between observations.

The models have been applied to the metropolitan area of Santander (Spain) with the goal of obtaining advanced trip generation models for this location. In the case of Santander, the models considering spatial dependency between observations show a substantial improvement in fit compared to the MLR models.

AGRADECIMIENTOS

Quisiera dedicar estas primeras palabras de afecto y de gratitud a mi director de tesis, D. Borja Alonso Oreña, por haberme aceptado para realizar esta tesis bajo su dirección brindándome toda su confianza a nivel científico, y a D. Rubén Cordera Piñera por su apoyo tanto profesional como personal y su capacidad no solamente en el desarrollo de esta tesis, sino también por su paciencia.

Mis más sinceros agradecimientos a D. Luigi dell'Olio por su apoyo, su paciencia y su ayuda, ha contribuido desde principio para que el trabajo desarrollado en esta investigación fuera de lo más gratificante y llegar este día a su fin.

De manera especial quisiera agradecer a D. Ángel Ibeas Portilla quien me abrió las puertas de este Departamento de Transportes, y por su invaluable apoyo desde principio, sin usted, nada de esto hubiese sido posible.

Quisiera, sin duda ninguna, agradecer a la gente con quien poseo una deuda intelectual y que ha influido de manera directa e indirecta en la elaboración de esta investigación como: D. José Luis Moura, D. Juan Pablo Romero, D. Roberto Sañudo, D. César Canales y a todos mis compañeros del Grupo de Investigación de Sistema de Transporte de la Universidad de Cantabria por su colaboración y apoyo, con quienes he pasado muchas horas a lo largo de esta tesis para que todo saliera adelante.

Mi agradecimiento también a mis amigos y amigas por todo lo que hemos compartido, las alegrías y penas de esta investigación y, siempre he tenido una respuesta de amabilidad como una sonrisa, un abrazo o un gesto que me lleva hasta sentirme feliz.

Finalmente quisiera extender mi gratitud a mi familia por su apoyo incondicional, por su comprensión y cariño. A mis padres, a mis hermanos, que me han animado a perseguir mis sueños y mis ilusiones.

GRACIAS.

ÍNDICE DE CONTENIDO

RESUMEN.....	v
ABSTRACT	vii
AGRADECIMIENTOS	ix
ÍNDICE DE CONTENIDO	xi
ÍNDICE DE TABLAS.....	xvii
ÍNDICE DE FIGURAS.....	xix
1 INTRODUCCIÓN Y OBJETIVOS	20
1.1 Antecedentes	20
1.2 Objetivos	23
1.3 Estructura de la tesis.....	23
1.4 Aportaciones	24
2 ESTADO DEL ARTE	28
2.1 Introducción a los modelos de generación de viajes	28
2.1.1 Los modelos y su función	30
2.1.2 Fundamentos de la generación de viajes.....	31
2.1.3 Variables que explican la generación de viajes.....	35
2.1.4 Clasificación de viajes.....	37
<i>Por tipo de persona</i>	37
<i>Por propósito del viaje</i>	37
<i>Para cada propósito y/o por hora</i>	38
2.1.5 Factores que afectan a la generación de viajes	38
<i>Producciones de viajes de personas</i>	38
<i>Atracciones de viajes de personas</i>	39
<i>Producciones y atracciones de viajes de mercancías</i>	39
2.2 Modelos Econométricos.....	39
2.2.1 Análisis de Modelos.....	40
2.2.2 Modelo Truncado de Regresión Lineal Múltiple (Maddala, 1983; Cotrus et al, 2005)	41
2.2.3 Escala de análisis de regresión	42
<i>Regresión lineal múltiple a nivel zonal</i>	42
<i>Regresión a nivel de hogares</i>	43
2.3 La econometría espacial (Baronio; Vianco; Rabanal; 2012).....	44
2.3.1 Modelos Regresión Lineal Múltiple con Correlación Espacial.....	46

2.3.2	El análisis espacial	50
2.3.2.1	La auto-correlación espacial o dependencia espacial	53
2.3.2.2	La heterogeneidad espacial.....	57
2.3.2.3	Especificación de un modelo con heterogeneidad espacial	59
2.3.3	Especificación de los efectos espaciales en modelos de regresión	60
2.3.4	Modelización en presencia de dependencia espacial.....	61
2.3.5	Fuentes de auto-correlación espacial	62
2.4	El análisis exploratorio de datos espaciales (AEDE)	62
2.4.1	Análisis exploratorio de datos.....	64
2.4.2	Técnicas gráficas.....	65
2.4.3	Perspectivas en el análisis de auto-correlación espacial univariante: global y local	68
2.4.4	Medidas de asociación espacial global	69
2.4.5	Medidas de asociación espacial local.....	72
2.4.6	Modelo de regresión lineal múltiple con heterogeneidad (Paez et al, 2005).....	74
2.4.7	Regresión geográficamente ponderada.....	75
2.4.8	Datos espaciales (Chasco, 2003)	76
	Naturaleza de los datos espaciales	76
	Sistemas de Información Geográfica (GIS).....	78
	Fuentes estadísticas de información micro-territorial.....	81
	Problemas en el tratamiento de datos espaciales	82
2.5	Análisis y cortes temporales de los modelos.....	86
2.5.1	Períodos de análisis.....	88
3	METODOLOGÍA	94
3.1	Modelos de producción de viajes	94
3.1.1	Generación de viajes basados en el hogar de ida (bhi).....	96
3.1.2	Generación de viajes basados en el hogar de retorno (bhr)	97
3.1.3	Generación de viajes no basados en el hogar (nbh)	98
3.2	Modelos de Atracción de Viajes	98
3.3	Modelo de Regresión Lineal Múltiple (RLM)	102
3.4	Los modelos espaciales	104
3.4.1	Modelo espacial lag.....	105
3.4.2	Modelo espacial error	106
3.4.3	Método de estimación	107

3.5	La matriz de pesos espaciales.....	109
3.5.1	Criterios habituales de contigüidad (Baronio; Vianco; Rabanal; 2012)	113
3.5.2	Otras especificaciones para la matriz de contactos.....	114
3.5.3	Matriz de pesos como instrumento para recoger las interdependencias	116
4	CASO DE ESTUDIO	120
4.1	Descripción del área de estudio (ciudad de Santander).....	120
4.2	Análisis y descripción de la base de datos	125
4.2.1	Análisis de los datos de partida.....	125
4.2.2	Zonificación del área de estudio	126
4.2.3	Descripción de las variables	130
5	ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS	134
5.1	Los resultados de weights (pesos) creados.....	136
5.2	Modelos de viajes estimados por zonas	137
5.2.1	Viajes producidos totales “PROD_TTRIPS” (estimación por zonas).....	137
	• Resultados de los modelos de viajes producidos totales “PROD_TTRIPS” (estimación por zonas).....	138
5.2.2	Viajes producidos basados en el hogar motivo trabajo (estimación por zonas) 141	
	• Resultados de los modelos de viajes producidos basados en el hogar motivo trabajo “PROD_HBW” (estimación por zonas)	142
5.2.3	Viajes producidos basados en el hogar por otros motivos (estimación por zonas) 144	
	• Resultados de los modelos de viajes producidos basados en el hogar por otros motivos “PROD_HBO” (estimación por zonas)	145
5.2.4	Viajes producidos no basados en el hogar (estimación por zonas)	147
	• Resultados de los modelos de viajes producidos no basados en el hogar “PROD_NHB” (estimación por zonas)	148
5.2.5	Viajes producidos (estimación por zonas).....	150
5.2.6	Viajes atraídos totales (estimación por zonas)	151
	• Resultados de los modelos de viajes atraídos totales “ATTRAC_TTRIPS” (estimación por zonas).....	152
5.2.7	Viajes atraídos basados en el hogar motivo trabajo (estimación por zonas) ...	154
	• Resultados de los modelos de viajes atraídos basados en el hogar motivo trabajo “ATTRAC_HBW” (estimación por zonas).....	155
5.2.8	Viajes atraídos basados en el hogar motivo otros (estimación por zonas).....	157

•	Resultados de los Modelos de viajes atraídos basados en el hogar por otros motivos “ATTRAC_ HBO” (estimación por zonas).....	158
5.2.9	Viajes atraídos no basados en el hogar (estimación por zonas)	160
•	Resultados de los modelos de Viajes atraídos no basados en el hogar “ATTRAC_ NHB” (estimación por zonas).....	161
5.2.10	Viajes atraídos (estimación por zonas)	162
5.3	Modelos de viajes estimados por hogares	163
5.3.1	Viajes producidos basados en el hogar motivo trabajo (estimación por hogares) 163	
•	Resultados de los modelos de viajes producidos basados en el hogar motivo trabajo “PROD_ HBW” (estimación por hogares).....	164
5.3.2	Viajes producidos basados en el hogar motivo otros (estimación por hogares) 166	
•	Resultados de los modelos de viajes producidos basados en el hogar motivos otros “HBO” (estimación por hogares).....	168
5.3.3	Viajes producidos basados en el hogar (estimación por hogares).....	170
6	CONCLUSIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN	172
6.1	Conclusiones	172
6.2	Líneas futuras de investigación	173
7	REFERENCIAS	176
8	ANEXOS	184
8.1	MODELOS DE PRODUCCIÓN DE VIAJES TOTALES (ESTIMACIÓN POR ZONAS).....	185
8.2	MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR MOTIVO TRABAJO (ESTIMACIÓN POR ZONAS).....	186
8.3	MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR OTROS MOTIVOS (ESTIMACIÓN POR ZONAS)	188
8.4	MODELOS DE VIAJES PRODUCIDOS NO BASADOS EN EL HOGAR (ESTIMACIÓN POR ZONAS)	190
8.5	MODELOS DE VIAJES ATRAÍDOS TOTALES (ESTIMACIÓN POR ZONAS) 192	
8.6	MODELOS DE VIAJES ATRAÍDOS BASADOS EN EL HOGAR POR EL MOTIVO TRABAJO (ESTIMACIÓN POR ZONAS).....	194
8.7	MODELOS DE VIAJES ATRAÍDOS BASADOS EN EL HOGAR POR OTROS MOTIVOS (ESTIMACIÓN POR ZONAS)	195
8.8	MODELOS DE VIAJES ATRAÍDOS NO BASADOS EN EL HOGAR (ESTIMACIÓN POR ZONAS)	196

8.9	MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR EL MOTIVO TRABAJO (ESTIMACIÓN POR HOGARES)	197
8.10	MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR OTROS MOTIVOS (ESTIMACIÓN POR HOGARES).....	201

ÍNDICE DE TABLAS

Tabla 1: Principales contrastes de auto-correlación espacial univariante – Fuente: Chasco, 2003	69
Tabla 2: Interpretación de los valores estandarizados de los estadísticos de	71
Tabla 3: Diversas especificaciones de la matriz de pesos espaciales – Fuente: (Chasco, 2003)	116
Tabla 4: Zonas de especiales características	128
Tabla 5: Resumen del modelo “PROD_TTRIPS_1” Viajes producidos totales (Estimación por zonas)	138
Tabla 6: The chi-square value for different probability level (0,01-0,99) for degrees of freedom ramging from 1 to 60	139
Tabla 7: Resumen del modelo “PROD_TTRIPS_2” Viajes producidos totales (Estimación por zonas)	140
Tabla 8: Resumen del modelo “PROD_HBW_1” Viajes producidos basados en el hogar motivo trabajo (Estimación por zonas)	142
Tabla 9: Resumen del modelo “PROD_HBW_2” Viajes producidos basados en el hogar motivo trabajo (Estimación por zonas).....	143
Tabla 10: Resumen del modelo “PROD_HBO_1” Viajes producidos basados en el hogar por otros motivos (Estimación por zonas).....	145
Tabla 11: Resumen del modelo “PROD_HBO_2” Viajes producidos basados en el hogar por otros motivos (Estimación por zonas).....	146
Tabla 12: Resumen del modelo “PROD_NHB_1” Viajes producidos no basados en el hogar (Estimación por zonas)	148
Tabla 13: Resumen del modelo “PROD_NHB_2” Viajes producidos no basados en el hogar (Estimación por zonas)	149
Tabla 14: Resumen del modelo “ATTRAC_TTRIPS_1” Viajes atraídos totales (Estimación por zonas)	152
Tabla 15: Resumen del modelo “ATTRAC_TTRIPS_2” Viajes atraídos totales (Estimación por zonas)	153
Tabla 16: Resumen del modelo “ATTRAC_HBW_1” Viajes atraídos basados en el hogar motivo trabajo (Estimación por zonas).....	155
Tabla 17: Resumen del modelo “ATTRAC_HBW_2” Viajes atraídos basados en el hogar motivo trabajo (Estimación por zonas).....	156
Tabla 18: Resumen del modelo “ATTRAC_HBO_1” Viajes atraídos basados en el hogar por otros motivos (Estimación por zonas)	158
Tabla 19: Resumen del modelo “ATTRAC_HBO_2” Viajes atraídos basados en el hogar por otros motivos” (Estimación por zonas)	159
Tabla 20: Resumen del modelo “ATTRAC_NHB_1” Viajes atraídos no basados en el hogar (Estimación por zonas)	161
Tabla 21: Resumen del modelo 1 - Viajes producidos basados en el hogar motivo trabajo “PROD_HBW_1” (estimación por hogares).....	165

Tabla 22: Resumen del modelo 2 - Viajes producidos basados en el hogar motivo trabajo "PROD_HBW_2" (estimación por hogares).....	165
Tabla 23: Resumen del modelo 1 - Viajes producidos basados en el hogar motivos otros "HBO" (estimación por hogares)	168
Tabla 24: Resumen del modelo 2 - Viajes producidos basados en el hogar motivos otros "HBO" (estimación por hogares)	169

ÍNDICE DE FIGURAS

Figura 1: Producción y atracción de viajes. Fuente: Ortuzar	32
Figura 2: Ejemplo de Zonas que Definen Correlación Espacial – Fuente: MIDEPLAN – FDC, 2010	47
Figura 3: Estrategias espaciales de contigüidad – Fuente: Baronio; Vianco; Rabanal; 2012	51
Figura 4: Autocorrelación espacial positiva: valor de las propiedades en una	56
Figura 5: Auto-correlación espacial negativa – Fuente: (Baronio; Vianco; Rabanal; 2012).....	56
Figura 7: Ejemplos de heterogeneidad espacial en situaciones centro-periferia (izquierda) y norte-Sur (derecha) – Fuente: Chasco, 2003	58
Figura 8: Ejemplo de objetos espaciales en el municipio de Santander	77
Figura 9: Ejemplo de datos vectoriales (izquierda) y raster (derecha) – Fuente: Chasco, 2003.	80
Figura 10: Ilustración de una matriz de pesos espaciales – Fuente: Baronio; Vianco; Rabanal; 2012.....	111
Figura 11: Criterios de contigüidad en la matriz de pesos espaciales. Retardo espacial de primer orden – Fuente: (Baronio; Vianco; Rabanal; 2012).....	113
Figura 12: Criterios de contigüidad en la matriz de pesos espaciales. Retardo espacial de segundo orden – Fuente: Baronio; Vianco; Rabanal; 2012.....	113
Figura 12: Situación del término municipal de Santander – Fuente: PGOU de Santander.....	121
Figura 13: Evolución del municipio de Santander (2001 – 2014) – Fuente: Padrón municipal. Instituto Nacional de Estadística (INE).....	123
Figura 14: Plano de la red de transporte (Incluyendo líneas de ALSA) – Fuente TUS.....	124
Figura 15: Plano de la red de transporte TUS (Excepto líneas de ALSA).....	125
Figura 16: Plano del área de estudio con la numeración de cada una de las zonas	129
Figura 17: Variables y categorías de viaje por zona	130
Figura 18: Variables y categorías de viaje por hogar	131
Figura 19: Construcción del Polígono de Thiessen con la ayuda de GeoDa.....	136
Figura 20: Plano estimación viajes producidos totales (estimación por zonas)	137
Figura 21: Plano viajes producidos basados en el hogar motivo trabajo (estimación por zonas)	141
Figura 22: Plano viajes producidos basados en el hogar por otros motivos (estimación por zonas)	144
Figura 23: Plano de viajes producidos no basados en el hogar (estimación por zonas).....	147
Figura 24: Figura 20 + Figura 21 + Figura 22 + Figura 23	150
Figura 25: Planos viajes atraídos totales (estimación por zonas)	151
Figura 26: Viajes atraídos basados en el hogar motivo trabajo (estimación por zonas)	154
Figura 27: Viajes atraídos basados en el hogar motivo otros (estimación por zonas).....	157
Figura 28: Viajes atraídos no basados en el hogar (estimación por zonas)	160
Figura 29: Figura 25 + Figura 26 + Figura 27 + Figura 28.....	162
Figura 30: Viajes producidos basados en el hogar motivo trabajo (estimación por hogares)..	163
Figura 31: Viajes producidos basados en el hogar motivo otros (estimación por hogares).....	167
Figura 32: Figura 30 + Figura 31	170



CAPÍTULO 1

Introducción y objetivos



1 INTRODUCCIÓN Y OBJETIVOS

1.1 Antecedentes

Los cambios acontecidos en las últimas décadas en el mundo han influenciado, como no podía ser de otra forma, al transporte. Esto ha hecho que muchos de los problemas del pasado no sólo se perpetúen sino que alguno de ellos ha incrementado su influencia. Algunos de ellos son, por ejemplo, la contaminación, los accidentes o el déficit financiero que desarrollan las empresas de transporte público, erigiéndose como nuevos desafíos que es necesario resolver para alcanzar un entorno de movilidad sostenible.

Nuestra historia proviene de un periodo en el que no se ha utilizado o si se ha hecho, sólo de forma escasa, la planificación, la inversión, el hincapié en el corto plazo, es decir, la falta de confianza en la modelización y en la toma de decisiones de forma estratégica. Todo esto ha producido una serie de aprendizajes como por ejemplo, que con la simple aplicación de mejores técnicas de gestión de tráfico no se solucionan los problemas básicos anteriormente enumerados, además tienden a reaparecer con más fuerza extendiéndose en zonas más amplias y con formas más complicadas y difíciles de manejar.

Actualmente, dentro del mundo desarrollado, se está invirtiendo más confianza en las propuestas técnicas para resolver los problemas que en décadas pasadas. También



teniendo en cuenta que la tecnología no nos suministra soluciones mágicas para los problemas sociales y económicos. No obstante, la informática y la electrónica debido a sus avances nos han posibilitado toda una serie de nuevas formas de infraestructura del transporte como son los sistemas guiados automáticos y de los sistemas de movimiento con los sistemas de navegación vial y trenes completamente automatizados. Son también de gran interés los avances conseguidos en la informática de bajo coste, tanto el software como sobre todo el hardware. Esto produce la situación de que las limitaciones más importantes las producen el lado humano y técnico.

En el caso de la mayoría de los países en desarrollo también encontramos graves situaciones de congestión y contaminación urbana, además de un problema con el transporte de carácter generalizado. Son problemas no sólo derivados de las conexiones entre núcleos rurales y los mercados sino que también se manifiestan aquellos que son típicos de los países industrializados. Sin embargo, como estos países tienen características diferentes demandan tratamientos distintos debido a sus ingresos bajos, crecimiento rápido, etc.

El principio del siglo XXI está marcado por dos hechos conceptuales de elevada relevancia que afecta a muchos aspectos vitales y, por lo tanto, al avance socio-económico. Primero, nos encontramos con el concepto de globalización y la potenciación de la “sociedad del conocimiento” y después, como soporte, están las telecomunicaciones. Combinando ambas se alcanza la integración en redes más amplias modificando la forma en cómo se abordan algunos de los problemas actuales. Todo esto afecta al mundo del transporte y alguna de las más importantes influencias es las variaciones importantes en las asignaciones de tráfico a las redes de transporte a través de la agilidad proporcionada por la rápida tramitación de instrucciones mediante el mundo de Internet.

En el último tercio del siglo XX, se puede decir que se alcanzaron importantes avances técnicos en el mundo de la modelización, dentro de los principales centros de investigación. Después todos estos avances han sido perfeccionados e implementados por grupos de expertos, resaltando que algunas de las innovaciones no han ido muchos más allá del ámbito académico. Sin embargo, después de todo esto se puede observar cómo se está creando una mayor comprensión del papel que pueden tener los modelos de transporte en el mundo de la planificación.



Es necesario resaltar finalmente que la modelización del transporte no es la planificación del transporte sino una herramienta que nos ayuda a tomar decisiones sobre esta actividad, aunque en ciertas ocasiones puede adquirir un rol de gran importancia pero no definitivo.

Los modelos de generación de viajes tienen un peso importante en la planificación de viajes ya que son la primera etapa del modelo clásico.

Los modelos sintéticos han sido desarrollados en la hipótesis de que cada viaje tuviera una producción y una atracción como fin. Los modelos esencialmente, ligan o relacionan las generaciones a las atracciones. En el caso de los viajes HB, la generación siempre es el hogar. Sin embargo, el origen de dichos viajes, sólo es el hogar para los viajes hacia el lugar de trabajo (o el lugar de estudio o de compras, etc.), pero en el viaje de regreso el hogar es ahora el destino del viaje (Ortuzar & Willumsen, 2001).

Los modelos de generación de viajes pueden considerar una desagregación más detallada, debido a que los viajes basados en el hogar de ida, retorno o no basados en el hogar, tienen normalmente variables explicativas distintas.

Desde el punto de vista de la demanda de transporte, la característica más relevante del usuario es su nivel socioeconómico. Pero dado que es difícil determinar este nivel para cada usuario en particular, en lugar de clasificar a los individuos normalmente se categorizan los hogares que habitan. Cada hogar tiene asociado un cierto ingreso familiar y una cierta tasa de motorización, variables que son utilizadas para categorizar los hogares y por extensión, a los individuos que viven en él. Para ello, el número de categorías de demanda es determinante para definir las dimensiones del modelo, sus requerimientos de información y calibración y finalmente, la precisión de sus resultados.

En este estudio se construyeron, entre muchas otras cosas, los modelos para los vectores de generación de viajes para Santander (España), para ello en primer lugar se especifican un modelo de regresión lineal múltiple (MLR), a partir de datos zonales. Se han calibrado diferentes modelos en función de las variables más significativas obtenidas previamente y tras un estudio más profundo del estado de arte que nos permita mejorar la calidad para un buen desarrollo de ésta tesis doctoral.



Estos modelos se comparan viendo las hipótesis de cada uno y los ajustes que reportan. A su vez, se plantean modelos de generación de viajes teniendo en cuenta las correlaciones espaciales (**Spacial Lag** y **Spacial Error**) sin olvidarse del **modelo básico** aplicándose a la regresión con el buen uso del programa **GeoDa** (Anselin and Bera, 1998), analizándose la mejora de estos nuevos modelos frente a los modelos que no consideran la correlación espacial.

1.2 Objetivos

El objetivo principal de esta tesis doctoral es proponer, desarrollar, analizar e implementar una metodología que permita modelizar, calibrar y obtener modelos más precisos de generación de viajes considerando las relaciones espaciales.

En particular, los objetivos específicos más relevantes son los que se detallan a continuación:

- Hacer un análisis de las variables más relevantes que influyen en la generación de viajes y que deben ser tenidas en cuenta a la hora de especificar los modelos.
- Profundizar en la aplicación de los modelos de generación de viajes con relaciones espaciales, para la obtención de modelos más ajustados, teniendo además en cuenta las características socio-económicas.
- Calibrar diferentes modelos tanto **Spacial Lag** como **Spacial Error** teniendo en cuenta las correlaciones espaciales y comparándolos con los mejores modelos obtenidos mediante la regresión lineal múltiple.
- Finalmente, los modelos son aplicados al área metropolitana de Santander (España) con el objetivo de obtener modelos de Generación viajes en dicha ciudad. Para el caso de Santander, los modelos que consideran la dependencia espacial entre observaciones obtienen una mejora sustancial en el ajuste frente a los modelos **MLR**.

1.3 Estructura de la tesis

La presente tesis doctoral se divide en 7 capítulos y un anexo a lo largo de los cuales se desarrollan un conjunto de modelos de generación de viajes considerando relaciones



espaciales que sirven para desarrollar los conocimientos y la aplicación de diferentes metodologías para el estudio de las variables que mejor se ajustan a los modelos tanto para viajes producidos como atraídos de estimación por zonas y por hogares para diferentes propósitos de viajes.

En el presente capítulo se describen los objetivos, los antecedentes y las aportaciones que esta tesis proporciona al estado del conocimiento.

En el capítulo 2 se realiza un profundo análisis del estado del arte, exponiendo el uso de modelos econométricos desde su especificación más básica a la consideración de interacción espacial en sus diversas formas.

El capítulo 3 describe la metodología seguida para la especificación de los modelos de generación de viajes, tanto a nivel zonal como a nivel de hogar, describiendo las posibles interacciones espaciales tratadas en esta tesis.

Seguidamente, en el capítulo 4 se describe el área de estudio sobre la que aplicar esta metodología, para posteriormente en el capítulo 5, exponer los resultados alcanzados así como la discusión de los mismos.

Finalmente, en el capítulo 6 se detallan las principales conclusiones obtenidas y las futuras líneas de investigación propuestas.

1.4 Aportaciones

La principal contribución de este trabajo se encuentra, precisamente, en la aplicación de los métodos de análisis espacial a la comprensión de las diferencias territoriales, especialmente el criterio de contigüidad que se ha aplicado para la obtención de los indicadores de análisis espacial.

Una de las técnicas más usuales para determinar cuáles son los atributos relevantes y la importancia relativa de cada uno de ellos, es el análisis de regresión de corte transversal, sin embargo existen interacciones espaciales que en la mayoría de estos trabajos no son tenidos en cuenta. Es por eso que este trabajo no solo hace los análisis de los efectos marginales, sino que muestra la interacción espacial de los datos, aprovechando la información georeferenciada. La Cuestión espacial propuesta por



Anselin (1988) en la nueva econometría espacial que permite conectar y relacionar los efectos marginales que se obtienen en el análisis hedónico con los efectos espaciales de vecindad como la imitación, los spill-overs, las externalidades, el contagio, etc. La econometría espacial facilita ensayar métodos globales y locales de análisis de autocorrelación espacial en las observaciones que se construyen y las variables que se utilizan. La no inclusión de las interacciones espaciales puede sesgar los resultados y disminuir la confianza en los test de significancia que se utilicen.

En el caso de Santander, se ha utilizado las matrices de vecindad para analizar las posibles relaciones espaciales presentes en los datos.





CAPÍTULO 2

Estado del arte



2 ESTADO DEL ARTE

2.1 Introducción a los modelos de generación de viajes

Lo primero que se necesita para la construcción de vectores de generación de viajes, es la definición del contexto espacial de su aplicación. Esto es independiente de si el tipo de modelo a considerar es agregado o desagregado.

La operación del modelo de transporte requiere como datos de entrada los vectores origen-destino de viajes para cada período de análisis, clasificados por propósitos de viaje y por categorías de demanda (MIDEPLAN – FDC, 2010). La estimación de tales vectores constituye el objetivo de los modelos de generación de viajes

Debiera estimarse un vector de orígenes de forma idealmente y un vector de destinos por cada propósito y categoría de demanda, pero en la práctica la clasificación por categorías de demanda no siempre es posible. Debido que éstas se definen a partir de los niveles de ingreso y tasa de motorización de los hogares, la categorización de los orígenes (producciones de viajes) es fácil de hacer cuando los viajes se originan en el hogar, lo cual es una característica de la mayoría de los viajes en el período punta de la mañana y una proporción importante en el período fuera de punta.



Sin embargo, durante estos mismos períodos la mayoría de los viajes se realizan hacia lugares distintos del hogar, por lo que una eventual categorización de los destinos (atracciones de viajes) resultaría arbitraria en el mejor de los casos.

Teniendo en cuenta lo anterior, el modelo propuesto supone que sólo los orígenes son clasificables por categoría-propósito y los destinos en cambio sólo por propósitos de viaje. Así mismo, el modelo de transporte recibe como datos de entrada un vector Origen por cada categoría de demanda y por cada propósito de viaje (O_i^{pn}); y un vector Destino por cada propósito de viaje, en el que todas las categorías de demanda están agrupadas (D_j^p). Además, debe cumplirse:

$$\sum_i \sum_p \sum_n O_i^{pn} = \sum_j \sum_p D_j^p \quad (1)$$

Donde:

O_i^{pn} = Número de viajes generados en la zona **i**, de la categoría **n**, con propósito **p**.

D_j^p = Número de viajes atraídos por zona **j**, con propósito **p**.

Por metodología, las generaciones de viajes (orígenes) son modeladas independientemente de las atracciones de viajes (destinos), aunque evidentemente sus resultados deben ser consistentes, tal como se desprende de la expresión anterior. Por otra parte, debido al mayor desarrollo conceptual de los modelos de generación de viajes, generalmente el analista tiende a confiar más en sus resultados y por lo tanto, habitualmente se ajustan las atracciones a las generaciones de viajes.

Un procedimiento sencillo de realizar este ajuste, es calcular un factor de corrección para **cada propósito** de viaje, de la siguiente manera:

$$f_p = \frac{\sum_i \sum_n O_i^{pn}}{\sum_j D_j^p} \quad (2)$$

Después se multiplica dicho factor por los componentes del vector de destinos del propósito correspondiente, obteniéndose los valores ajustados:



$$D_j^{p(a)} = f_p \cdot D_j^p \quad (3)$$

Eventualmente, para algún propósito de viaje, la calibración del modelo de atracción podría entregar resultados más confiables que la correspondiente a los modelos de generación. En este caso, es recomendable ajustar la generación a la atracción de viajes, calculando el factor f_{p1} como el cociente entre el total de viajes atraídos y el total de viajes generados en ecuación (4); para luego ajustar los orígenes en ecuación (5):

$$f_{p1} = \frac{\sum_j D_j^p}{\sum_i \sum_n O_i^{pn}} \quad (4)$$

$$O_i^{pn(a)} = f_{p1} \cdot O_i^{pn} \quad (5)$$

Se define dos tipos de modelos utilizables para aclarar la generación de viajes que son: el análisis por categoría y la regresión lineal. Para elegir un modelo u otro, depende de las características de los viajes cuyos orígenes o destinos se pretende aclarar. Si bien los modelos de análisis por categoría son conceptualmente más adecuados, su ámbito de aplicación se reduce básicamente a aquellos viajes originados en el hogar. Por otra parte, aunque los modelos de regresión lineal no son especialmente adecuados para aclarar la generación de viajes, en estos casos como las atracciones de viajes y las generaciones de viajes no son originados en el hogar, suelen ser la única herramienta metodológica disponible para estudiarlos.

2.1.1 Los modelos y su función

En resumida cuenta, un modelo es una representación simplificada de la realidad (sistema de interés). Consistente en una abstracción con el objetivo de obtener una mayor claridad desde el punto conceptual de la realidad, haciendo que tanto su variedad como la complejidad de niveles se reduzcan para que pueda ser comprendida y especificada para su posterior análisis. Generalmente, el modelo expresa de una forma sencilla las características más relevantes del fenómeno o situación real.



Considerando lo escrito anteriormente, para establecer una diferencia de los modelos se debe tener en cuenta estos dos tipos: los abstractos y los físicos. En el primer caso podemos encontrar por ejemplo las maquetas arquitectónicas o los túneles de viento que están claramente limitados por las características del diseño. En el segundo grupo existen desde los modelos mentales, que son los que utilizamos en la vida diaria, hasta las representaciones formales y abstractas (comúnmente analíticas) de una cierta teoría de un sistema de interés que se esté tratando y de su funcionamiento (modelos abstractos). Los modelos mentales tienen una función muy importante no sólo para comprender e interpretar el mundo real sino en el análisis de los modelos analíticos utilizados en la planificación del transporte. Su implementación está basada en la participación en discusiones, nuevos conocimientos y, ante todo, la experiencia; de ahí que sea de especial complejidad la comunicación y validación de los mismos.

Los modelos que se presentan en ésta tesis doctoral son de tipo abstracto, más concretamente son modelos matemáticos que tienen como objetivo replicar el sistema de interés y su comportamiento con una serie de ecuaciones matemáticas que se apoyan en una concreta hipótesis teóricas. Ellos son representaciones simplificadas, pero eso no impide que puedan ser complejos y que necesiten la utilización de cantidades ingentes de datos. Son un elemento muy valioso ya que constituyen “un ámbito común” de discusión en políticas de intervención, dando así la posibilidad de examinar con cierta objetividad, hecho que es requerido por la experiencia. Además los modelos matemáticos aportan otra importante ventaja, a través de su formulación, calibración y uso, se aprende mucho con la experimentación sobre el comportamiento y funcionamiento interno del sistema. Esto a su vez implementa los modelos mentales que pueden producir una más inteligente organización y gestión del sistema de transporte.

Un modelo es sencillamente una representación sensata de la realidad desde una perspectiva particular. Estos solamente son apropiados dependiendo del contexto donde se valla a aplicar. La capacidad de elección y de adaptación de los modelos a un contexto concreto es uno de los elementos más importantes en el equipaje total del planificador.

2.1.2 Fundamentos de la generación de viajes

La generación de viajes es un proceso mediante el cual se cuantifican los viajes realizados por las personas que residen o desarrollan actividad en una determinada área



urbana, o por vehículos relacionados con dicha área (Ortuzar & Willumsen, 2011). Para ellos, se deben determinar los elementos siguientes:

Viaje: representa un desplazamiento en una sola dirección de un punto de origen a un punto de destino (McLeod & Hanks, 1986). Por lo general existe interés por todos los desplazamientos efectuados en vehículo, aunque muchas veces se consideran también los viajes efectuados a pie y se ignoran frecuentemente analizar los viajes efectuados por los niños de edad inferior a los cinco años. En específico un viaje es un desplazamiento unidireccional desde un punto de origen hasta un punto de destino.

- **Viajes basados en el Hogar (HB):** son aquellos que tienen un extremo en el hogar de la persona que realiza el viaje, independientemente de si éste es el origen o el destino del viaje.
- **Viajes No basados en el Hogar (NHB):** son aquellos en los que ni el origen ni el destino del viaje es el hogar.
- **Producciones de viajes:** se definen como el extremo hogar en un viaje HB, o el origen de un viaje NHB.
- **Atracción de viajes:** se define como el extremo no-hogar en un viaje HB o el destino de un viaje NHB.

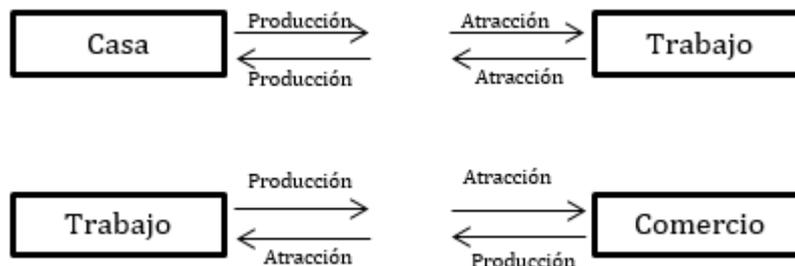


Figura 1: Producción y atracción de viajes. Fuente: Ortuzar & Willumsen, 2011

- **Generación de viajes:** es el número total de viajes, sean HB sean NHB, generados por hogares (Familias) de cierta zona. Éste es el valor que proporcionan la mayor parte de los modelos, y por ende resta el problema de asignar los viajes NHB a otras zonas como producciones de viajes.



La etapa de generación de viajes tiene como objetivo, conseguir una apropiada cuantificación e identificación de los viajes que tienen como límites las diferentes zonas en que ha sido dividida el área en estudio.

Generalmente, los volúmenes de viajes generados son muy complejos de definir y proyectar directamente. La información característica de las zonas puede ser evidentemente más fácil, de gran utilidad para los mismos extremos de viajes proyectados y constituyen lo que se denominan variables explicativas, de tal forma que explican la generación de viajes. Dicha información se refiere al uso del suelo, a las características socioeconómicas de las zonas del área de estudio y a las características del sistema de transporte. Los modelos de generación de viajes están constituidos por relaciones funcionales entre los viajes generados y las variables explicativas, de tal forma que conociendo el valor de las variables explicativas en una perspectiva futura se puede admitir la estimación de la demanda futura de viajes.

Los viajes pueden ser caracterizados por dos atributos que se deberán tener en cuenta durante el proceso de su estudio: el propósito y el horario del viaje.

El **propósito** de viaje está compuesto por estos dos elementos: **base** y **motivo**. Literalmente se entiende como **base**, el sitio en que comienza o termina un viaje, diferenciando entre los viajes basados en el hogar y no basados en el hogar. Naturalmente los viajes basados en el hogar, son aquellos que tienen uno de sus extremos en el hogar de la persona que viaja. Sin embargo, los viajes no basados en el hogar son los que en ninguno de sus extremos se encuentra el hogar. Hablando de **motivo**, se consideran los siguientes: **compra, trabajo, estudio y otros motivos**. Habitualmente en la práctica se consideran los siguientes propósitos de viajes:

- Basados en Hogar-Compras (BHC).
- Basados en Hogar-Trabajo (BHT).
- Basados en Hogar-Estudio (BHE).
- Basados en Hogar-Otros motivos (BHO).
- No Basados en el Hogar (NBH).

Los cinco propósitos enumerados anteriormente no son imprescindibles ser tenidos en cuenta en su totalidad ya que en algunos casos puede ser suficiente considerar



4 propósitos (BHT, BHE o BHC y NBH) y hasta 3 (BHT, BHO y NBH) según el tamaño y las características del área de estudio. Los BHT y NBH deberán estar siempre presentes mientras que los BHE, BHC y BHO pueden mantenerse separados o combinarse según su envergadura relativa. Así, para cada uno de los propósitos considerados se deberá desarrollar un modelo de generación que cuantifique los viajes realizados con ese propósito.

El atributo restante es el **horario** en que se lleva a cabo el viaje. Se distinguen viajes en hora pico (matutino y anochecido) y viajes diarios (total de viajes en el día sin considerar la hora en que se realizan). Los modelos de generación están diseñados para estimar los viajes en hora pico o los viajes diarios. Este último caso es el más común, ya que a partir de los viajes diarios y mediante factores horarios se pueden calcular los volúmenes en cualquier hora del día.

La generación de viajes puede dividirse en dos pasos:

- 1) Determinación del número de viajes originados en cada zona (producciones).
- 2) Determinación del número de viajes destinado a cada zona (atracciones).

Para toda el área de estudio el número de viajes producidos deberá ser obligatoriamente igual al número de viajes atraídos, aunque no necesariamente para una de las zonas en particular. Esto se debe a que por definición, los viajes basados en el hogar (BH) siempre son producidos por la zona que contiene al hogar y atraídos por la zona del otro extremo, en cualquier sentido que se realice el viaje. En cambio, los viajes no basados en el hogar (NBH) son producidos por la zona origen y atraídos por la zona destino.

Por tanto, teniendo en cuenta lo expuesto con anterioridad, cabe deducir que en las zonas principalmente residenciales las producciones serán mayores que las atracciones, mientras que en las zonas mayoritariamente comerciales, industriales o educacionales las atracciones serán mayores que las producciones.

Esta distinción entre los viajes de producción y atracción es importante cuando se utiliza un modelo gravitatorio para la etapa de distribución, aunque no así para los otros métodos de distribución, en los que solamente se tiene en cuenta el origen y el destino de los viajes.



2.1.3 Variables que explican la generación de viajes

- **Uso del suelo**

Con una precisión aceptable, se puede predecir y determinar con facilidad el uso del suelo. Dentro de las variables que explican la generación de viajes se pueden observar tres atributos que influyen en la generación de los mismos. Estos atributos son: tipo, intensidad y ubicación.

Los diferentes **tipos** de usos de suelos tienen distintas características de generación y por ello es importante identificarlos. Los tipos de usos del suelo más usuales suelen ser clasificados de la siguiente forma: residencial, comercial, industrial, educacional y de entretenimiento. El uso del suelo residencial produce más viajes que los otros usos, mientras que los restantes en general son mayores atractores de viajes que productores (Ortuzar & Willumsen, 2001).

La **intensidad** del uso del suelo expresa el nivel de actividad que caracteriza una determinada zona y generalmente se manifiesta en términos de cantidad o de densidad tal como el número total de viviendas en la zona o empleos por unidad de superficie. La intensidad del uso del suelo tiene un poder destacado en el número y tipo de viajes generado en una determinada zona. En general, a menor densidad habitacional mayor número de viajes generados por individuos. Lo mismo ocurre con la distribución por el modo en que se realizan los viajes, las zonas de menor densidad de hogares producen mayor cantidad de viajes en automóvil por vivienda.

La **ubicación** de las actividades se refiere a la distribución espacial de los usos del suelo y de las actividades dentro del área de estudio. Los modos de viajes por habitantes de un barrio de alta densidad, pero rodeado por zonas de baja densidad y alejado del centro de una ciudad, son distintos a los que tendría si ese mismo barrio estuviera próximo al centro.

- **Características socioeconómicas**



Las características socioeconómicas que influyen en la generación de viajes referente a los hogares, son las siguientes: tipo de vivienda, tamaño del hogar, ingreso familiar, posesión de automóvil y actividad de los integrantes del hogar.

La generación de viajes varía según el **tipo de vivienda**. Las viviendas unifamiliares en terrenos únicos generan más viajes por integrantes que las viviendas unifamiliares compartidas y éstas a su vez generan más viajes que las viviendas en edificios de apartamentos. Esta variable no es habitualmente utilizada en estudios a nivel de áreas urbanas sino para la determinación de la generación de viajes de desarrollos urbanos específicos, tales como grandes edificios, barrios cerrados, etc.

El tamaño del hogar: es el número de integrantes del hogar e influye positivamente en la generación de viajes. En otras palabras, la frecuencia de viajes por hogar aumenta con el tamaño del mismo.

Ingreso familiar: es una de las características más importantes en la determinación de la cantidad de viajes por hogar o por persona y la modalidad de los mismos. A mayor ingreso, mayor número de viajes por unidad de tiempo y mayor cantidad de viajes en automóvil.

La posesión de automóviles: está directamente relacionada con el nivel de ingreso familiar y con el tamaño del hogar. Generalmente una familia de menor grado de motorización genera menor frecuencia de viajes.

La actividad de los residentes: influye en la generación de viajes. A mayor número de personas ocupadas por hogar mayor cantidad de viajes generados. La principal influencia la tiene la ocupación del jefe de familia, ya que determina el nivel de ingreso del grupo familiar

- **Sistema de transporte**

El tipo y calidad de las facilidades de transporte disponibles en el área, determina la variable denominada accesibilidad. A mayor accesibilidad, mayor cantidad de viajes realizados. La accesibilidad se define de la siguiente manera:

$$ACC_i = \sum_{j=1}^n (A_j \times F_{ij}) \quad (6)$$



Donde:

ACC_i = accesibilidad de la zona i.

A_j = viajes atraídos por la zona j. Se adoptan las atracciones como de la importancia relativa de cada zona.

F_{ij} = Factor de impedancia entre la zona i y la zona j, a mayor factor mayor accesibilidad.

n = número de zonas

2.1.4 Clasificación de viajes

Los viajes pueden ser clasificados considerando a diferentes criterios, teniendo cada uno de ellos sus propias características y ventajas. Así se puede diferenciar entre:

Por tipo de persona

Las características socio-económicas influyen fundamentalmente en el comportamiento de los viajes de las personas y ésta particularmente forma parte de otra clasificación considerable. Las categorías que se utilizan normalmente son las siguientes:

- Posesión de auto
- Nivel de ingreso
- Tamaño y estructura del hogar

Por propósito del viaje

En la práctica se ha demostrado que se obtienen mejores modelos de generación si se modelizan independientemente los viajes por propósitos o motivos. En el caso de los viajes HB, las siguientes cinco categorías más utilizadas son las siguientes:

- Viajes al trabajo
- Viajes de estudio (al colegio o universidad)
- Viajes de compras
- Viajes sociales y recreacionales
- Viajes por otros motivos



Por general, se llaman viajes obligados de los dos primeros, mientras que se denominan viajes voluntarios (u opcionales) todos los demás. La última categoría confina todos los viajes efectuados por motivos menos frecuentes de forma particular como, por ejemplo: trámites burocráticos, salud (viajes para conseguir un certificado o un pasaporte), acompañar a una persona, etc. Sin embargo los viajes NHB, habitualmente no se subdividen en categorías ya que sólo representan el 15-20% del total de los viajes.

Para cada propósito y/o por hora

Se nota una variación muy importante de acuerdo a la proporción de viajes, por diferentes motivos, según la hora del día y se clasifican en:

- Punta
- Fuera de punta
- Total diario

2.1.5 Factores que afectan a la generación de viajes

En el momento de modelizar la generación de viajes, por lo general no sólo es de interés los viajes que realizan las personas sino que también los viajes de mercancías. Por este motivo se utilizan los modelos para estos cuatro principales grupos: producciones y atracciones, personas y mercancía. Se analizan algunos aspectos principalmente relevantes, dado el resultado de la experiencia de casos reales. Estos son los siguientes.

Producciones de viajes de personas

Se han considerado en muchos estudios prácticos, los aspectos siguientes:

- Posesión de vehículo
- Renta
- Estructura del hogar
- Tamaño del hogar
- Valor del suelo
- Densidad residencial
- Accesibilidad



La posesión de vehículo, la renta, la estructura y el tamaño del hogar son muy utilizados en los estudios sobre generación de viajes por hogares, siendo el valor del suelo y la densidad residencial característicos de los estudios zonales.

Atracciones de viajes de personas

Los factores más utilizados para especificar la atracción de viajes de personas son: el suelo industrial disponible, el comercio y otros servicios. El número de empleados es otro factor importante, sin embargo no se han realizado grandes avances introduciendo alguna medida de accesibilidad.

Producciones y atracciones de viajes de mercancías

Normalmente, la representación de este tipo de viajes es pequeña con respecto al total de viajes de un área determinada (las naciones industrializadas alcanzan como máximo un 20%), aunque pueden influenciar de forma importante en la congestión. Se puede destacar como variables importantes:

- Número de ventas
- Número de empleados
- Superficie total de las empresas
- Superficie cubierta de las empresas

En los estudios de transportes, ni el tipo de empresa ni la accesibilidad han sido considerados como variables explicativas, de acuerdo el conocimiento de los autores; aunque que parece razonable que productos diferentes tengan demandas de transporte diferentes, se nota una particularidad curiosa que no se ha tenido en cuenta el tipo de empresa.

2.2 Modelos Econométricos

Debido a su implementación y sencillez conceptual, la herramienta econométrica más utilizada en la construcción de modelos de generación de viajes corresponde al modelo de regresión lineal múltiple (RLM), (MIDEPLAN – FDC, 2010).

Sin embargo, en los últimos años, se ha extendido el uso de otros tipos de estructuras econométricas para estimar volúmenes y tasas de los viajes, tanto a nivel



agregados (hogares y zonas) como desagregado (individuos) con éxito más o menos considerable.

Como un objetivo principal es estimar el valor de una variable aleatoria (la variable dependiente: por ej.: el rendimiento de un cultivo), siendo que los valores de una o más variables asociadas (la o las variables independientes: por ej.: tipo de suelo, dosis de fertilizante N, elevación, etc.) son conocidos. La fórmula algebraica mediante la cual se determina el valor estimado de la variable dependiente o de respuesta (de forma cuadrática), es la siguiente ecuación de regresión:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i \quad (7)$$

Donde:

Y_i = Valor de la variable dependiente en la observación i

α = Primer parámetro de la ecuación de regresión, que indica el valor de Y cuando $X = 0$

β_1 = Estimador lineal o pendiente de la línea de regresión

β_2 = Estimador cuadrático de la línea de regresión. Es el coeficiente que captura los rendimientos decrecientes, es decir, la tendencia del N a producir menores y menores incrementos de rendimiento con niveles más y más altos de N.

X_i = Valor especificado de la variable independiente en el ensayo u observación i

ε_i = Error de muestreo aleatorio en el ensayo u observación i

2.2.1 Análisis de Modelos

Por lo general, los modelos RLM estiman el número de viajes atraídos por una determinada zona como una función lineal de ciertas variables relacionadas con un equipamiento y usos de suelos de la zona (habitualmente zona en lugares diferentes del hogar). Las variables que más nos llaman la atención son aquellas que habitualmente son recolectadas desde fuentes independientes y que generan actividades agregadas a cada propósito de viaje.



Por lo tanto la típica expresión de un modelo de regresión lineal múltiple para estimar la generación de viajes (y a atracción), a nivel agregado o desagregado, es la siguiente:

$$y_i^p = \beta_k^p + \sum_k \beta_k^p \cdot x_{ik}^p + \varepsilon_i^p \quad (8)$$

Donde $\varepsilon_i^p \sim N(0; \sigma^2)$ y además:

y_i^p : Número de viajes del tipo, propósito o categoría p generado por la unidad i .

La unidad puede ser el individuo, el hogar o la zona.

β : Vector de parámetros de calibración (por tipo de viaje).

x : Vector de variables explicativas (consistente con la unidad de análisis).

ε_i^p : Error de la estimación para la unidad i .

Aunque el uso del modelo de regresión lineal múltiple ha sido amplio, enfrenta una serie de críticas que ha estimulado el desarrollo de nuevos planteamientos.

Como primera crítica, el modelo de regresión lineal múltiple es el que trata como una variable continua a la variable explicativa (viajes), cuando realmente es una variable discreta. Por lo tanto, según el planteamiento de Tanto Long (1997) como Cameron y Trivedi (1998), la especificación de un modelo lineal para datos de conteo (como los viajes) produce resultados sesgados, inconsistentes e ineficientes de sus parámetros. Además surge una segunda crítica para una supuesta naturalidad del término de error, el cual podría generar valores negativos de la variable explicativa (viajes), lo cual es técnicamente imposible.

2.2.2 Modelo Truncado de Regresión Lineal Múltiple (Maddala, 1983; Cotrus et al, 2005)

Se puede truncar la distribución de la variable explicativa (viajes) con el fin de que no se pueda obtener valores inferiores a cero, ya que la misma no puede tomar valores negativos. Partiendo del modelo reseñado, cabe decir que el truncamiento de la variable endógena (o explicativa) en cero, conlleva por lo tanto, que la media del error ya no es cero, por lo que se tendrá que:(MIDEPLAN – FDC, 2010)



$$E(y_i^p / y_i^p \geq 0) = \beta_o^p + \sum_k \beta_k^p \cdot x_{ik}^p + E(\varepsilon_i^p / y_i^p \geq 0) = \beta_o^p + \sum_k \beta_k^p \cdot x_{ik}^p + \sigma \frac{\phi_i}{\Phi_i} \quad (9)$$

Dónde: ϕ_i y Φ_i son, respectivamente, las funciones de densidad y acumulada de una normal estándar evaluada en $\beta_o^p + \sum_k \beta_k^p \cdot x_{ik}^p / \sigma$.

Por el método de máxima verosimilitud o por el método de Heckmann se puede llevar acabo la estimación de este modelo.

2.2.3 Escala de análisis de regresión

Regresión lineal múltiple a nivel zonal

La regresión zonal múltiple se basa en la búsqueda de una relación lineal entre, viajes producidos o bien viajes atraídos por una determinada zona y el valor medio de una serie de características sociodemográficas de los hogares que existen en ella. Cabe destacar las siguientes reflexiones:

- El intercepto. Normalmente, la recta de regresión no es de esperar que pase por el origen. Sin embargo, en ciertas ocasiones, los valores del intercepto resultan bastante grandes si se comparan con la estimación de viajes obtenida del modelo, por lo que en ese caso la recta de regresión puede ser excluida; si por el contrario, el intercepto no es significativo y distinto de cero, puede ser valioso reestimar la regresión, haciendo que pase por el origen.
- Los modelos zonales sólo tienen capacidad para explicar la variación en el comportamiento de viaje entre zonas. Debido a esto, facilitarán resultados adecuados si las variaciones interzonales presentan de forma correcta los motivos reales que desencadenan la variabilidad de los viajes. Para ello, las zonas tienen que tener una composición sociodemográfica homogénea, además de que representen el espectro más amplio posible de requisitos. Sin embargo, las variaciones más relevantes en los datos de los viajes de las personas se obtienen de forma intrazonal.



- Zonas nulas. En determinadas circunstancias algunas zonas no aportan información sobre ciertas variables dependientes (ej: en las zonas no residenciales podrían no existir viajes HB). Entonces, las zonas nulas deben ser rechazadas del análisis, aunque si se tienen en cuenta no deberían afectar mucho en el valor de los coeficientes estimados. Cabe señalar, que a cuantas más zonas nulas tiende a producir estadísticas que sobrevaloran la precisión de la regresión estimada.
- Valores totales versus valores medios. Después de examinar el uso de variables totales o agregadas, tanto el número total de desplazamientos y de valores medios, como el número de viajes por hogar, se deduce que lo que se debe eludir es conjuntar ambas en un único modelo. Para finalizar, aunque se utilicen valores medios, las regresiones zonales siempre están dominadas por la naturaleza y la dimensión de las zonas causado por la agregación espacial. De este modo, al aumentar el tamaño de la zona, disminuye la variabilidad interzonal.

Regresión a nivel de hogares

Las variaciones intrazonales pueden ser aminoradas mediante la reducción del tamaño de las zonas. Es decir, cuantas más idénticas sean estas, se puede producir un aumento de las zonas que conlleva dos consecuencias:

- Los modelos son más costosos desde el punto de vista de recopilación de datos, utilización calibración.
- En los modelos de regresión lineal múltiple, los errores de muestreo son mayores, de los cuales se deduce su inexistencia.

Teniendo en cuenta estos argumentos, es evidente la necesidad de utilizar los modelos que sean independientes a los límites zonales. Por este motivo, el hogar se utiliza como unidad de análisis y en una aplicación basada en este, cada familia es interpretada como un vector de datos de entrada del modelo, con el fin de incorporar toda la variabilidad estudiada desde el punto de vista de las características de la familia y de sus viajes.



La calibración del modelo, como en los de tipo zonal, es un procedimiento que va paso a paso, testeando de forma iterativa cada una de las variables hasta la obtención del mejor modelo.

2.3 La econometría espacial (Baronio; Vianco; Rabanal; 2012)

La econometría espacial es una disciplina de la econometría general que encierra un grupo de técnicas de estimación, contraste, especificación y predicción imprescindibles para el desarrollo de los datos espaciales. Dicho de otro modo, es la parte de la econometría que se destina al estudio de los fenómenos económicos espaciales. No obstante, si bien todas las actividades económicas se desenvuelven en un espacio concreto, los fenómenos económicos espaciales pueden especificarse como aquellos en los que la variable espacio, entendida en sentido amplio y definida luego según algún criterio para alguna matriz de contacto, juega un papel tan destacable que su rechazo podría originar modelos econométricos con graves errores de especificación. La trascendencia del espacio, se expresa frecuentemente en la auto-correlación espacial o/y la heterogeneidad espacial en los modelos de regresión de corte transversal y de datos de panel (Paelink et al, 2004). Desde esta perspectiva, se trata de un concepto semejante a la estadística espacial o a la geoestadística, muy utilizada por geógrafos y físicos (Cressie 1993), aunque la econometría se diferencia de la estadística en general, de la misma manera, la econometría espacial es distinta de la estadística espacial.

Es evidente que la exposición que se ha realizado destaca aspectos conceptuales y prácticos más que de demostraciones matemáticas o de formalización técnica, puesto que nuestro objetivo no es aportar o desarrollar nuevos elementos teóricos a la econometría espacial, sino utilizarla como herramienta para contrastar de forma aplicada aspectos teóricos del crecimiento y los desequilibrios que antes hemos introducido. No obstante, se hará una clara referencia a la literatura que trata con más profundidad y formalidad algunos de los contenidos importantes de esta herramienta.

Se puede considerar cinco principios primordiales que conducen al análisis espacial (Paelinck & Klaassen, 1979):

- Asimetría: refleja la idea de concentración y desconcentración gradual en Áreas diferentes.



- Interdependencia: la dependencia recíproca entre las diferentes unidades de análisis deben ser debidamente incorporadas.
- Topología: Inclusión de variables de distancias entre dos localizaciones, coordenadas, densidades y otras.
- Alotopía: la causa de un fenómeno espacial debe buscarse en otro lugar.
- No linealidad.

De acuerdo con Corrado y Fingleton (2011), aunque muchos economistas se han opuesto a la econometría espacial, por estimar que las variables espaciales se agregan al análisis sólo por revelar significatividad estadística sin ningún argumento teórico, los estudios espaciales dentro de la ámbito principal de la economía han ido integrando variables del tipo espacial. En particular, la econometría espacial ha aportado herramientas apropiadas para el estudio de las externalidades en forma de spillovers (efectos indirectos) espaciales.

Se ha introducido el análisis de la econometría espacial (macroeconomía, economía regional y economía urbana, entre otras) a multitud de secciones de la economía. De acuerdo con esto, uno de los principales fundamentos para el análisis espacial de los fenómenos económicos es la teoría de los centros de desarrollo. Con este fin, la Teoría del Desarrollo Económico y de la denominada Teoría de la Organización Espacial (Posada, 1978), han aportado los elementos más significativos.

Según Anselin (1988) la Econometría Espacial es “la colección de técnicas que se ocupan de las peculiaridades causadas por el espacio en el análisis estadístico de modelos”. Esta tiene cuatro campos de interés:

- a) la especificación formal de efectos espaciales en modelos econométricos
- b) la estimación de modelos que incorporan efectos espaciales
- c) los tests de especificación y de diagnóstico para detectar la presencia de efectos espaciales
- d) la predicción espacial (interpolación).

Del análisis de la teoría anterior, de forma concisa, se puede determinar que la diferencia entre la econometría clásica y la econometría espacial es debido al tratamiento de los dos efectos espaciales, tal y como establece en su definición LeSage (1999): “la



econometría tradicional ha ignorado estos dos elementos que violan los supuestos de Gauss-Markov usados en la modelización de regresiones.”

2.3.1 Modelos Regresión Lineal Múltiple con Correlación Espacial

La definición de la correlación espacial y su integración de forma analítica en los modelos es el paso primordial para plantear y calibrar los diferentes modelos de corte transversal, concretamente en los modelos de generación de viajes a nivel agregado (zonas). (MIDEPLAN – FDC, 2010)

De una manera abreviada, se puede definir la correlación espacial como un impacto que estimule sobre la variable explicativa de un sector geográfico concreto, contiguos o vecinos. Anselin (1988) ofrece un tratamiento íntegro para muchos tópicos de la econometría espacial, integrando enfoques bayesianos aportando máxima verosimilitud.

Así, es posible identificar una matriz W de contigüidad que nos señale si un determinado par de zonas muestra algún grado de correlación espacial. Esta matriz W es de $n \times n$, y sus elementos w_{ij} valdrán 1 si la zona i está correlacionada espacialmente con la zona j , y 0 si no. Se puede definir la correlación espacial por los ejemplos siguientes:

$w_{ij} = 1$ si las zonas i, j son vecinas y 0 si no

$w_{ij} = 1$ si las zonas i, j pertenecen a la misma comuna y 0 si n

$w_{ij} = 1$ si las zonas i, j presentan el mismo uso de suelo y 0 si no

El mismo concepto se puede identificar de otras maneras. Aquellas zonas que dependan del modelador deberán ser empíricamente investigadas para contrastar si son válidas o no. En la Figura 2, se muestra un ejemplo de la zonificación.

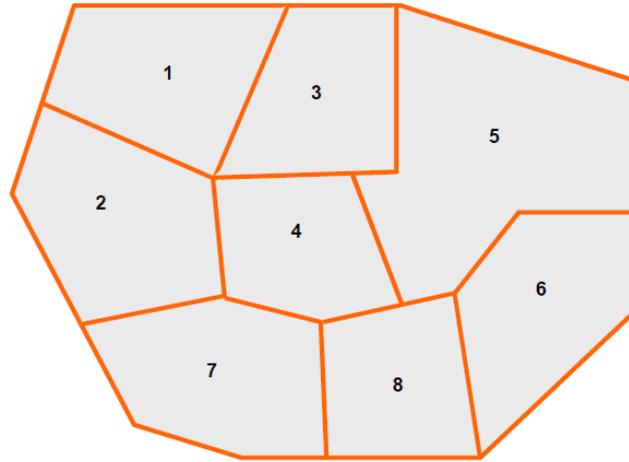


Figura 2: Ejemplo de Zonas que Definen Correlación Espacial – Fuente: MIDEPLAN – FDC, 2010

Si se ha definido la correlación espacial como $w_{ij} = 1$ si las zonas i, j son vecinas y 0 si no, se tendría que la zona 1 está correlacionada con las zonas 2, 3 y 4. De la misma forma, la zona 6 estaría correlacionada con las zonas 5 y 8, y la zona 4 estaría correlacionada con las zonas 1, 2, 3, 5, 7 y 8. Por lo tanto, la matriz W tendría la siguiente forma:

$$W = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (10)$$

A pesar de ello, por convención, se puede usar una estandarización de la matriz W anterior, de tal forma que sus columnas sumen 1:



$$\tilde{W} = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 0 & 1/4 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 1/6 & 0 & 1/6 & 1/6 \\ 0 & 0 & 1/4 & 1/4 & 0 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{bmatrix} \quad (11)$$

Otras formas para definir la matriz de contigüidad W se muestran en Kwizile y Teng (2009), correspondiéndose a la siguiente expresión:

$$W = [W_{ij}] = D_{ij}^n \quad (12)$$

Donde D_{ij} es la distancia entre la zona i y la zona j , y n es un parámetro de calibración. La expresión (12) es una representación continua para la contigüidad de zonas, y representa una medida de accesibilidad entre ambas zonas.

Determinada la matriz W , enlazada con las características espaciales entre las distintas zonas, es posible mostrar los modelos econométricos lineales que integran este fenómeno realización.

La expresión siguiente representa el modelo con correlación espacial más utilizado:

$$Y = \rho W_1 Y + X \beta + \varepsilon \quad (13)$$

$$\varepsilon = \lambda W_2 \varepsilon + u, \quad u \sim N(0; \sigma_u^2 I) \quad (14)$$

Donde:

$Y = [Y_1, Y_2, \dots, Y_n]$ representa el vector con las generaciones (o atracciones) de viajes a nivel zonal.

$X = [X_{k1}, X_{k2}, \dots, X_{kn}]$ es la matriz de variables explicativas

$\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]$ es el vector de perturbaciones correlacionadas espacialmente.

$u = [u_1, u_2, \dots, u_n]$ es el vector de perturbaciones ruido blanco (Independientes, normales e idénticamente distribuidas).



El modelador debe conocer y precisar las matrices de correlación espacial W_1 y W_2 las cuales no deben ser obligatoriamente diferentes. Los parámetros ρ , β y ε deben ser estimados; teniendo en cuenta que β puede ser concretado para cada variable explicativa e incluso para cada zona específica (Anselin, 2003).

Los casos más relevantes de modelos con correlación espacial son los siguientes:

a) Modelo autoregresivo espacial de Primer Orden (FAR)

La utilización de este tipo de modelo consiste sustancialmente en cuantificar los efectos y estructuras de interacciones espaciales entre áreas geográficas, determinando así las clases de fenómenos existentes. La generación o atracción de viajes son dos casos concretos en los que se puede observar determinados fenómenos. Dicho modelo valora que $\mathbf{X} = \mathbf{0}$ y $\mathbf{W}_2 = \mathbf{0}$. Dicho modelo no es práctico para predecir generaciones de viajes en cortes temporales futuros, debido a que no se tiene en cuenta ninguna variable explicativa.

En este modelo se intenta encontrar la fluctuación de las variables \mathbf{Y} (viajes generados en cada zona), sin ninguna variable explicativa, como una sucesión lineal de unidades vecinas.

b) Modelo regresivo-autoregresivo espacial orden mixto (SAR)

Adquiere su nombre ya que conjunta la variable de retardos espaciales (\mathbf{WY}) con la regresión estándar.

Este modelo valora $W_2 = 0$. Trata de encontrar la fluctuación de las variables \mathbf{Y} como una sucesión lineal de unidades vecinas, teniendo en cuenta también las variables exógenas (\mathbf{X}).

A su vez es extensión de la correlación espacial de primer orden del modelo FAR, teniendo en cuenta las variables explicativas \mathbf{X} perteneciente de un modo particular al modelo tradicional.

c) Modelo con correlación espacial en el error (SEM)



Método mediante el cual la estimación de este modelo puede ser elaborada es el Mínimo Cuadrado Ordinario (MCO), debido a que a través de él se crea estimadores insesgados y robustos. No obstante, la aparición de la correlación espacial en el error empuja su ineficiencia.

Ahora bien, para disminuir el problema es imprescindible hallar el valor del parámetro que tendrá que ser diferente de cero para precisar si existe perturbación en la correlación espacial.

La herramienta más utilizada para la estimar en este método es el test t sobre los parámetros de correlación (Anselin, 1988).

d) Modelo espacial de Durbin (SDM)

Recibe este nombre de acuerdo con su semejanza con el modelo de series de tiempo con residuos auto-correlacionados.

Es importante resaltar que el modelo anterior es un caso específico de éste

e) Modelo espacial general (SAC)

Este modelo estima la correlación espacial existente entre la variable endógena y los residuos del mismo.

Como una observación importante: Aunque W_1 y W_2 pueden tener el mismo valor, cuando coinciden sus valores se crea problemas para identificar sus parámetros. Por ello, se determina que $W_1 \neq W_2$.

La función de log-verosimilitud para este modelo puede ser maximizada utilizando la función de log-verosimilitud concentrada o condicional.

2.3.2 El análisis espacial

El análisis espacial es un conjunto de técnicas utilizadas para estudiar y analizar datos espaciales que dan valor a la información que ya se conoce, permitiendo la revisión de los modelos. Concretamente, su enfoque se basa en la localización de los siguientes efectos espaciales: la dependencia y la heterogeneidad espacial.



La econometría clásica, de manera intuitiva, ha incluido el concepto de retardo en el análisis de series temporales (los valores de una variable en el momento t dependen de los valores registrados en momentos anteriores). Dicho avance permitió una notable progresión de los métodos de análisis de las series temporales (Cliff y Ord, 1973). Como consecuencia, se ha dado a conocer la existencia de una posible auto-correlación espacial (los valores de un lugar están correlacionados con los de su entorno) contribuyendo este hecho al desarrollo del análisis espacial moderno.

El indicador Global Moran indica el grado de asociación espacial de una variable para el conjunto del territorio, es decir, la mayor o menor concentración de los valores de esa variable en dicho territorio (Anselin, 1995).

La contigüidad y la distancia son los criterios básicos para las relaciones de vecindad. Sí las unidades comparten los límites fronterizos se consideran vecinas, basándose en la contigüidad. Esta se puede clasificar en diferentes niveles basándose en el número de coronas que rodea la unidad de referencia (primer orden, segundo orden, tercer orden, etc). Existen diferentes maneras de interpretación dependiendo de si comparten un solo punto o un segmento de la frontera, respectivamente, el subcriterio de la 'reina' (**Queen**) y de la 'torre' (**Rook**), por semejanza al tipo de movimientos que pueden hacerse en el ajedrez, Figura 3. Esta figura muestra de manera esquemática y sobre unidades homogéneas las diferencias al aplicar cada subcriterio.

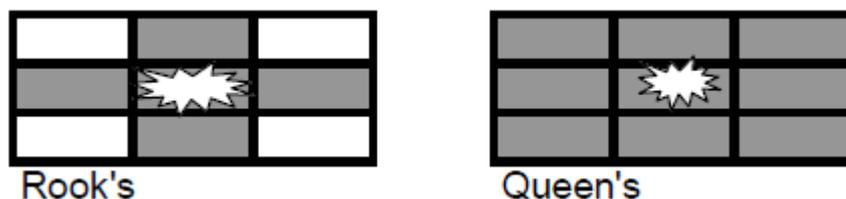


Figura 3: Estrategias espaciales de contigüidad – Fuente: Baronio; Vianco; Rabanal; 2012

El siguiente criterio se centra en la distancia que separa la unidad de referencia con el resto, obteniendo la relación de vecindad. Generalmente la distancia se mide entre centroides. La utilización de este criterio no nos garantiza encontrar unidades vecinas, situación que se soluciona usando el criterio de la contigüidad. Por este motivo, es necesario seleccionar la mínima distancia que nos garantice al menos una unidad vecina.



Las intervenciones que tiene el uso de estos criterios y las determinaciones complementarias que los mismos reivindican, indudablemente, afectan el cálculo de los indicadores de asociación espacial. Por esta razón, es necesario y práctico la utilización del criterio que maximice el nivel de asociación espacial observada y grado de significación. De esta manera, el criterio elegido por sí mismo podría tener un resultado satisfactorio e inmediatamente traducido claramente como una clave de extensión territorial de los clusters observados.

La valoración de la distribución y la ubicación espacial de los procesos sociales se evolucionan favorablemente mediante los métodos estadísticos, No obstante se ha observado unas restricciones considerables tanto en el proceso de desarrollo de los métodos en sí mismos como por la disponibilidad de datos georeferenciados. Los tres problemas más frecuentes en el análisis espacial son:

- a) La tendencia de áreas geográficas acotadas burocráticamente (estados, municipios, localidades) que pueden o no ser las más apropiadas para el fenómeno en cuestión. Es probable, por ejemplo, que los mercados de trabajo no sean bien capturados con estas fronteras administrativas, pero sí lo sean las políticas públicas de empleo que utilizan dichas unidades como referencia de planeación (Voss et al 2005, Messner y Anselin 2002). De cualquier forma, existe el peligro de que el fenómeno poblamiento estudiado no se ajuste a las unidades geográficas en las que la información esté disponible o no ocurra en la escala para la que tenemos datos. Se restringe no sólo el tipo de análisis que se puede emprender sino también las conclusiones que se puede obtener con los métodos anteriormente expuestos. Por ejemplo, es habitual que el Modelo Espacial del Error sea favorecido por los datos cuando existe dicho desajuste entre los datos que se tienen y la “verdadera” dimensión espacial en la que un fenómeno ocurre.
- b) Dado que la mayor parte de los métodos se abocan a modelar datos agregados (por área o polígono) se corre el riesgo de la **falacia ecológica**, es decir de tratar de inferir los comportamientos de los individuos a partir de las características del área en la que viven. Claramente, esto no puede hacerse. Los modelos sólo tienen sentido para entender que pasa en el nivel agregado y tratar de inferir comportamientos “colectivos” (Messner y



Anselin (2002). Es posible, sin embargo, mezclar métodos de análisis espacial con otros métodos que pueden descomponer ambos niveles de análisis, por ejemplo, los modelos multinivel o jerárquicos (Sampson et al 1997).

- c) Más allá del riesgo de falacia ecológica está el problema de cuál puede ser la interpretación substantiva de la autocorrelación o heterogeneidad espacial, particularmente cuando sólo tenemos información para un punto en el tiempo. En el caso de la dependencia espacial, sólo tendremos mayores certezas con datos de más de un punto en el tiempo, ya que con datos de un momento en el tiempo no es posible distinguir cuando un efecto espacial es aparente y cuando existe un efecto real de contagio, aprendizaje o “spillover” (Messner y Anselin 2002). El análisis de la heterogeneidad es un poco más permisivo al respecto, en este caso el mayor reto es identificar el tipo de variación en el espacio que se observa (continua o discreta) y desarrollar explicaciones de porque dicha heterogeneidad está presente.

Frente a estas restricciones, para los estudios poblacionales los métodos de análisis espacial pueden servir de gran apoyo. Teóricamente, existe la posibilidad de desarrollar políticas urbanas que aminoren la conglomeración geográfica.

De acuerdo a los estudios realizados, es posible que dicha conglomeración ayude a señalar de sus habitantes y un desgaste considerable de sus condiciones de vida. Sobre todo, nos sirve de gran ayuda para identificar las variables que más influyentes en la generación de viajes en cada zona (educación, empleo o estructuras familiares, ect).

Las interacciones espaciales más relevantes son las que se describen a continuación:

2.3.2.1 La auto-correlación espacial o dependencia espacial

La dependencia o auto-correlación espacial es uno de los objetivos principales de la econometría espacial ya que desde un principio, surge como consecuencia de la existencia de una relación funcional entre lo que ocurre en un punto determinado del



espacio y lo que ocurre en otro lugar Cliff y Ord (1973); Paelink y Klaassen (1979); Anselin (1988).

Es decir, una variable se encontrará espacialmente auto-correlacionada cuando los valores observados en un lugar determinado dependen, no sólo de ciertos factores externos (otras variables), sino de los valores observados en regiones vecinas. Por ejemplo, Cox (1969) demostró que existía una continuidad geográfica en las intenciones de voto de los norteamericanos, es decir, que si un estado votaba mayoritariamente al partido demócrata, se producía un aumento en la probabilidad de que los estados vecinos también votaran a los demócratas. Otro ejemplo, propuesto por Cliff y Ord (1981) presenta el análisis espacial como forma de conocimiento de las causas y formas de propagación de epidemias y enfermedades.

En este sentido, la dependencia espacial está determinada por una noción de espacio relativo o de localización relativa, que realza el efecto de la **distancia**. Cuando la noción de espacio va más allá del sentido euclídeo estricto, para incluir el “espacio general” de Isard (1969), que tiene en cuenta el espacio político, la distancia interpersonal, las redes sociales, etc., parece claro que la dependencia espacial es un fenómeno con un amplio campo de aplicación en las ciencias sociales.

La auto-correlación espacial implica que el valor de una variable se encuentra condicionado por el valor que esa variable asume en una región vecina. Como se verá más adelante, la vecindad no necesariamente quedará definida como contigüidad física, sino que existen una gran cantidad de criterios para definirla, a partir de una matriz de contactos (Buzai and Baxendale, 2009). Esta debilidad, ha sido una de las principales objeciones metodológicas a la econometría espacial y la robustez de los resultados que con ésta se pueden alcanzar.

La dependencia espacial surge principalmente por dos causas: la existencia de errores de medida y de fenómenos de interacción espacial.

Los errores de medida aparecen, entre otras razones, como consecuencia de una correspondencia limitada entre la extensión espacial del fenómeno que nos interesa y las unidades espaciales de observación. Los índices de auto-correlación espacial permiten vincular en forma conjunta la dependencia entre lugares y valores de variables o atributos



que importan. Al mismo tiempo, resultan apropiados para observar la configuración espacial fraccionada inherente a nuestros tiempos (Chasco, 2003).

Sí desde una perspectiva horizontal, se establece la correlación que una misma variable tiene en diferentes unidades espaciales contiguas, se puede registrar una de las siguientes tres posibilidades:

- **Auto-correlación espacial positiva:** las unidades espaciales vecinas presentan valores próximos. Indica una tendencia al agrupamiento de las unidades espaciales.
- **Auto-correlación espacial negativa:** las unidades espaciales vecinas presentan valores muy disímiles. Indica una tendencia a la dispersión de las unidades espaciales.
- **Sin auto-correlación:** no ocurre ninguna de las dos situaciones anteriores. Por lo tanto, los valores de las unidades espaciales vecinas presentan valores producidos en forma aleatoria.

Se da auto-correlación espacial positiva cuando lo que acontece en un determinado lugar lleva a que se extienda ese mismo fenómeno hacia el resto de lugares que lo rodean, posibilitando así la concentración del mismo. Cuando esto sucede, estaríamos en presencia del llamado efecto contagio o desbordamiento (“spillover”) que se produce en muchos fenómenos socioeconómicos de renta y desarrollo humano. Su presencia en una región sería causa de su extensión a regiones vecinas, favoreciendo la aglomeración del fenómeno en la zona. Dicho de otro modo, cuando la presencia de un fenómeno económico en una determinada unidad, se extienda a las regiones colindantes. Habitualmente, las características propias de cada inmueble, el precio de los mismos depende de factores tales como la ubicación y medida en función a la distancia de algún centro importante, como puede ser una plaza central, la disponibilidad de servicios, la concentración comercial del área en la que se sitúa, etc. De este modo, un inmueble tenderá a tener un valor más o menos similar al inmueble colindante (siempre que no tenga características propias que lo distingan), pero levemente decreciente en la medida que se encuentren más lejos de un punto central (que podrá ser una plaza importante “plaza central”, o algún otro punto relevante de similar importancia). La Figura 4 revela esta idea de forma resumida, puesto que, en una gran ciudad podría existir una gran cantidad de “puntos centrales” formando una cadena de cuadrados.

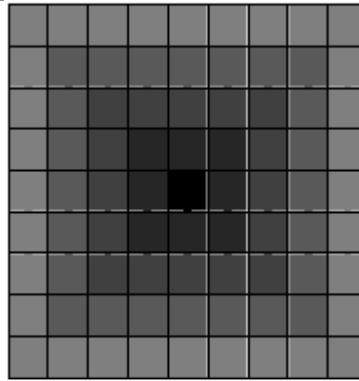


Figura 4: Autocorrelación espacial positiva: valor de las propiedades en una ciudad según la distancia a un punto central – Fuente: Baronio; Vianco; Rabanal; 2012

Se habla, de auto-correlación negativa cuando lo que sucede en un determinado lugar no puede aparecer en los lugares que lo rodean. Es decir, cuando la presencia de un fenómeno, haga imposible o disminuya significativamente las posibilidades de que ese mismo fenómeno pueda ocurrir en las regiones colindantes. Esta situación sería compatible con la existencia de jerarquías, por ejemplo, de tipo centro-periferia, donde una favorable evolución del centro podría verse acompañada de un empeoramiento de las regiones periféricas próximas. Por otro lado, la existencia de fenómenos de interacción espacial, de efectos desbordamiento y de jerarquías espaciales también tiene como consecuencia la aparición de un esquema de auto-correlación espacial. El ejemplo tradicional de auto-correlación negativa o yuxtaposición de valores es la disposición de las casillas blancas y negras en un tablero de ajedrez, como en la Figura 5.

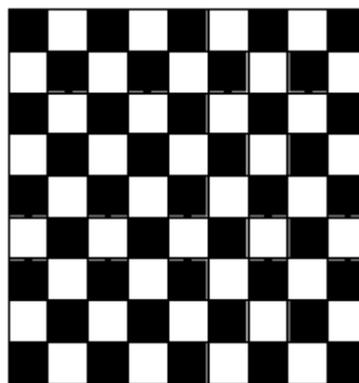


Figura 5: Auto-correlación espacial negativa – Fuente: (Baronio; Vianco; Rabanal; 2012)

La auto-correlación espacial positiva está relacionada con la semejanza de valores y la auto-correlación espacial negativa con la desemejanza de valores cercanos. Es imprescindible verificar que la aparición de un esquema de auto-correlación espacial es consecuencia de la presencia de fenómenos de interacción espacial y no a errores de



medida que haga que las unidades se relacionen. Finalmente, cuando la variable analizada se disponga de forma aleatoria, no estaremos en presencia de auto-correlación espacial.

2.3.2.2 La heterogeneidad espacial

La heterogeneidad espacial constituye la interacción del efecto espacial. Esta muestra la inestabilidad estructural y la heteroscedasticidad en el espacio de los datos en estudio y que pueden cambiar de acuerdo con la localización como el paisaje, la tasa de criminalidad, la contaminación, las cuales tienden a ser distintas en cada zona y se puede resolver con técnicas econométricas existentes para series temporales.

La heterogeneidad espacial, alude al concepto de variación en las relaciones que se establecen para los fenómenos económicos espaciales conforme se modifica el espacio de estudio. Cuestión que se manifiesta fundamentalmente de dos formas: parámetros que admiten distintos valores según se incluyan determinadas zonas o no, esto significa una inestabilidad estructural, o errores surgidos por especificaciones de modelos incorrectos, que pueden ocasionar una heteroscedasticidad.

A pesar de que la dependencia y heterogeneidad espacial son características propias tanto de los datos georreferenciados como de las series temporales, no pueden ser abordadas con los mismos instrumentos estadístico-económicos. La diferencia que existe entre ambos, radica en el tipo de instrumentos necesarios para investigar relaciones que tienen lugar sobre el espacio, dado que los desarrollados en un ámbito de series temporales presentan problemas de adecuación, debido a que la dimensión espacio no es homologable a la dimensión tiempo (Chasco, 2003).

Los contrastes ordinarios de Breusch-Pagan y White, frecuentemente utilizados en el análisis clásico, son también válidos aquí, siempre que no haya alguna otra especificación errónea, como la dependencia espacial o inestabilidad estructural. En estos últimos casos, los contrastes comentados con anterioridad ya no resultarán apropiados y requieren de algunos cambios.

El motivo por el que la contratación de la heterogeneidad se lleve a cabo en ausencia de algún tipo de dependencia espacial, se debe a que los modelos espaciales (con dependencia sustantiva o espacial) incorporan un componente espacial a cambio de alargar la homogeneidad en los efectos espaciales.



Evidentemente, la heterogeneidad espacial aparece cuando se trabaja con unidades espaciales (regiones, países, municipios, secciones censales) en las que un fenómeno se distribuye de manera distinta sobre el espacio, lo que suele suceder con localizaciones del tipo centro-periferia, este-oeste, norte-sur, etc., tal y como se observa en la Figura 6. Por este motivo, el efecto espacial suele estar directamente vinculado con la ubicación geográfica, el área o cualquier otra propiedad de las unidades espaciales (Anselin, 1988, Moreno y Vayá, 2000, Chasco, 2003). Según Anselin (2001a), la heterogeneidad espacial puede ser definida como “inestabilidad estructural en forma de varianza no constante de los residuos de una regresión (heteroscedasticidad) o en los coeficientes del modelo, que es posible abordar mediante técnicas de econometría tradicional o con herramientas propias de econometría espacial”.

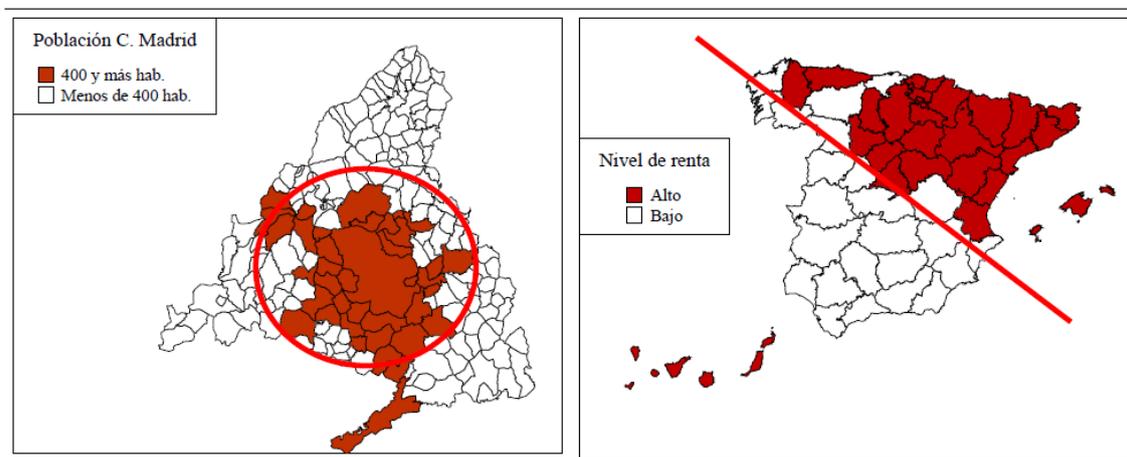


Figura 6: Ejemplos de heterogeneidad espacial en situaciones centro-periferia (izquierda) y norte-Sur (derecha) – Fuente: Chasco, 2003

Debido a que el efecto de heterogeneidad espacial puede ser tratado utilizando técnicas econométricas tradicionales, los estudios sobre econometría espacial dedican una menor atención a este tema que a la problemática en torno a la auto-correlación espacial. Sin embargo, Anselin (2001a) señala, que hay tres razones por las que se debería analizar este efecto de heterogeneidad a través de técnicas propias de econometría espacial:

- En primer lugar, la estructura que subyace en la inestabilidad espacial es de carácter geográfico, en el sentido de que la localización de las observaciones es fundamental para determinar la forma o especificación de dicha variabilidad. Éste sería, por ejemplo, el caso de la heteroscedasticidad de grupos (“groupwise”), que podría ser modelizada a través de tantos valores



de la varianza de la perturbación aleatoria como distintos grupos geográficos compactos puedan derivarse de los datos.

- En segundo lugar, dado que la estructura es espacial, la heterogeneidad suele producirse conjuntamente con el problema de auto-correlación espacial, no siendo ya adecuadas las herramientas de la econometría tradicional, dado que los contrastes habituales de heteroscedasticidad pueden estar sesgados en un contexto espacial.
- Por último, en tercer lugar, en un modelo de regresión de corte transversal, ambos efectos de auto-correlación y heterogeneidad espacial pueden ser, desde una óptica meramente observacional, totalmente equivalentes. Así, por ejemplo, un “cluster” o agrupamiento espacial (observado en localizaciones muy próximas) de los residuos con valores extremos podría ser interpretado como un problema de heterogeneidad espacial (heteroscedasticidad de grupos o “groupwise”), o también como un efecto de auto-correlación espacial. Por eso, deben estructurarse perfectamente ambos efectos espaciales para identificar correctamente los parámetros de un modelo con estos problemas y nunca considerar un aspecto independientemente del otro.

2.3.2.3 Especificación de un modelo con heterogeneidad espacial

Una de las soluciones más extendidas para modelos con heterogeneidad espacial, se basa en especificar y estimar un modelo de **expansión lineal espacial**. Considérese un modelo como el siguiente:

$$y = \alpha + \beta X + \varepsilon \tag{15}$$

La expansión se realiza para la constante, como así también para los β . Las coordenadas podrán actuar como variables de expansión, siendo γ la variable que considera la tendencia norte-sur y ψ la tendencia este-oeste. Bajo estas consideraciones, se tiene que:

$$\alpha_i = \alpha_0 + \alpha_1 \gamma_i + \alpha_2 \psi_i \tag{16}$$

$$\beta_i = \beta_0 + \beta_1 \gamma_i + \beta_2 \psi_i \tag{17}$$



Donde los valores de ordenada al origen y pendiente cambiarán según la región i .

Incorporando (16) y (17) en (15), se obtiene que:

$$y = \alpha_0 + \alpha_1\gamma_1 + \alpha_1\psi_1 + \alpha_2\gamma_2 + \alpha_2\psi_2 + \dots + \alpha_k\gamma_k + \alpha_k\psi_k + \beta_1\gamma_i + \beta_2\psi_i \quad (18)$$

En esta última expresión los parámetros varían en el espacio, dando lugar a estimaciones consistentes. Otra posibilidad, igualmente difundida, ha consistido en realizar la expansión según las regiones formen parte de una región núcleo o una región marginal (o centro vs. periferia en otros términos).

Donde $W_i = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & & \\ \vdots & & \ddots & 0 \\ 0 & \dots & & w_{in} \end{bmatrix}$ y $w_{ij} = \exp(\gamma d_{ij}^2)$, $\forall i=1, 2, \dots, n$ y $j=1, 2, \dots, n$. (19)

2.3.3 Especificación de los efectos espaciales en modelos de regresión

Como ha sido definido la econometría espacial, con palabras de Anselin (2001d), como “la parte de la econometría que se ocupa del tratamiento de la interacción espacial (auto-correlación espacial) y la estructura espacial (heterogeneidad espacial) en los modelos de regresión de corte transversal y de datos de panel”. Es decir, que podría decirse que la principal característica de la econometría espacial consistiría en el modo como los efectos espaciales de dependencia y heterogeneidad espacial son tenidos en cuenta en el contexto del análisis de regresión, para evitar problemas de mala especificación en los modelos. Obviamente, esto implica que las propiedades del espacio geográfico han sido previamente especificadas matemáticamente, por ejemplo, a través de una matriz de pesos espaciales.

Uno de los problemas más importantes planteados en este campo de la modelización espacial, ha sido la enorme variedad de especificaciones aplicadas a los procesos espaciales propios, por ejemplo, de la ciencia regional. Podría dar la impresión de que cada modelo particular necesita su propio esquema metodológico, cuando afortunadamente, gracias a las técnicas econométricas, es posible agrupar estas



especificaciones y realizar tipologías de modelos espaciales, como las que se presentan a continuación.

En primer lugar, el efecto de dependencia o auto-correlación espacial puede (o no) estar presente en la especificación de un modelo de regresión lineal de dos formas, que han sido denominadas como sustantiva o residual, ya sea como consecuencia de variables sistemáticas (endógena y/o exógenas) espacialmente auto-correlacionadas o por la existencia de un esquema de dependencia espacial en el término de la perturbación aleatoria, respectivamente. Por otro lado, el fenómeno de heterogeneidad espacial debe también especificarse en un modelo cuando la inestabilidad estructural, propia de este efecto, se manifieste como varianza no constante en los residuos de una regresión (heteroscedasticidad) o como variación en los coeficientes del modelo (inestabilidad paramétrica) que, a su vez, puede ser continua o discreta. Además, los efectos espaciales de auto-correlación y heterogeneidad pueden ser también especificados en otros esquemas econométricos, como los modelos de datos de panel, modelos multi-ecuacionales, modelos de variables truncadas, modelos no lineales.

2.3.4 Modelización en presencia de dependencia espacial

Cuando se trabaja con datos de corte transversal es muy habitual encontrarse con la existencia de dependencia espacial. Si el valor de una variable en un lugar está relacionado con el valor de la misma variable en otro lugar o lugares se habla de dependencia o auto-correlación espacial. Estas relaciones espaciales se caracterizan por la multidireccionalidad de manera que es necesario resolver los problemas que ocasiona la presencia de este tipo de efectos espaciales. Para ello, es necesario utilizar las técnicas específicas que proporciona la econometría espacial en la modelización y que son diferentes de las utilizadas en los modelos temporales.

En la década de los 90 la econometría espacial ha experimentado un fuerte desarrollo. Sin embargo, la mayor parte de los estudios de carácter empírico que utilizan dicha metodología se han vinculado a la economía regional y urbana y en mucha menor medida a la economía laboral. A diferencia de lo que ocurría en el pasado, la existencia de dependencia espacial es más la regla que la excepción, y en lo relacionado con la generación de viaje ocurre lo mismo. Hoy en día, existen razones teóricas y empíricas para incorporar este hecho a las investigaciones relacionadas con la generación de viajes.



En el caso de Santander es preciso señalar la escasez de viajes que incorporan la dependencia espacial en la modelización.

2.3.5 Fuentes de auto-correlación espacial

Las principales fuentes de auto-correlación espacial pueden ser los errores de medida y la propia interacción espacial de las unidades. En términos económicos, los efectos desbordamiento (spillovers) pueden generar la auto-correlación espacial. Esto se ha visto potenciado con los procesos de integración económica.

Es muy importante identificar los patrones espaciales para entender cuál de ellos describe mejor los datos y saber si vamos por el camino de la heterogeneidad espacial o por el de la correlación espacial. Para este procedimiento se han presentado algunos estadísticos globales como los índices de Moran y la C de Geary, y otros locales como Lisa y Glisa (Moreno, 2000).

La índice de Moran es una medida de auto-correlación espacial que es análoga al coeficiente de correlación, cuyo valor está entre -1 y 1, significando auto-correlación espacial negativa fuerte con el -1 y auto-correlación positiva fuerte con el 1. Antes de construir la Índice de Moran se debe definir la matriz de vecindad espacial. Este es uno de los temas operacionales cruciales en la econometría espacial, ya que se requiere expresar formalmente la estructura de dependencia espacial que va ser incorporada al modelo. Esta matriz representa el patrón espacial entre las observaciones que llamaremos localizaciones (Anselin 1992).

La multidireccionalidad que caracteriza las relaciones espaciales complica su introducción en los modelos empíricos. Hay que tener presente que la estimación de estas relaciones espaciales sería imposible puesto que se precisaría conocer al menos $N(N - 1) / 2$ elementos de la matriz de covarianzas de las observaciones espaciales asumiéndose que las relaciones son simétricas. La solución más habitual consiste en definir a priori lo que se conoce como matriz de pesos espaciales, W :

2.4 El análisis exploratorio de datos espaciales (AEDE)

En muchas ocasiones, el análisis de datos se realiza con el objetivo de contrastar determinadas hipótesis o modelos como, por ejemplo, cuando se quiere conocer a través



de una encuesta si la edad de las personas está relacionada positivamente con su capacidad de asumir riesgos o si, en situaciones más complicadas, se desea conocer si la relación entre un determinado número de variables es lineal o no lineal (Chasco, 2003). Sin embargo, otras veces, el investigador no dispone de una hipótesis definida a contrastar, por lo que la naturaleza del estudio es “exploratoria” y se deben utilizar ciertas técnicas especialmente diseñadas para encontrar algún “sentido” en los datos, es decir, para la detección en ellos de estructuras o conglomerados (“clusters”).

El análisis exploratorio de datos espaciales (AEDE) es una disciplina del análisis exploratorio de datos, también denominado minería de datos “data mining”, que ha sido diseñada para el tratamiento específico de los datos espaciales o geográficos. Este análisis se utiliza para identificar relaciones sistemáticas entre variables cuando no existen expectativas claras sobre la naturaleza de estas relaciones. Esto último resulta muy habitual en el ámbito de los estudios del mercado geográfico (geomarketing), donde se suele trabajar con grandes bases de datos cuya estructura no siempre es bien conocida. Por ésta razón GeoDa nos proporciona una serie de técnicas para visualizar y estimar la auto-correlación espacial; es comúnmente considerado como un subconjunto del análisis exploratorio de datos (AED) que se ocupa de las características distintivas de datos geográficos, con especial énfasis en los problemas de auto-correlación espacial y heterogeneidad espacial (Anselin, 1998; Haining, 1990).

El AEDE puede definirse como “una colección de técnicas para describir y visualizar distribuciones espaciales, identificar localizaciones espaciales atípicas o outliers espaciales, descubrir patrones de asociación espacial, clusters o puntos calientes, y sugerir regímenes espaciales u otras formas de heterogeneidad espacial” (Anselin, 1998; p. 4).

De este modo, es posible obtener una estructura explicativa de los datos a través de unas técnicas que combinan herramientas de la estadística básica (descriptivos, correlaciones, tablas de frecuencias o de correlación cruzada) con análisis multivariante avanzado, especialmente diseñado para identificar formas en grandes bases de datos (análisis cluster, escalas multidimensionales, análisis logit, correlación canónica, análisis de correspondencias, árboles de clasificación, etc.).



De esta forma, al igual que el AED, el AEDE procura detectar en los datos posibles patrones y sugerir hipótesis sobre posibles relaciones entre las variables involucradas en el análisis. Dentro de la colección de técnicas incluidas en el AEDE se encuentran:

- Visualización de distribuciones espaciales
- Visualización de asociación espacial
- Indicadores locales de asociación espacial (LISA - Local Indicator of Spatial Association)
- Indicadores multivariados de asociación espacial

El desarrollo de nuevos paquetes informáticos y el mejoramiento de otros que ya existían, han permitido a dichas técnicas alcanzar su potencial como procedimiento para proponer posibles conexiones entre las variables. En el caso de ésta aplicación, hemos trabajado con el software libre GeoDa desarrollado por Luc Anselin en la Arizona State University.

2.4.1 Análisis exploratorio de datos

El análisis exploratorio de datos (AED), a diferencia de los contrastes de hipótesis tradicionales diseñados para la verificación a priori de conexiones entre variables, se utiliza para identificar relaciones sistemáticas entre variables cuando no existen expectativas claras sobre la naturaleza de estas relaciones (Chasco, 2003). El clásico AED consiste en el tratamiento y comparación de una batería de variables utilizando técnicas que identifican en ellas formas estables. El AED podría definirse como “el conjunto de herramientas gráficas y descriptivas utilizadas para el descubrimiento de patrones de comportamiento en los datos y el establecimiento de hipótesis con la menor estructura posible”. Esta definición fue propuesta por Tukey (1977), que hizo posible la extensión de este tipo de análisis multivariante y del “software” estadístico moderno (Anselin, 1998; Anselin y Bao, 1997; Moreno y Vayá, 2000).

De este modo, es posible obtener una estructura explicativa de los datos a través de unas técnicas que combinan herramientas de la estadística básica (descriptivos, correlaciones, tablas de frecuencias o de correlación cruzada) con análisis multivariante avanzado, especialmente diseñado para identificar formas en grandes bases de datos



(análisis cluster, escalas multidimensionales, análisis logit, correlación canónica, análisis de correspondencias, árboles de clasificación, etc.).

2.4.2 Técnicas gráficas

Las técnicas de AED suelen estar acompañadas de métodos de visualización gráfica capaces de identificar conexiones, tendencias o sesgos presentes en las bases de datos inicialmente desestructuradas (Chasco, 2003).

Una de las más conocidas técnicas propias del AED es el “cepillado” o “brushing”, un método interactivo que permite la selección en pantalla de un determinado número de datos para la identificación de características comunes o el examen de efectos o relaciones entre variables relevantes. Estas relaciones entre variables suelen visualizarse mediante el ajuste de funciones (líneas, en un plano de dos dimensiones o superficies tridimensionales) y sus intervalos de confianza, de forma que, por ejemplo, puedan examinarse los cambios que producen sobre dichas funciones la eliminación o añadido (temporal) de un determinado número de datos.

Otras técnicas gráficas del AED incluyen el ajuste y representación gráfica de funciones, alisado de datos, superposición o fusión de múltiples “vistas” (“views”), añadido de datos en gráficos, identificación y señalización de subconjuntos de datos que cumplen determinadas condiciones, representación gráfica de intervalos o áreas de confianza, técnicas de reducción de imágenes, etc. La exploración de los datos es sólo un primer paso y sus resultados deben ser considerados como una etapa previa a su “confirmación” (modelización). Si los resultados de la fase exploratoria sugieren la utilización de un modelo determinado, entonces su validación puede verificarse aplicando dicho modelo a un nuevo conjunto de datos y contrastando el ajuste, es decir, su capacidad predictiva.

Existen dos perspectivas para llevar a cabo un AEDE a partir de técnicas gráficas, según que se trate de un análisis desarrollado por la geoestadística o por la econometría espacial (Cressie; 1993). El objeto del análisis geoestadístico se encuentra, por lo general, en el entorno de las ciencias medioambientales (física, geología, hidrología, etc.) y se centra en una muestra de datos puntuales procedentes de distribuciones geográficas continuas (por ejemplo, precipitación atmosférica, humedad de la tierra, altura del



océano, etc.). Por su parte, la econometría espacial analiza localizaciones geográficas discretas de puntos o polígonos (provincias, municipios, etc.). Es lo que se denomina perspectiva de retícula o “lattice”, y se encuentra mucho más centrado en el análisis de los fenómenos socioeconómicos (distribución de la renta, clientes, votantes, etc.) Dicho de otro modo por un lado, la aproximación geoestadística y por el otro, **la aproximación lattice**. De acuerdo con Anselin (1998) la principal diferencia entre estas dos técnicas se encuentra en la forma en la que la proximidad espacial es formalizada.

a) Abordaje geoestadístico

En este, el supuesto de procesos espaciales continuos conduce al uso de la distancia métrica como medio para organizar las observaciones.

Al respecto, Anselin (1998, p. 5) sostiene: “Dado que la asociación espacial se asume como una función suave de la distancia, una medida formal de la (no)similitud entre dos observaciones, como la diferencia al cuadrado, es comparada con la distancia que los separa. Un mayor grado de auto-correlación espacial implica pequeñas diferencias en distancias cortas y diferencias crecientes en distancias mayores. La función formal que operacionaliza esta noción es el variograma [ver, e.g., Cressie (1993) para detalles técnicos]. Las técnicas del AEDE desde la geoestadística radican en las formas en las que el variograma puede ser visualizado, resumido y probado para la presencia de no estacionariedades locales u otra conducta atípica. Un importante aspecto de esta visualización es que las entidades en un variograma pertenecen a pares de observaciones (separadas por una distancia dada), y no a localizaciones individuales.”

Las técnicas gráficas de esta perspectiva se concentran en:

- La función de distribución acumulativa
- El Box Plot del variograma
- La nube del variograma multivariado

b) La aproximación lattice

En este método las observaciones tienen carácter discreto. En consecuencia, la formalización de las similitudes espaciales se desarrolla de una manera diferente a la empleada por el anterior abordaje.



“El concepto clave aquí es la noción de vecino espacial, que conduce a la construcción de matrices de pesos espaciales y variables espaciales rezagadas” (Anselin, 1998; p. 6).

De esta forma, las herramientas del AEDE en la perspectiva lattice intentan reflejar la asociación entre las variables y sus rezagos espaciales, para diferentes definiciones de valores similares y para diferentes pesos espaciales.

Los instrumentos más útiles son:

- El box map
- El histograma regional
- El scatterplot de Moran
- Los mapas LISA
- El scatterplot multivariante

El AEDE y GeoDa: puede definirse como el conjunto de técnicas que describen y visualizan las distribuciones espaciales, identifican localizaciones atípicas o “atípicos espaciales” (“spatial outliers”), descubren esquemas de asociación espacial, agrupamientos (“clusters”) o puntos calientes (“hot spots”) y sugieren estructuras espaciales u otras formas de heterogeneidad espacial (Anselin, 1999). Por tanto, el AEDE se correspondería con los métodos de estadística descriptiva espacial.

De manera particular, cuando no existe un marco formal o teoría previa acerca del fenómeno que se analiza, deben de utilizarse las técnicas del AEDE. Esta situación se plantea muy a menudo en el campo de las ciencias sociales, cuando se analizan grandes bases de datos geográficos cuya distribución no se conoce a priori. Por ejemplo, parece obvio que la distribución del nivel de instrucción en las provincias españolas sigue una tendencia de norte (mayor nivel) a sur (menor nivel). Sin embargo, esta distribución suele ser más desconocida cuando se analiza este fenómeno para el ámbito de secciones censales en el interior de un municipio. En los últimos años, el AEDE ha sido introducido en algunas aplicaciones de economía regional (por ejemplo, Moreno y Vayá, 2000), así como en los ejercicios de predicción-extrapolación de datos (Chasco, 2003). Sin embargo, son prácticamente nulas las aplicaciones realizadas en el campo del marketing.



2.4.3 Perspectivas en el análisis de auto-correlación espacial univariante: global y local

En la taxonomía de técnicas del AEDE propuesta por Anselin (1998), se ponía de manifiesto la existencia de una doble perspectiva en el análisis del fenómeno de asociación o dependencia espacial, lo que, también se ha denominado, en Wise et al. (1999), como alisado o “smooth” (perspectiva global) y asperezas o “rough” (perspectiva local). La perspectiva global del fenómeno de auto-correlación espacial tiene por objeto el contraste de la presencia de tendencias o estructuras espaciales generales en la distribución de una variable sobre un espacio geográfico completo, mientras que el fenómeno de dependencia local vendría definido por una concentración, en un lugar del espacio global analizado, de valores especialmente altos o bajos (“puntos calientes/fríos”, “picos”, “bolsas” de valores atípicos) de una variable en comparación con el valor medio de la misma (Chasco, 2003).

Aunque la mera observación en un mapa de la distribución de una variable espacial permite captar de forma intuitiva la existencia o no de patrones de comportamiento en dicha variable, esta información será siempre subjetiva y altamente dependiente de, por ejemplo, el número de intervalos establecido para la representación de dicha variable en el mapa. Por eso, resulta fundamental contar con una combinación de medidas o instrumentos gráficos y estadísticos, capaces de detectar la presencia significativa de auto-correlación (global y/o local) en una variable espacialmente distribuida. En esta línea, los contrastes de auto-correlación espacial, en combinación con las herramientas del AEDE, ofrecen criterios estadísticos objetivos que permiten confirmar o rechazar la presencia de tendencias o estructuras espaciales en la distribución de una variable.

Los estadísticos de auto-correlación espacial global son las primeras formulaciones propuestas en la literatura como medida estadística del efecto de auto-correlación espacial. Estos contrastes tienen la capacidad de resumir el esquema general de dependencia presente en una variable espacial en un único indicador (Moreno y Vayá, 2000). Recientemente, se han propuesto también los tests de auto-correlación espacial local, muy útiles para la detección de agrupaciones (“clusters”) de este fenómeno en el espacio geográfico general. En la Tabla 3, se incluyen los estadísticos de auto-correlación, global y local, más utilizados y mencionados en la literatura.



De todos ellos, los tests I de Moran (1948) y c Geary (1954) son probablemente los contrastes de auto-correlación espacial más conocidos. En el trabajo con datos ordinales y de intervalo, ambos contrastes son de uso generalizado, aunque también puede añadirse el estadístico G (d), que fue propuesto por Getis y Ord (1992) junto con otros indicadores de auto-correlación local, que han sido denominados conjuntamente como “familia de estadísticos G de Getis y Ord”. Por su parte, los tests de auto-correlación espacial local permiten contrastar la presencia de subzonas de dependencia espacial dentro de un espacio general dado, que Anselin (1995b) denomina Indicadores LISA (Indicadores Locales de Asociación Espacial), para los que establece una serie de buenas propiedades estadísticas.

Autocorrelación global	Autocorrelación local
<i>I de Moran (1948)</i>	<i>I_i de Moran (Anselin, 1995B)</i>
<i>c de Geary (1954)</i>	<i>c_i de Geary (Anselin, 1995B)</i>
<i>Γ de Mantel (1967)</i>	<i>Γ_i de Mantel (Anselin, 1995B)</i>
<i>G(d) de Getis y Ord (1992)</i>	<i>G_i(d) de Getis y Ord (1992)</i>
	<i>G_i[*](d) de Getis y Ord (1992)</i>
	<i>Nueva-G_i de Ord y Getis (1995)</i>
	<i>Nueva – G_i[*] de Ord y Getis (1995)</i>
	<i>O_i de Ord y Getis (2001)</i>

Tabla 1: Principales contrastes de auto-correlación espacial univariante – Fuente: Chasco, 2003

2.4.4 Medidas de asociación espacial global

El análisis de auto-correlación espacial global realiza un examen conjunto de todas las unidades que componen la muestra para determinar si las unidades espaciales se encuentran distribuidas aleatoriamente o si, por el contrario, lo hacen conforme a un patrón determinado (Chasco, 2003).

Existen tres contrastes desarrollados para ese propósito, que se describen a continuación: la I de Moran, la C de Geary y la G (d) de Getis y Ord.

La expresión analítica del contraste de Moran es:



$$I = \frac{N \sum_{ij=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^N (x_i - \bar{x})} \quad (20)$$

con $i \neq j$, donde x_i representa el valor de la variable cuantitativa x para la región i , N el tamaño muestral, w_{ij} los pesos espaciales de una matriz de contactos W y S_0 la sumatoria de los pesos espaciales. La distribución de la I de Moran es asintótica normal [$N \sim (0,1)$] cuando N es grande.

Para el caso de la C de Geary, la formulación es:

$$C = \frac{N-1 \sum_{ij=1}^N w_{ij} (x_i - x_j)}{2S_0 \sum_{i=1}^N (x_i - \bar{x})} \quad (21)$$

con $i \neq j$, donde los elementos tienen el mismo significado que en la expresión anterior. La C de Geary también es asintótica normal [$N \sim (0,1)$] cuando N es grande.

Finalmente, la expresión de la $G(d)$ es:

$$G(d) = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} d(x_i, x_j)}{\sum_{i=1}^N \sum_{j=1}^N x_i x_j} \quad (22)$$

con $i \neq j$, siendo i y j vecinos siempre que se encuentren a una distancia d . El contraste se distribuye normal [$N \sim (0,1)$] para N grande. Nótese que este contraste es aplicable sólo a variables positivas y naturales.

La formulación de los contrastes descriptos anteriormente puede llevarse a cabo con cualquier especificación de matriz W . No obstante, los resultados finales de los contrastes, respecto a la determinación de auto-correlación global o no, son sensibles a esas especificaciones.

Otro factor de sensibilidad con los resultados de los contrastes son las transformaciones que pueden sufrir las variables. Por estas razones, resulta útil ensayar varias formulaciones alternativas de matriz W .



Los tres contrastes de auto-correlación espacial global, I, c y G(d), proporcionan una información complementaria y pueden ser utilizados conjuntamente (Moreno y Vayá, 2000). Sin embargo, deben conocerse también las diferencias existentes entre ellos, algunas de las cuales se resumen en la Tabla 2. Por ejemplo, el test I es una medida similar a la covarianza (diferencia de cada x_i respecto del valor medio, \bar{x}) de los valores de X en cada punto del espacio, mientras que el contraste c de Geary podría ser considerado como un análisis de la varianza, más afectado por la distribución de los datos muestrales que I (Cliff y Ord, 1981). Por su parte, el test G(d) es en realidad una medida de concentración o asociación de una variable geográfica X, que presenta un gran parecido formal con el test I, sobre todo en el numerador de ambas expresiones matemáticas: mientras que G(d) es proporcional a la suma de los productos $x_i \cdot x_j$, el contraste I mide la correlación de cada valor x_i con sus vecinos (Getis y Ord, 1992).

Una importante limitación, propia de la familia de estadísticos G, consiste en que sólo pueden aplicarse a variables positivas, por lo que el contraste G (d), a diferencia del test I, no podrá ser utilizado para analizar la presencia de auto-correlación espacial, por ejemplo, en los residuos de una regresión. Además, debe tenerse en cuenta que la matriz W, cuando se aplica a la obtención del test G(d), debe ser simétrica por lo que no podrá realizarse, en estos casos, ningún tipo de estandarización por filas en la misma. Estas restricciones, sin embargo, no tienen lugar en los tests I y c, que pueden ser aplicados a valores positivos y negativos de cualquier variable continua, así como a cualquier tipo de matriz de pesos.

Test	Hipótesis nula (z no significativo)	Hipótesis alternativa (z significativo)	
		z > 0	z < 0
I de Moran	No autocorrelación espacial	Autocorrelación espacial POSITIVA	Autocorrelación espacial NEGATIVA
c de Geary	No autocorrelación espacial	Autocorrelación espacial NEGATIVA	Autocorrelación espacial POSITIVA
G(d) de Getis y Ord	No autocorrelación espacial	Autocorrelación espacial POSITIVA (valores de Y altos)	Autocorrelación espacial POSITIVA (valores de Y bajos)

Tabla 2: Interpretación de los valores estandarizados de los estadísticos de Auto-correlación espacial global – Fuente: Chasco, 2003



Por último, tampoco debe olvidarse la importancia que tiene la matriz de pesos espaciales en el proceso de contraste del fenómeno de auto-correlación espacial global, pues está bien demostrado que los resultados obtenidos por los diferentes estadísticos pueden variar, a veces de forma sensible, en función de la matriz W especificada. Por eso, se recomienda la utilización, en la aplicación de estos contrastes, de diversas expresiones de W . También suele recomendarse calcular el test I de Moran para matrices de contigüidad de órdenes o potencias sucesivas porque permite contrastar si el esquema de auto-correlación espacial detectado entre regiones vecinas es extensible a regiones alejadas en el espacio (Anselin y Smirnov, 1996; Moreno y Vayá, 2000).

2.4.5 Medidas de asociación espacial local

Los contrastes analizados en la sección anterior presentan una fuerte limitación: no son capaces de considerar situaciones de aglomeración (clusters) en un área determinada, para la que cabría esperar valores más bajos o más altos si existiese una distribución homogénea (Moreno y Vayá, 2000). Esto implica que un determinado esquema de auto-correlación espacial, detectado mediante contrastes globales, pueda no mantenerse para toda la muestra.

El fenómeno de auto-correlación, dependencia o asociación espacial local puede ser definido como una concentración, en un lugar del espacio global analizado, de valores especialmente altos o bajos de una variable en comparación con el valor medio esperado (o media de la variable considerada). Este fenómeno se produce en procesos espaciales no estacionarios, lo que contrasta con el hecho de que el análisis de la asociación espacial ha estado, durante mucho tiempo, asentado sobre el principio de estacionariedad espacial, hipótesis ésta que ha sido reconocida como bastante irreal, sobre todo cuando se utiliza un gran número de unidades geográficas (Aznar et al., 1996). Efectivamente, en muchas ocasiones las propiedades estadísticas (función de distribución conjunta) de un proceso espacial no dependen sólo de la posición relativa (ángulo) y las distancias respectivas de los puntos en el espacio, sino también de cuestiones como el contexto económico. Es decir, no siempre puede aceptarse un tipo de dependencias similar entre cualquier par de puntos situados a la misma distancia en un mapa, por ejemplo, entre las ciudades de Madrid y Barcelona, por un lado, y Madrid y Ourense, por otro (Chasco, 2003).



Este problema de la dependencia espacial local puede plantearse desde dos puntos de vista (Vayá y Suriñach, 1996):

- Existe la posibilidad de que, en un espacio dado, no se detecte la presencia de auto-correlación espacial global en la distribución de una variable aunque, de hecho, existan pequeños “clusters” espaciales en los que dicha variable experimenta una concentración (o escasez) importante.
- Existe también la posibilidad de que, habiéndose detectado dependencia a nivel global en una variable, no todas las regiones del espacio considerado contribuyan con igual peso en el indicador global, es decir, que coexistan unas zonas en las que la variable se distribuya de forma aleatoria junto a otras con una importante contribución a la dependencia existente.

Para la detección de aglomeraciones (**clusters**) resulta útil el análisis de **auto-correlación espacial local**. Este procedimiento permite medir la auto-correlación espacial para la ubicación de cada observación y está basado en el Indicador Local de Moran.

De acuerdo con Anselin (1995) un indicador LISA (Local Indicator of Spatial Association) es un estadístico que satisface dos requerimientos: por un lado, proporciona una cuantificación del grado de agrupamiento significativo de valores similares alrededor de una observación, y por otro lado, la suma de los LISA's para todas las observaciones es proporcional a un indicador global de asociación espacial, por lo que resulta útil para medir la contribución de cada observación al valor del contraste global (sólo para el caso de la I de Moran).

Este análisis es también particularmente útil para detectar la posible presencia de localizaciones que muestren valores altos respecto a la media de otros emplazamientos colindantes, y para las variables de interés, o viceversa.

Los contrastes más relevantes para el análisis de la auto-correlación local son: la I de Moran y la G (d) de Getis y Ord.

La expresión analítica del contraste de Moran es:



$$I = \frac{z_i}{\sum_i z_i / N} \sum_{j \in J_i} z_j \quad (23)$$

Donde Z_i representa el valor de la región i para la variable normalizada y J_i el conjunto de regiones vecinas a i . N es el tamaño muestral. La distribución de la I de Moran es asintótica normal [$N \sim (0,1)$] cuando N es grande. Finalmente, si el valor de la I es positivo, denotará la existencia de un cluster de valores similares.

La expresión de la $G(d)$ es:

$$G(d) = \frac{\sum_{j=1}^N w_{ij} d(x_j)}{\sum_{j=1}^N x_j} \quad (24)$$

con $i \neq j$, siendo i y j vecinos siempre que se encuentren a una distancia d , x la variable de interés no normalizada. El contraste se distribuye normal [$N \sim (0,1)$] para N grande. Nótese que este contraste es aplicable sólo a variables positivas y naturales.

2.4.6 Modelo de regresión lineal múltiple con heterogeneidad (Paez et al, 2005)

Este modelo busca incorporar la eventual heterogeneidad espacial entre zonas (es por lo tanto un modelo agregado) estimando un vector de parámetros propio para cada una de las n zonas. La hipótesis que hay detrás de este tipo de modelos es que el efecto de las diferentes variables que explican la generación de viajes varía en función de la ubicación geográfica de las zonas donde se producen (o atraen) los viajes; es decir, el efecto marginal depende de la localización de los viajes. Ello puede ocurrir debido a la escala de las zonas o a decisiones administrativas que varían entre los distintos sectores del sistema de actividades estudiado.

Esta heterogeneidad puede generar sesgo e incluso invalidar contrastes de significancia estadística en los modelos convencionales de regresión lineal.



El modelo, que es aplicable principalmente en análisis agregados (a nivel zonal), es el siguiente:

$$Y_i^p = \beta_{0i}^p + \sum_k \beta_{ki}^p \cdot X_{ik}^p + \varepsilon_i^p \quad \forall i = 1, 2, \dots, n \quad (25)$$

Cada vector de parámetros es estimado de la siguiente forma:

$$\hat{\beta}_i^p = (x' W_i x)^{-1} x' W_i y_i^p \quad (26)$$

Y_i^p representa la variable endógena del modelo, que puede ser la generación de viajes de la unidad i (a nivel de zonas, hogares o individuos), para el propósito p . X_{ik}^p corresponde a la k -ésima variable explicativa, para la unidad i y el propósito p . β_{ik}^p son los parámetros de calibración y ε_i^p es la perturbación aleatoria. La variable d_{ij} representa la distancia entre los centros de gravedad de las zonas i y j . Este modelo se estima directamente por mínimos cuadrados ordinarios, y la elección del parámetro $\hat{\beta}_i^p$ es tal que se minimice la suma de errores al cuadrado.

2.4.7 Regresión geográficamente ponderada

Mientras que GeoDa se aboca a modelar la auto-correlación espacial, la regresión geográficamente ponderada (GWR) busca analizar la no estacionariedad de los datos. Así esta técnica hace posible explorar o identificar variaciones por zonas. Esto es posible porque una regresión geográficamente ponderada permite la estimación de parámetros locales y no sólo globales. Un estimado local es computado “tomando prestada” información de las unidades dentro de una distancia previamente establecida, donde las unidades más cercanas tienen mayor peso que las más lejanas (Fotheringham et al. 2001:52). Como tal, esta técnica cuestiona el supuesto implícito en las regresiones estándares de que un modelo explicativo puede aplicarse por igual a toda el área geográfica analizada, cuando en realidad puede haber importantes variaciones tanto en el modelo completo como en la relación específica entre la variable dependiente y una de sus variables explicativas (Charlton et al. 2003, Fotheringham et al. 2001). Así, nuestro modelo explicativo podría funcionar mejor en ciertas zonas de la ciudad que en otras y algunas variables importar más en ciertas regiones que en otras o incluso tener efectos opuestos.



Además de la idea de que las variables tienen efectos diferenciados a través de la ciudad, los mapas de GWR permiten identificar la presencia de los llamados **regímenes espaciales** que pueden definirse como regiones geográficas donde se observan cambios discretos en la distribución de las variables (media, varianza) y donde un tipo distinto de relación parece establecerse entre las variables consideradas

2.4.8 Datos espaciales (Chasco, 2003)

Naturaleza de los datos espaciales

El espacio geográfico es una realidad de naturaleza compleja y distribución continua imposible de manejar sin realizar un esfuerzo de abstracción encaminado a reducir o dividir esta continuidad en entidades numéricas discretas, observables y susceptibles de medición matemática. En esta línea, podría definirse el dato espacial como la observación de una variable asociada a una localización del espacio geográfico. Esta localización puede ser expresada mediante tres tipos de objetos espaciales: puntos, líneas y polígonos (Anselin, 2001b, MapInfo, 1995).

- Los puntos se encuentran determinados por las coordenadas terrestres de latitud y longitud, pudiendo corresponderse con individuos, empresas, ciudades, delitos cometidos o accidentes acontecidos.
- Las **líneas** son objetos abiertos que cubren una distancia dada y comunican varios puntos o nodos (dada la naturaleza esférica de la Tierra, las líneas son en realidad “arcos”). Éste sería el caso de las líneas de transmisión telefónica, infraestructuras viarias, calles de una ciudad, etc.
- Por último, los **polígonos** son figuras planas conectadas por distintas líneas u objetos cerrados que cubren un área determinada, como es el caso de países, provincias, secciones censales o áreas comerciales.

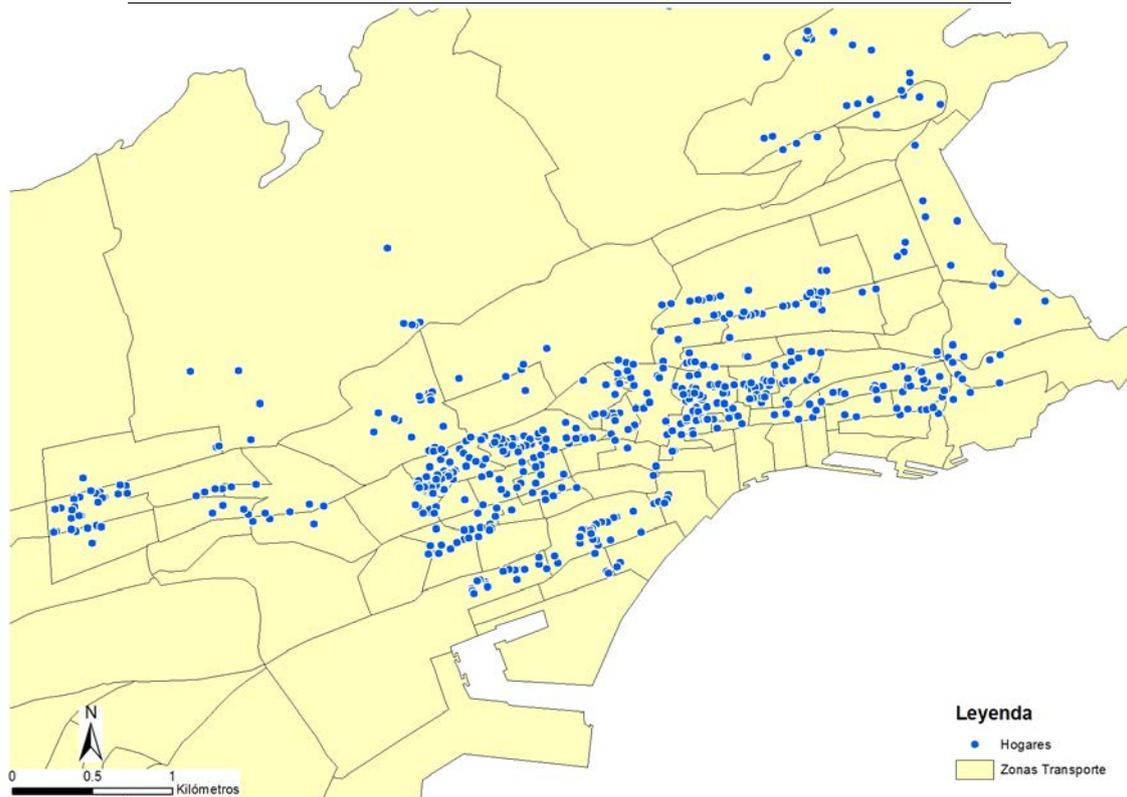


Figura 7: Ejemplo de objetos espaciales en el municipio de Santander

En la Figura 7 , se han representado a modo de ejemplo algunos objetos espaciales en el municipio de Santander: los polígonos son las zonas de transporte de la ciudad, los puntos son los hogares presentes en los mismos. Para las zonas de transporte es posible medir variables como la población residente.

Como ya se ha indicado, los datos espaciales se caracterizan por su naturaleza georreferenciada, multidireccional y multidimensional.

- La **georreferenciación** pone de manifiesto que la posición relativa o absoluta de cualquier elemento sobre el espacio contiene una información valiosa, pues la localización (el lugar donde acontecen los fenómenos) debe considerarse explícitamente en cualquier análisis de datos espaciales.
- La **multidireccionalidad** aflora en situaciones de dependencia o interacción espacial, así como en otros fenómenos como externalidades, efecto contagio, desbordamiento (“spillovers”) o réplica (“copy-cating”). La primera ley de la geografía de Tobler (1979) indica que “todo tiene que ver



con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas”.

- La **multidimensionalidad** de las áreas geográficas significa que, en ellas, no es posible distinguir entre pasado, presente, futuro, sino que todo es presente, todo es pasado y todo es futuro.

En la geografía, el proceso de “discretización” o partición del espacio es más sencillo cuando se trata de objetos con localización o fronteras claramente identificadas (edificios, carreteras o provincias), a diferencia de otros fenómenos que varían de forma continua, como la elevación del terreno, la temperatura del suelo o la densidad de la vegetación. En estos casos, la continuidad del fenómeno puede “aproximarse” mediante una muestra de puntos situados en lugares representativos de la superficie que se desea registrar, de forma que sea posible obtener los valores situados entre dos puntos a partir de alguna función de alisado de los mismos. Existe además otro modo de expresar algunas realidades continuas sobre el espacio geográfico que consiste en determinar un grupo de polígonos contiguos con características homogéneas en su interior, pero diferentes entre sí. Los sistemas de información geográfica (GIS), como se verá a continuación, utilizan dos estructuras de datos, vectoriales y “raster”, para representar esta compleja realidad espacial.

Sistemas de Información Geográfica (GIS)

Los sistemas de información geográfica (GIS o “Geographic Information System”) hacen posible la visualización, exploración, almacenamiento eficaz, recuperación rápida y visualización interactiva de las formas correspondientes a conjuntos de datos geográficos. Los nuevos productos GIS no requieren de una gran especialización informática, pues son de fácil manejo para cualquier usuario que desee trabajar con datos espaciales. A partir de ficheros estándar grabados, por ejemplo, en programas de texto (.txt, .prn), bases de datos (.dbf, .mdb) u hoja de cálculo (.xls, .wks), es posible introducir en el GIS la información geográfica en forma de tablas cuyos datos podrán ser visualizados, manejados o tratados por un sistema de “query” (mediante lenguaje “SQL”), resumidos y organizados geográficamente. Así, con un fácil aprendizaje, podrán utilizarse los datos geográficos de un modo completamente nuevo, siendo posible el descubrimiento de estructuras o relaciones hasta ahora ocultas, que ofrecen nuevas perspectivas y conclusiones en cualquier investigación o negocio.



Los avances recientes en “hardware” y “software” informático han modificado la naturaleza del análisis espacial llevado a cabo por los GIS. Nuevos conceptos como almacenamiento de datos y minería de datos espaciales superan ya con creces la revolución cuantitativa que sufrió la geografía en los años 60, cuando se establecieron la mayoría de los fundamentos del actual análisis espacial.

Las ubicaciones geográficas de los datos suelen ser almacenadas en dos tipos de estructuras, vectorial o de datos “raster”, mientras que los datos propiamente dichos se guardan en un conjunto de tablas relacionadas geográficamente con las formas o figuras a que corresponde, lo que suele ser conocido como estructura de datos georrelacionados.

- Los **datos vectoriales** registran la información a modo de coordenadas geográficas (x,y) en un sistema plano o rectangular de coordenadas, de forma que las formas o figuras de puntos se registran como localizaciones simples del tipo (x,y), mientras que las formas lineales, incluyendo los contornos de los polígonos, se recogen como series ordenadas de coordenadas (x,y). Este tipo de datos vectoriales son muy apropiados para registrar la ubicación de formas geográficas discretas, con localizaciones precisas, como calles, ríos, límites fronterizos de parcelas, postes telefónicos, etc. Los datos vectoriales son altamente dependientes de la estructura de coordenadas (x,y), por lo que la representación gráfica de los vectores tiene forma de segmentos de líneas rectas, que pueden ser visualizados realizando una aproximación (“zoom”) sobre este tipo de formas, como puede apreciarse en la Figura 9 (imagen izquierda), con los límites de dos barrios de Madrid, en negro, y la representación de la M-30, en rojo.
- Los **datos “raster”**, por su parte, registran la información espacial en una cuadrícula regular o matriz organizada en filas y columnas, de forma que cada celda contiene un número que representa una forma geográfica determinada, como un tipo de suelo, la elevación, el uso de la tierra, una pendiente, etc. Este tipo de datos “raster” suele ser utilizado para almacenar información sobre formas geográficas que varían de forma continua sobre una superficie, como la elevación del terreno, reflectancia, profundidades marinas, etc. Los datos procedentes de imágenes son también un modo de



datos “raster” en los que cada celda o “pixel” almacena un valor registrado como un dispositivo óptico o electrónico. Los datos “raster” son muy dependientes de la resolución del entramado regular en el que se graban, siendo fijo el tamaño de las celdas del entramado, de modo que si se realiza una aproximación sobre estos datos, podrá observarse la forma de las celdas, tal como se aprecia en la imagen de la derecha de la Figura 8, correspondiente a la imagen de un mapa de carreteras Michelin, incorporada a un GIS.

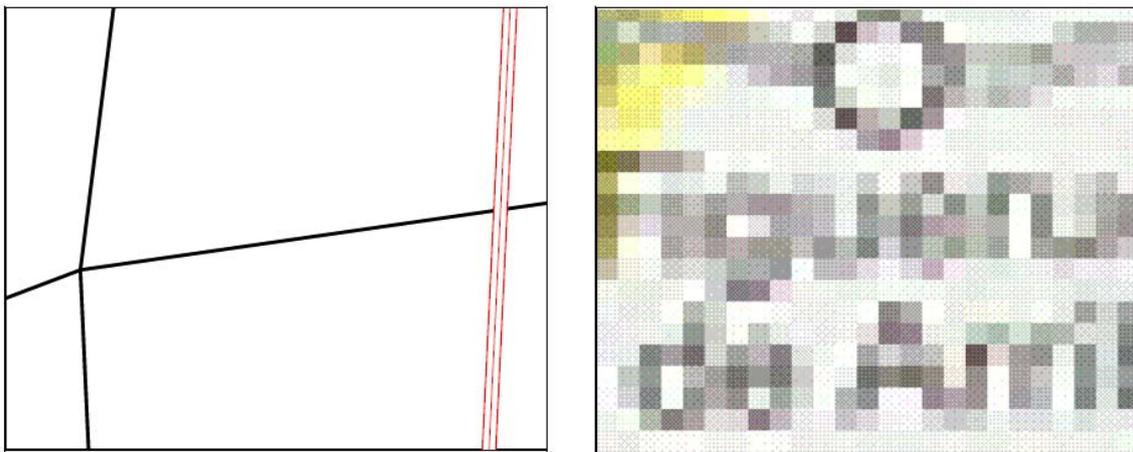


Figura 8: Ejemplo de datos vectoriales (izquierda) y raster (derecha) – Fuente: Chasco, 2003

La utilización de un GIS, junto con el análisis de datos y modelos espaciales, es cada vez más común en los campos de la economía aplicada, política económica, economía urbana y medioambiental, y economía del desarrollo. En los últimos años, se ha prestado especial atención a la integración del análisis espacial en el entorno GIS, tanto desde el campo académico como desde el comercial, planteándose diversos enfoques sobre el contenido concreto que debería tener el paquete de herramientas de análisis espacial de un sistema de información geográfica (ver, entre otros, Haining et al., 2000; Anselin, 1998; Fischer y Getis, 1997).

La aplicación de las herramientas de análisis espacial y el GIS hace posible la obtención y tratamiento estadístico-econométrico de una gran variedad de datos geográficos de diferentes escalas o ámbitos territoriales. Efectivamente, como se verá a continuación, en los últimos años se está produciendo una auténtica explosión en la disponibilidad de bases de datos socioeconómicas geo-codificadas, es decir, con información acerca de la localización geográfica de las observaciones, lo que impulsa el



desarrollo del análisis espacial en el conjunto de las ciencias sociales y del conjunto de la denominada Ciencia GI o Ciencia de la Información Geográfica (Anselin, 1999a).

Fuentes estadísticas de información micro-territorial

La creciente disponibilidad de fuentes estadísticas de información geográfica (censos, directorios, encuestas, etc.), así como de mejores herramientas informáticas de tipo estadístico-econométrico para su correcto tratamiento y visualización, han dado lugar a una demanda creciente de este tipo de análisis espacial, sobre todo en el campo micro-territorial. Hasta donde las leyes de la privacidad lo permiten, va siendo posible la distribución de datos de una amplia gama de variables socioeconómicas, desde la actividad económica o el mercado de trabajo en la salud pública, la educación o la religión, distribuidos en formatos accesibles al análisis geográfico. Todo esto está motivando una actividad cada vez mayor, en el sector privado, en forma de venta de datos, más o menos públicos, o indicadores propios, especialmente depurados y geocodificados.

En concreto, tanto en los Estados Unidos como en la Unión Europea, los institutos de estadística oficiales ofrecen grandes bases de datos, con información de ámbito regional y local, en formatos accesibles y de bajo costo. Por ejemplo, en los EEUU, la base REIS (“Regional Economic Information System”), incluye una gran cantidad de datos de ámbito provincial, publicada en formato CD-ROM por el Ministerio de Comercio; además, existe en este país una extensa infraestructura de redes viarias espacialmente georreferenciadas, los ficheros “Tiger” de la Oficina del Censo de los EEUU, y mapas digitales de una extensa gama de unidades administrativas. De forma similar, el banco de datos REGIO, publicado por Eurostat, contiene datos regionales de los países integrantes en la UE.

En España, el Instituto Nacional de Estadística dispone de bases de datos regionales fundamentalmente, aunque está realizando un esfuerzo importante por proporcionar también variables de ámbito provincial. Sin embargo, exceptuando la información procedente de los censos de población y padrones, existe aún un importante déficit de información oficial de ámbito micro-territorial para todo el territorio nacional. Algunas instituciones llevan realizando, desde hace unos años, un esfuerzo en esta línea, como es el caso de los Anuarios Económico y Social elaborados por el Instituto L. R. Klein para



“la Caixa” (“laCaixa”, 2001a y b). También debe destacarse la meritoria labor que realizan algunos institutos de estadística regionales para proporcionar información micro-territorial de sus correspondientes comunidades autónomas. Debe advertirse, no obstante, que las bases de datos disponibles de forma gratuita para el ámbito micro-territorial, en España, no siempre se ofrecen convenientemente depurados ni geo-codificados, lo que dificulta especialmente el trabajo a este nivel.

Problemas en el tratamiento de datos espaciales

Como indican Anselin (1999a) y Haining (1995), muchos de los análisis de tipo espacial que se realizan actualmente gracias a la creciente disponibilidad de datos geográficos, tienen que contrastarse con la literatura científica existente, dado que la gran dimensión de las bases de datos y la multitud de formas de obtenerlas dan lugar a un número de cuestiones que no siempre se resuelven satisfactoriamente y requieren de un mayor esfuerzo de investigación. Además, como ya se ha indicado, en la mayoría de los casos, el tratamiento de las bases de datos espaciales (sobre todo, las grandes bases micro-territoriales) no debe llevarse a cabo directamente, tal como es proporcionada por los organismos o fuentes originales, sino que exige un tratamiento que haga posible la obtención de una información lo más precisa, depurada y operativa posible.

Por eso, se resumen a continuación algunos problemas relativos al tratamiento de datos espaciales, entre los que deben destacarse la elección de escala o nivel de agregación territorial, el proceso de obtención y almacenamiento informático de los datos y el análisis exploratorio y confirmatorio de la información micro-territorial.

❖ Elección de la escala o nivel de agregación territorial:

La elección de la escala o ámbito espacial adecuados se ha convertido en un tema fundamental en la investigación espacial, máxime cuando en la actualidad se dispone con relativa facilidad de grandes bases de datos geocodificadas con información recogida desde ámbitos más o menos individuales (individuos, hogares, secciones censales) hasta niveles globales (regiones, países). Además, las cada vez más poderosas herramientas de los GIS hacen posible al usuario pasar de una escala de análisis a otra, así como la integración de datos recogidos en diferentes ámbitos geográficos. Muchas veces, las observaciones procedentes de un nivel de análisis agregado (por ejemplo, países) no



proporcionan necesariamente información útil sobre ámbitos inferiores (municipios, hogares), sobre todo cuando existe heterogeneidad espacial en el fenómeno que se analiza.

Otras veces sucede que algunos estadísticos cambian de valor según se obtengan en un ámbito u otro de agregación territorial, tal como se ha demostrado en los coeficientes de correlación y las medidas de auto-correlación espacial (Arbia, 1989). Es éste un problema metodológico ya viejo y familiar, conocido con diferentes denominaciones: los políticos y sociólogos hablan de “inferencia ecológica” (King, 1997), mientras que en geografía (Openshaw y Taylor, 1979) se trata del “problema de la unidad espacial modificable” (MAUP, “Modifyable Areal Unit Problem”), en estadística es la “estimación indirecta de unidades pequeñas” (Rao, 2000), y los geólogos lo denominan problema del “aumento de escala” o “cambio de soporte” (Chilès y Delfiner, 1999).

Otra cuestión relacionada tiene que ver con la diferencia existente entre el ámbito espacial del proceso a considerar y la escala geográfica concreta en la que se mide la observación, cosa que suele producirse muchas veces en estudios que combinan fenómenos físicos y económicos (Anselin, 2001a). Por ejemplo, en el análisis de los factores socioeconómicos que subyacen al proceso de deforestación tropical (Wood y Skole, 1999), la variable dependiente, la deforestación, se calcula a partir de la interpretación de imágenes vía satélite que dan lugar a datos correspondientes a enrejados regulares de cuadrados o “pixels”. Este enrejado no se corresponde, ni en el tamaño ni en las fronteras geográficas, con las unidades administrativas de las que proceden las variables estadísticas de censos o encuestas. Como normalmente las unidades administrativas son, en estos casos, las de mayor tamaño, son también las que determinan el mínimo ámbito geográfico de análisis, lo que conlleva la agregación (o alisado) de los valores “pixel” y la pérdida de variabilidad espacial en cada unidad geográfica finalmente utilizada. Y ya se sabe que, cuanto mayor es la variabilidad (es decir, cuanto menor es la escala espacial en la que se trabaja), menos precisa será la información agregada como estimador de la variable dependiente. Es más, la variabilidad interna no siempre permanece constante entre unidades, lo que da lugar a problemas de heteroscedasticidad.

❖ **Proceso de obtención y almacenamiento informático de los datos:**



Una vez decidido el ámbito espacial de trabajo, deberán afrontarse los problemas relacionados con el proceso de recogida y tratamiento informático de los mismos. En esta línea, Haining (1995) apunta como problemas a solucionar la identificación y corrección de errores, valores poco fiables o no disponibles para algunas unidades espaciales, los posibles errores de muestreo y la correcta asignación de datos a su correspondiente unidad geográfica.

Efectivamente, en la mayoría de los casos, la información procedente de un determinado organismo público o privado adolece de errores u omisiones que deben ser previamente detectados y depurados, según unos determinados criterios. Cuando sea posible habrá de procurarse la obtención de la información correcta, solicitándola al propio organismo generador o a otro que disponga también de dicha estadística (como sucede en el caso de la información municipal con los ayuntamientos). En caso contrario, podrán aplicarse modelos ingenuos que estimen los datos incorrectos o inexistentes. En concreto, deberán estimarse también los datos estadísticos correspondientes a entidades geográficas de nueva creación (municipios, distritos, secciones, etc.), dado que muchas fuentes de información ignoran la existencia de dichas entidades.

Otras veces se producen errores de muestreo espacial, como se pone de manifiesto en Anselin (1999a). La estratificación de las encuestas en las que se basan las estadísticas y estudios socioeconómicos habituales, suelen ignorar el papel de los efectos espaciales y pueden, por tanto, ser inadecuadas para los propósitos del análisis espacial. En el muestreo espacial debe tenerse en cuenta que el fenómeno de auto-correlación espacial suele estar presente en muchas variables geográficas, lo que obligará, en estos casos, a seleccionar en la muestra observaciones “suficientemente alejadas” unas de otras (o incorrelacionadas), para que puedan ser consideradas como representativas de una población.

Además, el tratamiento de la información micro-territorial procedente de muchos organismos generadores de estadísticas, exige muchas veces la identificación y agrupación de direcciones y localidades pertenecientes a una determinada área poblacional, tomada como unidad de investigación (municipio, distrito, barrio, etc.).

Por último, pueden producirse también problemas relacionados con el proceso de informatización de datos, como los errores de teclado, redondeo, cálculo, geo-



codificación, etc. El análisis de las grandes bases de datos micro-territoriales exige un conocimiento suficiente de los programas informáticos gestores de bases de datos (Microsoft Access, dBase IV) y hojas de cálculo (Microsoft Excel), a los que el investigador deberá recurrir siempre que sea posible. La introducción manual de datos deberá ser la última opción, debido al riesgo más que probable de errores de teclado cuando se trabaja con un volumen grande de información. Hay que tener en cuenta que, en muchas ocasiones, la información facilitada por las fuentes estadísticas se encuentra sin codificar o con una codificación diferente de la establecida por el Instituto Nacional de Estadística (INE), en cuyo caso, habrá de ser geo-codificada para hacer posible un tratamiento posterior por programas gestores de bases de datos o GIS.

❖ **Análisis exploratorio y confirmatorio de la información micro-territorial:**

El tamaño de las bases de datos geoespaciales disponibles constituye también un reto para el análisis espacial. La mayor parte de las técnicas tradicionales de análisis de datos espaciales fueron inicialmente desarrolladas para situaciones en las que los conjuntos de datos tuvieran menos de 100 observaciones¹. Sin embargo, lo normal hoy día es trabajar con bases de datos muy superiores, como sucede con el análisis exploratorio denominado “minería de datos espaciales”, fundamental para la detección de estructuras, agrupaciones, asociaciones y otras interesantes situaciones de no aleatoriedad geográfica. Aunque muchas de las técnicas actualmente en uso, como los indicadores LISA, son en principio aplicables, su implementación en bases de datos muy grandes puede constituir un problema desde el punto de vista informático o conceptual. Por este motivo, el papel de la geo-computación o geo-informática ha llegado a ser más importante que nunca, dado que muchos problemas espaciales son intrínsecamente de orden N^2 (por ejemplo, las matrices de pesos espaciales) por lo que no pueden ser abordados de una forma eficaz a través de los sistemas informáticos habituales.

Otra cuestión relacionada con las bases de datos micro-territoriales de gran tamaño es la elección del paradigma inferencial. Es decir, la teoría asintótica clásica sobre la que se asientan los métodos de estadística y econometría espacial ha sido desarrollada para obtener las propiedades de los estimadores y contrastes estadísticos en muestras finitas, pero no tienen sentido cuando la muestra utilizada prácticamente tiende a infinito. Por eso, en los últimos años, se han desarrollado otros paradigmas prometedores basados en nociones bayesianas o de puro cómputo, como los estimadores de simulación, los



métodos de remuestreo o las aproximaciones permutativas, aunque su implementación para bases de datos territoriales muy extensas está aún cada vez más cerca de conseguirse satisfactoriamente.

Todas estas cuestiones ponen, por tanto, de manifiesto la necesidad de una adecuada formación para llevar a cabo cualquier análisis de datos espaciales, sobre todo, en el ámbito micro-territorial. Como se ha expuesto, quedan aún problemas técnicos o científicos sin una solución satisfactoria. Sin embargo, la demanda de información y análisis micro-territorial es cada vez mayor, por lo que debe estimularse la investigación interdisciplinar sobre en temas relacionados con esta materia: econometría y estadística espacial, geo-informática, geografía cuantitativa, geo-marketing, geopolítica, etc.

2.5 Análisis y cortes temporales de los modelos

En este trabajo se especifican modelos de generación de viajes considerando relaciones espaciales. En primer lugar se especifican dos modelos de regresión lineal múltiple (MLR), a partir de datos zonales. Para considerar las situaciones de aglomeración (clusters) en un área determinada se especifica un modelo de correlación espacial local que permite medir la auto-correlación espacial para la ubicación de cada observación. Los modelos son aplicados al área metropolitana de Santander (España) con el objetivo de obtener modelos avanzados de Generación de viaje en dicha ciudad.

Normalmente los modelos de generación de viajes son requeridos para analizar la operación del sistema en ciertos años representativos, que incluyen el presente (habitualmente llamado año base) y algunos años futuros, todos los cuales se denominan colectivamente cortes temporales. En cada uno de estos cortes temporales, es necesario contrastar la operación del sistema de transporte en una situación base de comparación (situación base) con la operación del sistema después de introducir modificaciones estructurales en sus características fundamentales (situación con proyecto).

Sin embargo, el modelo de transporte que aquí se considera es una herramienta de análisis de corto plazo, incapaz de simular el comportamiento del sistema de transporte a lo largo de un año, ni siquiera de un día. La operación del sistema en un día completo es un fenómeno cuya complejidad y dinamismo desborda las capacidades de las actuales herramientas de análisis.



En consecuencia, la metodología habitual analiza sólo algunos períodos representativos de un día típico y utiliza el modelo de transporte para simular la operación del sistema dentro de tales períodos. Los resultados por período se extienden al día completo y posteriormente son extrapolados para obtener el total anual.

Los cortes temporales futuros son simulados introduciendo al modelo las variables de entrada correspondientes de cada período, con los valores que se estima que tales variables tomarán en el futuro. Luego se simula la operación del sistema en los períodos definidos y se sigue el procedimiento descrito anteriormente.

La metodología plantea varias interrogantes respecto al contexto de validez de aplicación del modelo. En primer lugar, es claro que en el mejor de los casos, el modelo se calibra y valida con información actual, lo cual significa que aún un buen modelo sólo garantiza una buena simulación del comportamiento actual del sistema de transporte. Si el modelo está correctamente formulado y calibrado, debería también ser adecuado para predecir los cambios en la operación del sistema, por efecto de modificaciones estructurales de sus características (proyectos y políticas) si ellas se produjeran en el presente.

La llamada calibración del modelo tiene como principal objetivo capturar los patrones de comportamiento de los distintos elementos que interactúan en el sistema, entre los cuales (y muy importantes) están los usuarios. La hipótesis básica de que tales patrones de comportamiento permanecerán constantes en el futuro, se hace claramente discutible a medida que se consideran escenarios más distantes en el tiempo. Por ejemplo, no es posible garantizar que las tasas de generación de viajes para los hogares de ingreso medio con un automóvil, serán en 20 años más, las mismas que las actuales. Tampoco es posible saber si la valorización que los usuarios de distintos niveles de ingreso dan al tiempo de viaje por ejemplo, permanece constante en el futuro.

Estos ejemplos reflejan un problema tradicional de cualquier modelo que trate de simular la operación futura de un sistema en que está involucrado el comportamiento de individuos. En el caso del modelo de transporte, el comportamiento de los usuarios es una de las variables fundamentales del sistema, por lo que la incertidumbre del futuro es una característica importante a considerar cuando se aplica el modelo. Es inmediato que mientras más cercano sea el horizonte temporal de aplicación, menor es la incertidumbre



respecto a la validez del modelo, debido a que se reducen las posibilidades de alteración de los patrones de comportamiento y por lo tanto, de los parámetros que fueron calibrados para el modelo.

De la discusión anterior se concluye la conveniencia de que los cortes temporales de análisis estén lo más cercanos posible del año de calibración. Habitualmente el modelo de transporte es utilizado para simular escenarios a diez, quince y hasta veinte años plazo, lo que parece revelar un exceso de confianza en su capacidad predictiva. Estos horizontes tan lejanos son necesarios para la evaluación de proyectos, pero es posible limitar estos requerimientos para evitar entregar una responsabilidad excesiva al modelo.

En términos generales, es aconsejable que el corte temporal máximo no se proyecte más allá de diez años plazo (y en lo posible menos de diez años). Adicionalmente debe definirse un corte temporal intermedio entre el año base y el corte temporal máximo.

Debe tenerse presente además, que la definición de los cortes temporales está fundamentalmente relacionada no sólo con el contexto de validez conceptual del modelo, sino también con la posibilidad de estimar correctamente (a futuro) sus datos de entrada.

2.5.1 Períodos de análisis

Tradicionalmente la modelación de los viajes generados se ha limitado a definir dos períodos básicos de análisis: período punta de la mañana y período de fuera de punta.

Normalmente, cada uno de estos períodos se prolonga entre una y dos horas, espacio de tiempo dentro del cual se supone que todos los viajes que se producen en algún origen, llegan a su destino. Además, debe considerarse que la definición de estos períodos debe incluir todos los horarios que se definan al incorporar un modelo de elección de horario de viaje. El análisis de estos períodos representativos arroja ciertos resultados operacionales (matrices de viaje por modo, niveles de servicio, flujos por arco de cada red) que son valorados económicamente para efectos de evaluación, obteniéndose finalmente costos y beneficios por período. Dichos costos y beneficios son posteriormente extrapolados para obtener totales diarios y anuales.

Es indudable que con sólo dos períodos representativos, la extensión al total diario resulta menos realista que lo deseable y debido a ello, algunos estudios han aumentado el



número de períodos de análisis, incluyendo punta del mediodía, punta de la tarde, fuera de punta de la mañana y de la tarde, etc. Desafortunadamente, la definición del número de períodos no es sólo una cuestión de más modelos y de más información de calibración y validación. Un problema adicional, se presenta cuando se considera la efectiva capacidad de los modelos disponibles para simular realísticamente el comportamiento del sistema de viajes (MIDEPLAN – FDC, 2010).

Este problema se manifiesta principalmente en una de las tareas básicas del modelo: las estimaciones de la demanda. Los modelos de demanda con mayor fundamento conceptual son aplicables a aquellos viajes que se originan en el hogar, entre otras cosas porque la mayoría de estos viajes son habituales y autónomos. Es decir, para los viajes originados en el hogar, las decisiones de los usuarios son relativamente típicas (a dónde viajar, con qué propósito, en qué modo, por cuál ruta), las alternativas de elección son también relativamente claras y en general las opciones de los usuarios no están condicionadas por decisiones tomadas en períodos anteriores. Estas dos características de habitualidad y autonomía, unido a la mayor facilidad de obtener información de los usuarios y de su comportamiento para este tipo de viajes, hacen que la tarea de modelarlos sea más abordable.

Considérese en cambio, el problema de modelar los viajes de punta tarde. En este caso la habitualidad de los viajes es mucho más difusa (distintos destinos de viaje en distintos días: regreso al hogar, diversión, compras, social, etc.); se presentan fenómenos complejos de explicar y simular (por ejemplo los viajes concatenados: origen en el lugar de trabajo, destino intermedio con propósito compras y destino final en el hogar); y la autonomía de los usuarios puede estar condicionada por decisiones de períodos anteriores, lo que tiene obvias implicancias para explicar su comportamiento en el período punta de la tarde (por ejemplo, si un usuario eligió automóvil para viajar al trabajo en la mañana, en la práctica no tiene alternativa modal para su viaje de regreso en la tarde y ningún modelo podría predecir adecuadamente su comportamiento modal, a menos que fuera informado de la decisión de la mañana y se condicionara exógenamente su predicción).

En términos de oferta, un mayor número de períodos, básicamente significa un mayor esfuerzo de definición y calibración de las redes involucradas, aunque conceptualmente no existen problemas mayores (algunas dificultades de modelación podrían presentarse no obstante, si existieran alternativas multimodales de viaje). En



resumen, se puede concluir de la discusión anterior que el actual estado de desarrollo de los modelos (particularmente los modelos de demanda) hace poco aconsejable incluir en el análisis, períodos en los cuales el comportamiento del sistema es aún demasiado complejo para las herramientas modelísticas disponibles.

En consecuencia, en el contexto de esta metodología, parece razonable considerar (al menos para el tratamiento con el modelo) solamente los dos períodos tradicionales: punta mañana y fuera de punta.

Un primer elemento básico para el modelo, que requiere una suerte de calibración, es la definición horaria y la extensión de cada período de análisis (punta de la mañana y fuera de punta). A este respecto no es posible entregar una recomendación general, porque las definiciones mencionadas dependen de cada ciudad particular.

Sin embargo, la Encuesta Origen-Destino de Viajes (EOD) provee la información necesaria para definir los períodos: a partir de los datos de la hora media de realización de los viajes (promedio entre hora de salida y hora de llegada) es posible construir histogramas en intervalos de 15 minutos (cuartos de hora) de viajes totales, viajes en transporte público y viajes en transporte privado. En estos histogramas se pueden identificar las variables que influyen en los horarios de mayor y menor demanda de viajes durante el día (horas punta y no punta).

El período de punta mañana generalmente está localizado entre 7:00 y 9:00 horas y su extensión normal es de una hora (o una hora y media). Este período de una hora de extensión, debe ser elegido de manera que queden incluidos dentro de él, al menos los dos cuartos de hora más cargados (entre 7:00 y 9:00 horas) de cada uno de los histogramas de viajes totales, viajes de transporte público y viajes de transporte privado.

La definición del período fuera de punta, tiene una connotación metodológica distinta a la del período de punta mañana. A diferencia de este último, cuya definición está temporalmente asociada con el inicio de actividades que generan viajes muy concentrados en el tiempo (principalmente trabajo y estudio), en el caso de fuera de punta tal asociación no existe. A lo largo de un día típico, existen muchas horas que pueden ser consideradas fuera de punta, en el sentido de que no son horas punta y de que los viajes allí materializados no tienen un horario obligado ni común para todos los viajes.



Otro problema que dificulta la definición, está relacionado con la no habitualidad temporal de los viajes fuera de punta dentro de un día típico (por ejemplo, una dueña de casa que viaja todos los días al mercado a hacer sus compras, pero a distintas horas cada día).

Dado que los viajes con este tipo de características son comunes en fuera de punta, es claro que cualquiera sea la definición temporal del período, se corre el riesgo de subestimar tales viajes (es decir, que el período no sea representativo). Desde este punto de vista, la definición estricta de una hora para representar el período fuera de punta, es necesariamente cuestionable cualquiera sea la hora elegida.

Esta discusión sugiere un enfoque distinto al tradicional para definir el período de fuera de punta. Este consiste en definir un período de una hora de extensión para fuera de punta, pero la especificación precisa de la hora de inicio y término de dicho período queda indeterminada. En otras palabras, se reconoce el hecho de que aunque existen varias horas de fuera de punta durante el día, ninguna de ellas es suficientemente representativa del resto, y por lo tanto una hora promedio de todas las horas fuera de punta es la mejor opción.

Este enfoque tiene la ventaja de permitir el mejor aprovechamiento de los datos de la EOD para efectos de la calibración del modelo de viajes, ya que se incluye en el análisis una mayor cantidad de viajes que efectivamente se producen en fuera de punta y cuyos patrones de comportamiento interesa determinar.

Para identificar las horas que pueden ser consideradas fuera de punta, se debe excluir del mencionado histograma de viajes totales, los períodos estimados como punta de mañana, punta tarde y punta mediodía, junto con los intervalos inmediatamente anteriores y posteriores (debido a que se está tratando de aislar la influencia de las horas punta sobre la fuera de punta). Además, se excluyen aquellas horas donde no existe actividad de viajes o ésta es mínima, tal como las horas de madrugada.

Desde el punto de vista del modelo que sólo modela una hora promedio de fuera de punta la suposición implícita es que en cualquiera de estas horas el sistema de viajes se comporta de manera semejante. De hecho, el modelo recibe como datos, los vectores de orígenes y destinos para una hora, que pueden ser calculados como el promedio



ponderado de los viajes realizados en cada propósito para el lapso horario mayor y el lapso de una hora seleccionado.

Resulta deseable que la definición horaria de los períodos punta y fuera de punta sea contrastada con el histograma que se deduce de los conteos continuos realizados. Dicho análisis debe ser realizado en forma cuidadosa ya que el concepto de hora media de viaje usado para construir el histograma obtenido a partir de la EOD es distinto al concepto de un conteo en un punto particular. Para reducir la distorsión espacial y temporal que se puede producir, se recomienda construir el histograma para una envolvente de los puntos de conteo continuo. La envolvente puede construirse como la suma vertical (en todos los puntos), del flujo en cada lapso temporal.



CAPÍTULO 3

Metodología



3 METODOLOGÍA

En este capítulo se desarrolla lo relacionado al Modelo de Regresión Lineal Múltiples, teniendo en cuenta sus variables, sus parámetros y los resultados de todos los demás modelos con que se ha trabajado; el análisis y la teoría relacionada a los modelos espaciales tan Espacial Lag como Espacial Error, los cuales se suelen tratar por alguno de los siguientes métodos: Máxima Verosimilitud, Estimación de las Variables, etc. como estructuras principales a modelizar; que son los objetivos de éste trabajo, basados en regresiones genéricas como método de estimación estándar.

3.1 Modelos de producción de viajes

El objetivo de la producción de viajes consiste en estimar el número de viajes que se generan en cada zona del área de estudio, generalmente correlacionada con las características socioeconómicas de los residentes en la zona (Ortuzar, 2000). De un modo, la generación es suma de los extremos hogar de los viajes basados en el hogar (BH), más los orígenes de los viajes no basados en el hogar (NBH).

Es interesante hacer notar que para algunos propósitos, la tendencia creciente según ingreso y motorización es estricta. Esto es especialmente cierto en el propósito trabajo, lo cual en teoría es razonable.



En vista de que solo el 5% de los viajes de los residentes no son basados en hogar, las variables socioeconómicas utilizadas para el ajuste de los modelos de generación están estrictamente relacionadas con la población de las zonas de transporte y las características de los hogares.

Con la información obtenida de los usos del suelo y la relación observada entre generación de viajes (detectada en las encuestas origen y destino) se obtiene la relación en la que se fundamenta el método.

Para cada una de las zonas del área estudiada se detectan las superficies comprendidas por cada tipo de uso del suelo y se calculan los extremos de viajes correspondientes.

Dicho de otro modo, las producciones de viajes se diferencian en tres tipos principales: generación de viajes basados en el hogar tanto de ida como de retorno, y la de viajes no basados en el hogar. Los primeros serán estimados mediante modelos de Análisis de Clasificación Múltiple (ACM), en tanto que para los segundos y terceros se puede utilizar modelos de regresión lineal múltiple (RLM). Lo que es lo mismo, los orígenes de una zona se muestra a continuación:

$$O_i^{pn} = O_{i(bhi)}^{pn} + O_{i(bhr)}^{pn} + O_{i(nbh)}^{pn} \quad (27)$$

Donde:

- O_i^{pn} = Número total de viajes con propósito p, producidos por usuarios de la categoría n, en la zona i.
- $O_{i(bhi)}^{pn}$ = Número de viajes basados en el hogar de ida (bhi), para la misma clasificación anterior.
- $O_{i(bhr)}^{pn}$ = Número de viajes basados en el hogar de retorno (bhr), para la misma clasificación anterior.
- $O_{i(nbh)}^{pn}$ = Número de viajes no originados en el hogar para la misma clasificación anterior.

Esta diferencia es metodológicamente significativa por los motivos siguientes. En primer lugar, la importancia de cada tipo de viaje se ve influenciado en función del período de modelización. Es así como, los viajes basados en el hogar de ida se realizan



principalmente en las horas punta de la mañana. En segundo lugar, la generación de los viajes basados en el hogar de ida se explica por las variables socioeconómicas asociadas al lugar de origen del viajero. Por consiguiente, la generación de viajes no basados en el hogar y basados en el hogar de retorno se describe por aquellas variables asociadas a las actividades desarrolladas en las zonas. En el período fuera de punta, los dos últimos tipos de viajes son de suma importancia.

3.1.1 Generación de viajes basados en el hogar de ida (bhi)

La metodología más empleada confía en la precisión de algunas tasas de generación de viajes por hogar y en el conocimiento del número de hogares en una determinada área para los viajes originados en el hogar. Dichas tasas deben ser destacadas dentro de cada período de análisis, para cada propósito de viaje y categoría del hogar según el matiz de la demanda.

$$O_{i(bhi)}^{pn} = H_i^n \cdot t^{pn} \quad (28)$$

Donde:

$O_{i(bhi)}^{pn}$ = Número de viajes con propósito p generados por los hogares de la categoría n de la zona i.

H_i^n = Número de hogares de la categoría n en la zona i.

t^{pn} = Tasa de generación de viajes con propósito p de los hogares de la categoría n.

Este modelo requiere conocer el número de hogares por categoría en cada zona, lo cual debe ser determinado o estimado a partir de información socioeconómica independiente: normalmente del Censo Poblacional o de otros catastros urbanos. Además se debe recordar, que es imprescindible conocer la distribución de hogares por categoría no sólo en el año base de análisis; también se necesita la distribución futura de los hogares para cada uno de los cortes temporales. Dichas proyecciones forman parte del ámbito de especialización de otras disciplinas, por lo que relacionado con la metodología, la distribución de hogares por categoría se examina como un dato exógeno.

El problema ya se restringe a encontrar las tasas de generación de viajes para cada propósito y categoría de hogar. Dicha tarea ha sido generalmente realizada con los llamados modelos de **análisis por categorías**, que diagnostican las tasas de generación



requerida a partir de una muestra de hogares, sencillamente basta por dividir cada categoría ingreso-tasa de motorización, por el número de viajes observados de un propósito por el número de hogares en la muestra.

Debido a sus ventajas teóricas y su consideración, la aplicación práctica de este modelo presenta un menudo graves problemas. El primero de ellos se relaciona con la descripción de las tasas calculadas para las categorías extremas, donde el número de hogares en la muestra puede ser muy pequeño (hogares de bajos ingresos con dos o más autos); otro problema más frecuente es la obtención de tasas contra improvisadas. A todo esto se debe añadir los obstáculos para calificar la bondad estadística de estos modelos.

3.1.2 Generación de viajes basados en el hogar de retorno (bhr)

En este tipo de viajes (no habitual en el período punta mañana y más común en el período fuera de punta), la generación debe ser conformada con regresión lineal múltiple a nivel zonal, Dicho tipo de viajes juega un papel importante en función de las variables asociadas con el uso de suelos y las actividades de una zona. De esta forma, las variables explicativas del modelo RLM (en la generación) serán casi las mismas utilizadas por los modelos de RLM de atracción de viajes.

A diferencia de los modelos de atracción de viajes, la generación de viajes debe clasificarse por categoría de demanda, de forma que se someta a dos alternativas. La primera consiste en calibrar un modelo RLM por categoría, mientras que la segunda consiste en calibrar un modelo RLM que no distingue categorías (modelo conjunto) y aplicar posteriormente factores que representen adecuadamente la proporción de cada tipo de usuarios.

Siendo la primera opción la más deseable, su uso y grado de fiabilidad sigue siendo limitado al número de viajes contemplados por categoría demandada. Debido a ello, en este método podrán aparecer dificultades en la calibración y seguramente será complicado conseguir modelos de RLM estadísticamente contundentes. En la segunda opción, es necesario saber porcentualmente el número de viajes basados en el hogar de retorno, generados de acuerdo a la zonificación según la clasificación demandada. Dicha información es difícil de obtener, salvo que se disponga de un banco de datos como la Encuesta Origen Destino.



3.1.3 Generación de viajes no basados en el hogar (nbh)

Como en el apartado anterior, esta clase de viajes debe ser configurada con regresión lineal múltiple a nivel zonal. La generación de estos viajes dependerá de las variables asociadas con el uso de suelos y las actividades de una zona. Por ello, se utilizarán esencialmente las mismas variables explicativas, tanto para los viajes generados como para los viajes atraídos del modelo RLM.

Contrariamente, en atracciones de viaje, la generación de viajes se clasificará por categoría de demanda, de tal forma que se calibre el modelo RLM para cada categoría. Esto podría generar problemas de calibración, ya que concretamente en el periodo punta de mañana, el número de viajes no originados en el hogar podría ser excesivamente reducido. Si además se suma a este hecho que la reducción del número de viajes se debe diferenciar por propósito y categoría, se sobreentiende que estadísticamente será más complicado obtener modelos RLM porcentualmente más consistentes. También es necesario estimar propósito, categoría y el porcentaje de viajes originados y no originados en el hogar por zonas. Dichos datos son difíciles de obtener, salvo que se disponga de un paquete de datos como Encuesta Origen-Destino.

3.2 Modelos de Atracción de Viajes

Tanto en la producción de viajes basados en el hogar de ida, como en la atracción de los viajes basados en el hogar de retorno se utiliza el análisis de clasificación múltiple (MIDEPLAN – FDC, 2010).

Habitualmente, la técnica más utilizada para modelar este tipo de viaje es a través de modelos de regresión lineal múltiple, ya que posibilita mejor ajuste y la obtención de indicadores estadísticos de la bondad del modelo.

De igual modo que en la producción, la atracción es la operación de sumar los extremos no hogar de los viajes basados en el hogar (BH), más el destino de los viajes no basados en el hogar (NHB). La diferencia radica que las variables que explican los desplazamientos para los modelos de producción de viaje, van en función del motivo del viaje, pudiendo ser distintos para los desplazamientos obligados por cuestiones laborales y estudio que los no obligatorios por otras causas (trámites, compras, salud, ocio, restauración y asuntos personales).



Para efectos de calibración de modelos de atracción de viajes, se recomienda tener en cuenta, al menos, estas dos opciones:

- Para los viajes atraídos basados en el hogar de ida (bhi) y viajes no basados en el hogar (nbh) las variables explicativas generalmente tienen correspondencia con equipamientos por zona, destinados a cada actividad y no los hogares (el destino de este viaje no es el hogar). Como se ha expuesto anteriormente respecto a la atracción de viajes, con excepción de los modelos de regresión lineal, no se hallan prácticamente alternativas metodológicas de análisis. Por ello, para esta clase de modelos se calibrará para cada zona el propósito y periodo de análisis definidos usando técnicas de regresión lineal múltiple (RLM).
- Para modelar la atracción de viajes basados en el hogar de retorno (bhr). Ya que el destino del viaje es el hogar, la única variable explicativa posible será el número de hogares por zona. En este caso, será posible usar la técnica de regresión lineal simple (RLS). Además será posible tener en cuenta modelos de tasas en Análisis de Clasificación Múltiple (ACM) de atractividad.

La diferencia expuesta anteriormente se recomienda para modelos que corresponden al periodo fuera punta, ya que en éste se comprueba un número importante de viajes basados en el hogar de retorno (bhr). Como se expuso anteriormente, en las situaciones del período punta mañana, los viajes de esta clase son limitados. Por esto no se recomienda que se separen. Es evidente que la atracción total de viajes es igual a la suma de los resultados de los dos modelos.

Para los modelos de RLM, la expresión típica de un modelo de atracción, para un propósito determinado, es la siguiente:

$$D_j^p = \theta_0 + \sum_k \theta_k \cdot X_{jk} + E_j \tag{29}$$

Donde:

D_j^p = Número de viajes con propósito **p** atraídos por la zona **j**

θ_k = Parámetros de calibración

X_{jk} = Variables explicativas zonales



E_j = Error de la estimación para la zona j

Se debe tener en cuenta las hipótesis básicas para que un modelo de RLM esté bien especificado y son las siguientes:

- I. El valor esperado del error es cero, con $i \neq j$.
- II. La varianza del término del error es constante, con $i \neq j$.
Esta propiedad se conoce como “homocedasticidad”.
- III. No existe correlación, es decir, el término del error de una observación no se encuentra correlacionado con el término del error de otra observación. Las dos últimas propiedades implican que la matriz de varianzas y covarianzas del error se puede escribir como el producto entre la varianza del error por la matriz identidad:

$$\sum (\varepsilon) = \sigma^2 \cdot \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ \dots & & & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma^2 \cdot I \quad (30)$$

- IV. Las variables explicativas toman un valor fijo o determinístico. Es decir, ellas no corresponden a un valor aleatorio (esto no impide que ellas pudieran ser valores fijos producto de un proceso aleatorio).
- V. Las variables explicativas no están correlacionadas con el término de error.
- VI. Las variables explicativas no presentan una relación lineal exacta entre sí. La existencia de este problema se denomina normalmente como “colinealidad” y produce problemas de estimación al no poder cuantificar de manera exacta la influencia de cada variable incluida en el modelo.
- VII. Las variables explicativas son medidas sin error. Es decir, la medición de ellas no incorpora un error adicional al modelo, producto de fallas en toma de los datos, etc.
- VIII. En el modelo no se excluyen las variables relevantes y tampoco se incluyen las variables irrelevantes, a la hora de explicar el comportamiento de la variable endógena.
- IX. Los parámetros poblacionales se mantienen constantes a lo largo de toda la muestra. Esta condición se conoce como hipótesis de permanencia



estructural, lo que significa que no hay variación en la “influencia” de una variable explicativa sobre la variable explicada.

Lo correcto sería estimar un vector de orígenes y un vector de destinos para cada propósito y categoría de demanda, pero debido a que la definición de categoría de demanda se asocia a las características de los hogares (niveles de ingresos y tasas de motorización), las atracciones de viajes son difícilmente clasificables por categorías, de tal manera que no se manifiestan en los hogares en los intervalos de tiempo apreciados (punta mañana y fuera de punta).

Conviene señalar que los modelos deben cumplir las siguientes condiciones:

- 1) Los modelos deben ser de tipo “predictivos”, más que explicativos. Este punto es clave, refiriéndose a que las variables explicativas deben ser fácilmente predecibles en el tiempo.
- 2) Los modelos deben estar correctamente “especificados”. Esto significa que, tanto las variables explicativas como la variable explicada deben estar estrechamente relacionadas.
- 3) Los modelos deben tener indicadores estadísticos razonables y signos adecuados.

Estos modelos estiman el número de viajes atraídos por una zona, suponiendo una relación lineal de esta variable con ciertas características de la zona. Habitualmente, estas características se refieren a la infraestructura existente en la zona, en términos de las actividades relevantes según los propósitos de viajes que son:

- Viajes con propósito **trabajo**: Estos son atraídos por las actividades que ofrecen empleos. Las más relevantes son el comercio, oficinas, servicios y la industria.
- Viajes con propósito **estudio**: Son atraídos por la presencia de establecimientos educacionales (número de matrículas por nivel de educación: básica, media y superior).
- Viajes con **otros** propósitos: Corresponden a los viajes de compras, diligencias, y salud entre otros. Las actividades relevantes serán el comercio, los servicios y las atenciones de salud. Además, para incluir el



efecto de los viajes con motivos sociales suele incluirse como variable explicativa al total de hogares existente en una zona.

Para obtener los valores de estas variables, habitualmente se recurre a distintas versas fuentes oficiales (el INE o Ministerios – Consejerías). Lo ideal sería que el modelo determinase el número de viajes atraídos por zona, para cada período, propósito y categoría de demandada. Pero, debido a que en el caso de las atracciones de viajes no se conocen otras formas de deducir el fenómeno de forma más desagregada, normalmente se debe considerar en el análisis del modelo de forma global, tanto cada zona como sus características. Este hecho hace que no sea tan fácil clasificar las atracciones por categoría de demanda.

3.3 Modelo de Regresión Lineal Múltiple (RLM)

Dentro de la disciplina de la Estadística, los Modelos de Regresión constituyen un área de investigación clásica utilizados con frecuencia desde hace muchos años. Estos Modelos de Regresión estudian la relación estocástica cuantitativa entre una variable de interés y un conjunto de variables explicativas. De ahí que podamos decir que el modelo de regresión lineal múltiple y el modelo lineal simple sean idénticos con la única diferencia o salvedad de que en la primera aparecen un mayor número de variables explicativas.

Por tanto, el método de regresión lineal múltiple es un procedimiento estadístico en el que se establece una relación lineal entre una variable dependiente y variables independientes o explicativas, de acuerdo a la siguiente expresión:

$$T = k + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_n \cdot X_n \quad (31)$$

Donde T es la variable dependiente (viajes producidos o atraídos por una zona), X_1 a X_n son las variables independientes, k es el término independiente que representa la parte de la variable dependiente no explicada por las variables independientes, y b_1 a b_n son los coeficientes de la regresión lineal.

Mediante software estadístico se determinan los parámetros de la ecuación de regresión, es decir los coeficientes de regresión y el término independiente. Pero además de los parámetros mencionados se debe conocer también el grado de ajuste de la ecuación



de regresión con respecto a los datos. Dicho grado se mide a través de los siguientes indicadores estadísticos:

- R^2 (coeficiente de determinación), cumpliéndose que $0 \leq R^2 \leq 1$. Este indicador representa la proporción en que las variables independientes explican la variabilidad de la variable dependiente. Cuanto más próximo a la unidad, mayor grado de ajuste.
- Error estándar de la estimación:

$$S_T = \sqrt{\frac{\sum_i (T_i - T_{est})^2}{N - m}} \quad (32)$$

Donde:

T_i = valores observados de la variable dependiente (viajes zonales).

T_{est} = valores de T_i calculados mediante la ecuación de regresión.

N = número de casos.

m = número de parámetros estimados de la ecuación de regresión (coeficientes y término independiente).

Como resultado, el error estándar de la estimación (S_T) debe ser lo más pequeño posible.

Así mismo deben ser sometidos a tests de hipótesis del estadístico “t” los valores obtenidos de los parámetros de la ecuación de regresión (b_1 a b_n) a través de la siguiente fórmula:

$$t = \frac{b_j}{S_{bj}} \quad (33)$$

Siendo S_{bj} el desvío estándar del coeficiente. Para que el valor del coeficiente b_j sea significativo, es decir que la variable X_j tenga importancia en la explicación, se debe cumplir que t sea mayor que 2, de otro modo es conveniente eliminar la variable de la ecuación.

Tajantemente el término independiente k debería ser cero, ya que representa la porción de los viajes que no dependen de ninguna variable explicativa. Si dicho valor es mayor con respecto al valor medio de T , la ecuación debería ser destacada. Por lo



contrario si el valor de k es menor, se puede re-estimar la función obligándole a pasar por el origen.

De acuerdo con su facilidad de proyección y su disponibilidad, las variables independientes a considerar deberían ser seleccionadas. En general son variables que representan las características socioeconómicas de cada zona.

3.4 Los modelos espaciales

En el ámbito de la econometría espacial, se procede a comentar, sin ánimo de exhaustividad, los principales modelos de corte transversal que incorporan la existencia de dependencia espacial, tomando como referencia el artículo de Elhorst (2010) en el cual se destacan las aportaciones del libro de LeSage y Pace (2009).

El modelo más completo e inicial es el ofrecido por Manski que incorpora tanto la matriz de variables explicativas modificadas espacialmente como la variable dependiente. De mismo modo, se recoge a su vez dicha modificación espacial en el término de error.

$$\begin{aligned} Y &= \rho WY + X\beta + WX\theta + u \\ u &= \lambda Wu + \varepsilon \end{aligned} \tag{34}$$

Según Manski (1993) existen tres tipos diferentes de efectos interacción espacial que explican por qué el valor de una variable en una zona específica depende de lo que ocurre en otras zonas. El primer tipo es una interacción endógena de forma que el comportamiento de la variable a explicar en una unidad espacial depende del comportamiento de esa misma variable en otras unidades espaciales (WY), el segundo engloba interacciones exógenas, donde el comportamiento de la variable a explicar depende del comportamiento de las variables explicativas en otras unidades espaciales (WX) y para terminar, se determinan los efectos de interacción entre las perturbaciones aleatorias de las diferentes unidades espaciales (Wu).

Los modelos más utilizados son el **espacial error** y el **espacial lag**. La diferencia principal entre ambos modelos radica en la forma en la que se especifica la dependencia espacial, si la dependencia espacial está incluida en el término de error (espacial error) o si es la variable dependiente la que se modifica espacialmente (espacial lag).



Para estudiar la dependencia espacial tras comprobar otras características destacables es importante aplicar modelos multivariados de regresión espacial autorregresivo. Más importante aún, si las medidas exploratorias desvelan la presencia de dependencia espacial, entonces debemos aplicar dichos modelos, porque ignorarla podría comprometer la fiabilidad de nuestro análisis (Anselin, 1992a).

Según Anselin (1992a) existen dos tipos de dependencia espacial: una es ruido o ajena para nuestro análisis y la otra es sustantivamente relevante. La primera está en el término del error y viola el supuesto de la no correlación entre los errores en la regresión de mínimos cuadrados. Si se ignora la auto-correlación del error los estimadores de nuestra regresión serán ineficientes, aunque no sesgados (Anselin, 2002). Se considera como ruido porque los coeficientes del modelo no se ven afectados y solo nos interesa para mejorar nuestras estimaciones. La auto-correlación en el término del error puede ser utilizada a través de un **Modelo Espacial Error**, de esta forma, se supone que la dependencia espacial encontrada en la variable dependiente es resultado de la distribución geográfica de nuestras variables explicativas y de la auto-correlación del término del error (Anselin, 2002; Baller et al., 2001).

3.4.1 Modelo espacial lag

La auto-correlación espacial substantiva se presenta cuando el valor que toma la variable dependiente en cada unidad geográfica está realmente determinado por el valor que dicha variable toma en las unidades vecinas. Es decir, la auto-correlación estará presente después de controlar otras variables explicativas. Los coeficientes estimados con la regresión de mínimos cuadrados estarán sesgados si no se tiene en cuenta ese tipo de dependencia espacial, con los consecuentes problemas de la estimación y el signo de los coeficientes, la significancia de los mismos y la bondad de ajuste del modelo.

El **Modelo Espacial Lag** es la opción para modelar este tipo de auto-correlación, que considera la dependencia espacial introduciendo una variable espacial. De acuerdo con Baller et al. (2001), un Modelo Espacial Lag representa la relación interactiva entre la variable dependiente y las variables independientes en las unidades vecinas.

La especificación resultante en el caso del **modelo espacial lag** es la siguiente:

$$Y = \rho WY + X\beta + \varepsilon \tag{35}$$



Donde:

WY = es la variable dependiente modificada o retardada espacialmente

ρ = es el parámetro espacial auto-regresivo

β = es un vector de coeficientes asociados al resto de variables explicativas recogidas en la matriz X .

Ésta ecuación nos permite entender mejor cómo se estima un modelo de rezago espacial como un complemento a la explicación anterior, donde ρ es el parámetro espacial de rezago que se estima, W es la matriz de pesos geográficos, X es la matriz de covariante, y ε es el vector de errores no correlacionados.

Los modelos de retardo espacial son similares a los modelos de variable dependiente retrasada de auto-regresión en análisis de series de tiempo, pero el problema es que el coeficiente de correlación no se puede estimar fácilmente (Anselin, 2003). Eso es un problema porque para estimar el coeficiente, se necesita una matriz de pesos espaciales, pero a menudo no está claro lo que la matriz debe ser similar, es decir, cuál es la relación espacial real.

3.4.2 Modelo espacial error

Por otro lado la auto-correlación en el término del error puede ser manejada a través de un Modelo Espacial Error (estimada con máxima verosimilitud). Este modelo asume que la dependencia espacial encontrada en nuestra variable dependiente es resultado de la distribución geográfica de nuestras variables explicativas y de la auto-correlación del término del error, esto último sugiere que nuestro modelo no está bien especificado (Anselin 1992a, Baller et al 2001).

Surge auto-correlación espacial error si los términos de error están correlacionados a través de sus observaciones, es decir, el error de una observación afecta a los errores de sus vecinos. Es similar a la correlación serial en el análisis de series de tiempo y deja de Mínimos Cuadrados Ordinarios (MCO) coeficiente imparcial, pero las hace ineficientes (Anselin, Syabri and KHO, 2004). Debido a que es un problema tan molesto, los errores espaciales también se llama "dependencia molestia en el error."

La especificación resultante en el caso del **modelo espacial error** es la siguiente:



$$\begin{aligned} Y &= X\beta + u \\ u &= \lambda Wu + \varepsilon \end{aligned} \tag{36}$$

Donde λ es el parámetro espacial autorregresivo que acompaña al retardo espacial del error (Wu). Se asume que ε no está correlacionado y es homocedástico.

Como se puede observar en ésta ecuación, el modelo es similar a un modelo de regresión estándar, pero se estima el parámetro λ que calcula el grado de auto-correlación de los errores, dada una matriz de pesos geográficos (W):(Anselin 2001, Anselin 2002)

Tanto la variable dependiente como las independientes han sido consideradas en términos logarítmicos para facilitar la interpretación de los resultados.

Hay una serie de casos en los que puede surgir de error espacial. Por ejemplo, similar a lo que puede suceder en la serie de tiempo, una fuente de correlación puede provenir de las variables no medidas que se relacionan a través del espacio. La correlación también puede surgir de la agregación de variables espacialmente correlacionadas y error de medición sistemática.

Entonces, ¿qué hacer si hay una buena razón para creer que no hay error espacial? Tal vez la prueba más famosa es I de Moran que se basa en los residuos de la regresión y también está relacionada con la dispersión de Moran de los residuos que se pueden utilizar para detectar el problema de forma gráfica. Hay otras estadísticas como multiplicador de Lagrange y las pruebas de coeficiente de riesgo, y cada uno de ellos tiene diferentes maneras de llegar al mismo problema. Si hay una buena razón para creer que el error espacial es un problema, entonces el camino a seguir es o modelar el error directamente o utilizar métodos autoregresivos.

Un Modelo Espacial Error (MEE) considera la dependencia espacial y mejora la eficiencia de nuestra estimación.

3.4.3 Método de estimación

La estimación de los modelos espaciales con las variables dependientes se ha llevado a cabo mediante la vía por Máxima Verosimilitud (MV) que ha sido una de las principales vías usadas para este trabajo como se ha visto en el capítulo 5, se asume



normalidad de los errores en este casos y la batería de estadísticos se aplica principalmente al método de MV.

1. Método Máxima Verosimilitud (MV)

En general la teoría de la estimación de los parámetros por el método de máxima verosimilitud es útil tanto para modelos lineales como no lineales. Para ello se presentan pruebas estadísticas para la comparar los parámetros en distintos modelos. El método de máxima verosimilitud aplicado a un modelo paramétrico, consiste en asignar un valor al conjunto de parámetros tal que los datos observados adquieran la máxima probabilidad de haber resultado bajo el modelo asumido. Por esta razón se ha realizado una correcta estimación de los parámetros con objetivo de aportar una herramienta para poder comparar las curvas de crecimiento.

Las curvas de crecimiento y su comparación han sido un tema de gran importación para los autores como: (Rao, 1958; Allen, 1976; Gallucci y Quinn, 1979; Kimura, 1980; Kappenman, 1981; Pauly y Munro, 1984; Cerrato, 1990; Hansen et al., 1993 entre otros) que lo han ido ampliando con el paso de los años.

En este trabajo se ha utilizado la razón de verosimilitudes para probar las hipótesis sobre los parámetros y para los mismos construir sus límites de confianza con el fin más bien práctico para facilitar la aplicación de la misma en estudios de crecimiento. Tal finalidad no sería posible si los autores citados anteriormente no lo hubiesen tratado. Así se describe de una forma particular el método de máxima verosimilitud para la estimación de los modelos tanto de regresión lineal múltiple como espaciales.

Dados los supuestos de normalidad en los errores, se puede derivar una función de verosimilitud, es decir una función no lineal de los parámetros que deberá maximizarse.

Tests de contraste

Medidas de ajuste:

La tradicional R^2 no es aplicable a un modelo espacial, por lo que se proponen dos medidas más también llamadas pseudo R^2 : la primera es el ratio simple de la varianza de los valores predichos sobre la varianza de los valores observados de la variable dependiente, la segunda es la correlación al cuadrado entre los valores predichos y



observados. Sin embargo las medidas más adecuadas para estos modelos son las basadas en la función de verosimilitud, entre las que se encuentran: el Log likelihood (LIK), el criterio de información de Akaike (AIC) y el criterio de Schwartz (SC). El modelo con el mayor Log Likelihood, o con los AIC o SC más bajos serán los mejores.

Heteroscedasticidad:

Se podrían presentar dos estadísticos, uno es el Breusch-Pagan (BP) espacial basado en los residuos de la estimación de Máxima Verosimilitud y el otro BP basado en la regresión estándar del modelo y aunque este último no se ha sido considerado muy adecuado, en realidad en términos prácticos no hay mucha diferencia entre ellos.

Dependencia espacial:

En términos generales se tiene el test LR para el coeficiente autoregresivo espacial ρ . Este estadístico se distribuye como una χ^2 con un grado de libertad.

Dependencia Espacial (en los errores):

Para este caso se propone el ratio de verosimilitud (likelihood Ratio) LR para el coeficiente espacial autorregresivo λ . Se reportan los grados de libertad y las correspondientes probabilidades, de acuerdo a una distribución de χ^2 y los diagnósticos de dependencia espacial se han incluido sólo para estimaciones de MV que ha sido en este caso.

3.5 La matriz de pesos espaciales

La matriz de pesos espaciales es una generalización de la matriz de interacciones o contigüidades, por lo que suele ser designada también con estos nombres, aunque se la conoce habitualmente como matriz de pesos, ponderaciones, retardos o contactos espaciales. Hay autores que también la denominan “matriz de ponderaciones Cliff-Ord”, en honor a quienes la formularon por primera vez (Anselin, 1988a). El uso de esta matriz de pesos espaciales, a diferencia de la matriz binaria de interacciones, permite al investigador la elección del conjunto de ponderaciones que él considere apropiado para cada fenómeno, lo que supone una mayor flexibilidad en la definición de la estructura de interdependencias de un sistema regional y permite considerar cuestiones como las



barreras naturales o el tamaño de las regiones. Es más, cuando sea necesaria la consideración de hipótesis acerca del grado de vinculación existente entre áreas vecinas, deben utilizarse distintos conjuntos de ponderaciones que permitan contrastar dichas hipótesis (Chasco, 2003).

En el análisis de series temporales es usual utilizar un operador de retraso por el tiempo de ejecución, para capturar la influencia de las observaciones pasadas en la dinámica de la serie y el valor de la realización contemporánea. Otra razón por la que resulta relevante tiene que ver con la posibilidad de realizar pronósticos. De esta manera, la influencia del rezago temporal es unidireccional: las realizaciones pasadas afectan a las presentes, y harán lo propio con las futuras, conforme a una estructura dinámica. Ahora bien, en el marco del análisis espacial se establecen relaciones multidireccionales, por lo que resulta necesario construir una matriz que permita incluirlas de manera adecuada al análisis.

$$W = \begin{pmatrix} 0 & w_{12} & w_{1N} \\ w_{21} & 0 & w_{2N} \\ w_{N1} & a_{N2} & 0 \end{pmatrix} \quad (37)$$

La matriz de pesos espaciales (también denominada matriz de contactos o matriz de proximidad espacial) y simbolizada con W , es una matriz cuadrada de $N \times N$ (siendo N el número de unidades espaciales), no estocástica cuyos elementos (w_{ij}) reflejan la intensidad de la interdependencia entre cada par de regiones i, j , (Moreno y Vayá, 2000).

El valor que asume cada w_{ij} (ponderación) se basa en las siguientes determinaciones de adyacencias. De manera simplificada:

- ❖ $w_{ij} = 1$, si dos regiones son contiguas, o sea sería indicativo de existencia de una interacción espacial entre las observaciones i, j que podría ser expresada como simple contigüidad binaria (teniendo una frontera común), como contigüidad de distancias, a través de centroides, dentro de una banda de distancia mínima (considerándose contiguos dos puntos cuando estén situados a una distancia inferior a la prefijada como mínima), o como función inversa de la distancia simple o cuadrática.



- ❖ $w_{ij} = 0$ en caso contrario, o sea sería indicativo de ausencia de auto-correlación espacial entre las observaciones i, j (por convenio, los elementos de la diagonal principal de la matriz de pesos serán, como en la matriz de interacciones, igual a cero). Típicamente los elementos de la diagonal principal son cero, pues ninguna región puede ser vecina de sí misma. La Figura 9 ilustra un ejemplo hipotético:

		A	B	C	D	E
	A	0	1	1	0	0
	B	1	0	1	1	0
	C	1	1	0	1	1
	D	0	1	1	0	
	E	0	0	1	1	0

Figura 9: Ilustración de una matriz de pesos espaciales – Fuente: Baronio; Vianco; Rabanal; 2012

En algunos casos suele normalizarse la matriz W por filas. Esto es que, se divide cada elemento w_{ij} por la suma de fila a la que pertenece. De esta forma, la suma de los pesos asociados a cada área es igual a 1. Por otra parte, de acuerdo con Kapoor et. al. (2007) la matriz W debe ser delimitada de manera uniforme y en valor absoluto, lo que implica la existencia de una constante $c < \infty$ tal que

$$\overbrace{1 \leq i \leq N \sum_{j=1}^N |w_{ij}|}^{\text{máximo}} \leq c \text{ y } \overbrace{1 \leq j \leq N \sum_{i=1}^N |w_{ij}|}^{\text{máximo}} \leq c \quad (38)$$

para producir los resultados asintóticos exigidos por una estimación consistente.

La matriz de pesos espaciales desempeña un papel fundamental en la incorporación de las relaciones espaciales de las variables al modelo. En el contexto de una única ecuación, por ejemplo, es posible generar una matriz de rezago espacial de la variable endógena (WY), a partir de multiplicar la matriz W por un vector de variables endógenas Y , de orden $N \times 1$. Otra posibilidad consiste en incorporar otras variables espacialmente rezagadas, desde una matriz de variables X de orden $N \times k$, reflejado en las columnas de la matriz WX (Baronio; Vianco; Rabanal; 2012).



La especificación adecuada de los elementos de esta matriz w_{ij} , es uno de los puntos metodológicos más difíciles y controvertidos en la econometría espacial. Efectivamente, será necesario tener cuidado a la hora de elegir los pesos espaciales para evitar posibles correlaciones espurias. Los factores más importantes a tener en cuenta para determinar las ponderaciones dependerán de cada estudio en particular.

Por ejemplo, el grado de vinculación existente entre dos regiones puede depender de la distancia entre sus centros geográficos o demográficos, o también puede estar relacionada con la longitud de la frontera común de dichas regiones, etc (Stetzer, 1982, Anselin y Rey, 1991, Florax y Rey, 1995). Por ejemplo, en áreas urbanas, la relación entre dos zonas podrá depender de la frecuencia de los servicios de transporte público.

De otro modo, la especificación de todo modelo de econometría espacial requiere de una matriz de pesos espaciales, W , que exprese las relaciones de vecindad o interacción existente entre las unidades geográficas del espacio considerado. Son varias las especificaciones que han sido utilizadas en la literatura económica: desde las sencillas matrices binarias de interacciones espaciales, hasta matrices generales de distancias, basadas o no en puntos críticos de cortes. En el contexto general de las regiones europeas, suelen utilizarse las matrices de distancias, sobre todo para evitar que la existencia de regiones / provincias insulares deje las correspondientes filas / columnas de la matriz con ceros, lo que podría afectar a los resultados de la inferencia. La distancia geográfica es considerada como una buena variable “proxy” cuando no se dispone de variables de vecindad económica como la red de infraestructuras o de transportes entre provincias / municipios.

Siguiendo el criterio de la proximidad física, diversos autores han propuesto definiciones de la matriz W basadas en la utilización de la distancia entre regiones (Cliff y Ord, 1973, 1981; Bodson y Peeters, 1975), de manera que la intensidad de la interdependencia entre dos regiones disminuye conforme aumenta la distancia que separa sus respectivos centros. En este sentido, se recurre de forma habitual a la matriz inversa de distancias al cuadrado sugerida por Anselin (1980). Por último, cabe destacar que en los últimos años han sido propuestas matrices de pesos alternativas que tratan de huir del concepto de contigüidad o proximidad física para acercarse más al objeto de estudio analizado en cada caso. Así, cabe mencionar los trabajos de Case et al (1993) donde se define una matriz W basada en un concepto de distancia económica entre regiones o de

Vayá et al (1998a) y López-Bazo et al (1999), donde los pesos de W recogen el grado de intercambio comercial entre las regiones analizadas. En cualquier caso es preciso remarcar que, en el proceso de contrastación y estimación, la matriz de pesos ha de ser considerada exógena, de manera que se deberá asegurar la exogeneidad de W cuando se recurra a indicadores socioeconómicos para definir sus elementos (salvo que se considere explícitamente la endogeneidad de W en la especificación del modelo).

3.5.1 Criterios habituales de contigüidad (Baronio; Vianco; Rabanal; 2012)

Usualmente los criterios de contigüidad (o de vecindad) más utilizados son los que se resumen en la Figura 10, a continuación:

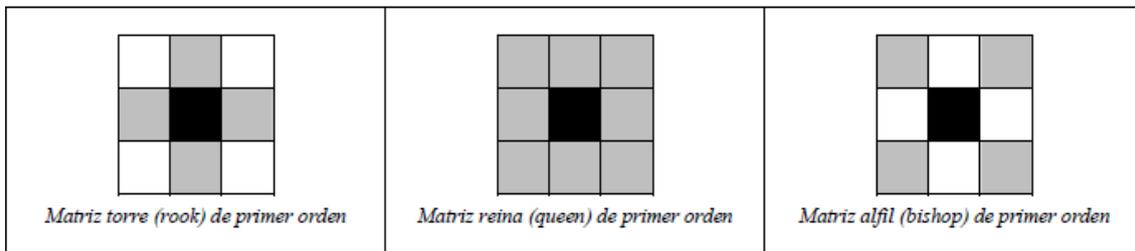


Figura 10: Criterios de contigüidad en la matriz de pesos espaciales. Retardo espacial de primer orden – Fuente: (Baronio; Vianco; Rabanal; 2012)

Donde los casilleros grises son vecinos del casillero negro, en tanto que los casilleros blancos no son considerados contiguos respecto a los negros. Los criterios presentados ilustran situaciones de contigüidad de primer orden. No obstante, en algunos problemas puede resultar particularmente útil el abordaje a partir de contigüidades de segundo orden.

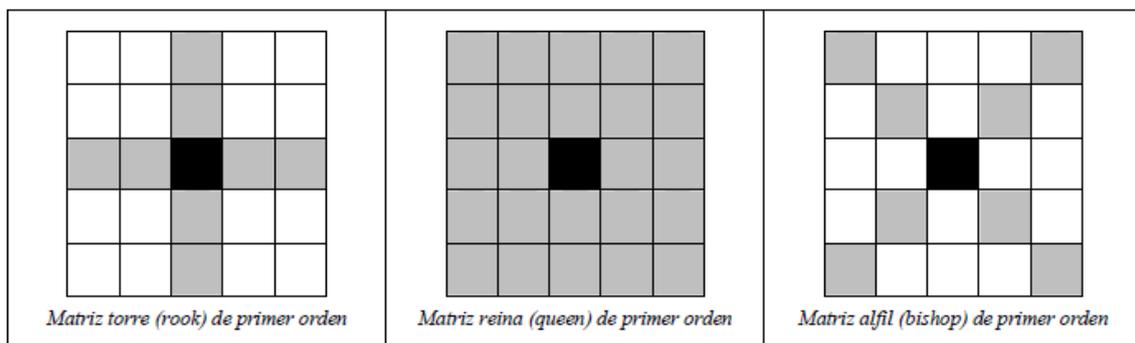


Figura 11: Criterios de contigüidad en la matriz de pesos espaciales. Retardo espacial de segundo orden – Fuente: Baronio; Vianco; Rabanal; 2012



En general, no existen situaciones que requieran la adopción de un criterio de vecindad que vaya más allá del segundo orden. En todo caso, sería necesario explorar otras hipótesis de vecindad que respondan a otras formulaciones sobre la base de especificaciones económicas o geográficas alternativas, como pueden ser: la distancia entre dos unidades, el nivel de intercambio comercial entre diferentes regiones o países, etc. En todos los casos, lo que resultará importante aquí, será escoger variables que sean operativas.

3.5.2 Otras especificaciones para la matriz de contactos

Existen numerosos criterios para definir cuándo dos regiones pueden considerarse vecinas. La reseña de procedimientos presentada a continuación no pretende ser exhaustiva, pero los mismos constituyen las alternativas más difundidas (Chasco, 2003).

- 1) Cliff y Ord (1981) construyeron una matriz sobre la base de la distancia que separa a dos regiones i y j , (d_{ij}), ponderada por el tamaño de la frontera que presenta en común (β_{ij}). En este caso la formulación da lugar a una matriz asimétrica. De esta forma los elementos de \mathbf{W} se reducen a:

$$w_{ij} = (d_{ij})^{-a} (\beta_{ij})^b \quad (39)$$

Donde a y b son parámetros a estimar. Anselin (1980), comenta que dichos parámetros suelen ser dados a priori y no estimados conjuntamente.

- 2) Bodson y Peeters (1975) han sugerido que la matriz \mathbf{W} considere un criterio de accesibilidad general. Esto es, que tenga en cuenta y combine los distintos canales de comunicación entre regiones. En consecuencia, esta formulación permite redefinir el concepto de vecindad, ya que para serlo habrá de haber buenos canales de comunicación entre las regiones. La formulación analítica de esta propuesta se expresa como:

$$w_{ij} = \sum_{n=1}^N k_n \left\{ \frac{a}{[1 + b * \exp(-c_j d_{ij})]} \right\} \quad (40)$$

Siendo:



k_n = la importancia relativa de la vía de comunicación n

N = el número de vías de comunicación

d_{ij} = la distancia entre las regiones i y j

a , b y c_j = son parámetros a estimar

- 3) Case et al. (1993) han propuesto que la matriz W se sustente en distancias económicas. En este caso, se asumirá que una región es vecina de otra cuando esas distancias, en términos de alguna variable económica operativa, sea pequeña (lo que deberá ser definido por el investigador). A menudo, desde esta perspectiva, la distancia económica es entendida como la reducción de los costos de transacción asociados a la interrelación económica de regiones lejanas entre sí (no contiguas). Esto es presumiblemente provocado por mejoras que se asocian a condiciones económicas estructurales similares y costos de información más bajos.
- 4) A continuación, se presenta un resumen con algunas especificaciones para la matriz de pesos espaciales sugeridos en la literatura.



Referencia	Modelo	Descripción
Dacey (1968)	$w_{ij} = d_{ij} \cdot \alpha_i \cdot \beta_{i(j)}$. d_{ij} : distancia entre los puntos o regiones (i, j) . α_i : proporción de i sobre el área total de regiones. . $\beta_{i(j)}$: proporción del perímetro de i en contacto con j
Cliff y Ord (1973)	$w_{ij} = d_{ij}^{-a} [\beta_{i(j)}]^b$. a, b: parámetros positivos
Bodson y Peeters (1975)	$w_{ij} = \sum_{n=1}^N K_n \left\{ \frac{a}{ 1 + b \cdot e^{-c_i d_{ij}} } \right\}$. K_n : importancia del medio de comunicación n . N: total de medios de comunicación considerados . a, b, c _j : parámetros a estimar.
Anselin (1980)	$w_{ij} = d_{ij}^{-2}$	
Cliff y Ord (1981)	$w_{ij} = (c + d_{ij})^{-a}$. c: término constante positivo
Case <i>et al.</i> (1993)	$w_{ij} = \frac{1}{ x_i - x_j }$. x: variable socioeconómica (ej., PIB per cápita).
Molho (1995)	$w_{ij} = \frac{E_j^{-ad_{ij}}}{\sum_{k \neq i} E_k^{-ad_{ik}}}$; $\forall i \neq j$. E: volumen de empleo
Ma <i>et al.</i> (1997)	$w_{ij} = e^{-d_{ij}^2}$ $w_{ij} = (l_{ij}/l_i)^a$ $w_{ij} = \frac{(l_{ij}/l_i)^a}{d_{ij}^{-b}}$. l_{ij} : longitud de frontera entre las regiones (i,j) . l_i : perímetro de la región i
Toral (2000A,B)	$w_{ij} = \delta_{ij} \frac{k_i k_j P_i P_j}{d_{ij}^a}$. $\delta_{ij} = 1$, si las unidades espaciales i, j tienen una frontera en común y cero, si no la tienen. . p: población; k: longitud (km) de carreteras . d_{ij} : distancia por carretera entre las capitales de i, j . a: parámetro positivo, con valores 0, 1 ó 2.
Van der Kruk (2001)	$W = \sum_{d=1}^D W_d$. d: orden de vecindad . D: número máximo de órdenes de vecindad existentes

Tabla 3: Diversas especificaciones de la matriz de pesos espaciales – Fuente: (Chasco, 2003)

3.5.3 Matriz de pesos como instrumento para recoger las interdependencias

Tras la definición anterior, es posible detectar una cierta similitud entre los conceptos de auto-correlación espacial y temporal en la medida en que, en ambos casos, se produce un incumplimiento de la hipótesis de independencia entre las observaciones muestrales, se hallen éstas referidas a unidades de corte transversal o a series temporales. Sin embargo, una importante diferencia aparece entre ellas: la dependencia temporal es únicamente unidireccional (el pasado explica el presente), mientras que la dependencia espacial es multidireccional (una región puede estar afectada no sólo por otra región contigua a ella sino por otras muchas que la rodean, al igual que ella puede influir sobre



aquéllas). En este caso, la solución al problema de la multidireccionalidad en el contexto espacial pasa por la definición de la denominada matriz W de pesos espaciales, de retardos o de contactos (Vayá and Serrano; 2000).

Respecto a cómo construir la matriz W , cabe destacar que no existe una definición de W unánimemente aceptada, si bien se ha de cumplir que sus pesos sean no negativos y finitos (Anselin, 1980). A pesar de ello, de forma habitual se recurre al concepto de contigüidad física de primer orden, utilizado inicialmente por Moran (1948) y Geary (1954), donde w_{ij} es igual a 1 si las regiones i y j son físicamente adyacentes o a 0 en caso contrario (se asume por definición que $w_{ij} = 0$). Si bien dicha matriz de contigüidad es habitualmente utilizada por su simplicidad, ciertamente presenta algunas limitaciones importantes que la hacen ser excesivamente restrictiva como, por ejemplo, su carácter simétrico (no es posible incorporar influencias no recíprocas) o la consideración de la adyacencia física como único determinante de las interdependencias regionales (descuidando con ello, por ejemplo, posibles influencias mutuas entre regiones que, aun estando alejadas, mantienen estrechas relaciones comerciales).

Siguiendo el criterio de la proximidad física, diversos autores han propuesto definiciones de la matriz W basadas en la utilización de la distancia entre regiones (Cliff y Ord, 1973, 1981; Bodson y Peeters, 1975), de manera que la intensidad de la interdependencia entre dos regiones disminuye conforme aumenta la distancia que separa sus respectivos centros. En este sentido, se recurre de forma habitual a la matriz inversa de distancias al cuadrado sugerida por Anselin (1980). Por último, cabe destacar que en los últimos años han sido propuestas matrices de pesos alternativas que tratan de huir del concepto de contigüidad o proximidad física para acercarse más al objeto de estudio analizado en cada caso. Así, cabe mencionar los trabajos de Case et al (1993) donde se define una matriz W basada en un concepto de distancia económica entre regiones o de Vayá et al (1998a) y López-Bazo et al (1999), donde los pesos de W recogen el grado de intercambio comercial entre las regiones analizadas. En cualquier caso es preciso remarcar que, en el proceso de contrastación y estimación, la matriz de pesos ha de ser considerada exógena, de manera que se deberá asegurar la exogeneidad de W cuando se recurra a indicadores socioeconómicos para definir sus elementos (salvo que se considere explícitamente la endogeneidad de W en la especificación del modelo).



Por último, cabe destacar que de forma habitual se recurre a la estandarización de la matriz W , dividiendo cada elemento w_{ij} por la suma total de la fila a la que pertenece, haciendo que la suma de cada fila sea igual a la unidad. Si bien no existe una razón contundente que justifique este hecho, la posibilidad de ponderar por igual la influencia total que recibe cada región de sus vecinos, con independencia del número total de vecinos de cada una de ellas, explicaría dicha transformación. Asimismo, dicha estandarización facilita la interpretación de los coeficientes autorregresivos del modelo estimado al asimilarlos a un coeficiente de correlación, asegurando además que los parámetros espaciales estimados sean comparables entre los distintos modelos propuestos. Evidentemente, tras la estandarización de W la matriz resultante será asimétrica (siempre que $\sum_j w_{ij} \neq \sum_i w_{ij}$) complicando los cálculos de algunos estadísticos y estimadores.



CAPÍTULO 4

Caso de estudio



4 CASO DE ESTUDIO

En este capítulo se aplicará la metodología desarrollada en el capítulo anterior para determinar las variables que más se ajustan a los modelos de generación de viajes en la ciudad de Santander.

Con ello, se plantea un primer acercamiento cuantitativo al proceso de crecimiento y desequilibrios de los viajes producidos y atraídos en Santander, y de contrastación de la teoría a través del uso de cuadros con información estadística y ratios determinados, que se complementarán posteriormente con un análisis más cualitativo de los resultados de los modelos.

4.1 Descripción del área de estudio (ciudad de Santander)

Santander es una ciudad costera española, capital de Cantabria y situada en la costa norte de la península ibérica. El municipio de Santander limita con los de Santa Cruz de Bezana y Camargo. Ambos municipios, con el paso del tiempo, han quedado adheridos al centro de la ciudad. Al sur cierra con la bahía homónima y con el mar Cantábrico por el norte. Es el municipio más poblado de Cantabria siendo además la capital de esta comunidad autónoma y una de las ciudades más importantes del norte de España; Santander es, además, la capital de comunidad autónoma y también de provincia, más



septentrional de España. Este municipio presenta una extensión relativamente pequeña de 36 km^2 y supone menos del 1% de la superficie de la comunidad (ver Figura 12).



Figura 12: Situación del término municipal de Santander – Fuente: PGOU de Santander

El municipio de Santander cuenta, según el nomenclátor de 2009, con la ciudad de Santander como casco urbano y cuatro pueblos colindantes como: Cueto, Monte, Peñacastillo y San Román, localidades que antaño eran pueblos pero que actualmente son asimiladas por el centro urbano debido a la expansión de la población hacia la zona periurbana.

En el centro urbano de Santander vive la mayor parte de la población, aunque es una de las zonas en las que más decae el crecimiento natural debido al envejecimiento de la población y al elevado coste de la vivienda, que provoca el desplazamiento de la población joven hacia zonas de la periferia. En Santander también se encuentran los barrios como: Cabildo de Arriba, Centro, Cuatro Caminos, Puertochico. Castilla-Hermida, aunque no es el centro de la ciudad, está bastante cerca de él y bien comunicado. El Antaño barrio de pescadores (barrio pesquero) y clase trabajadora, hoy está integrado en la ciudad.



El Alisal es un barrio bastante nuevo y uno de los que tiene los mejores servicios de la ciudad. En él se encuentra el Parque Científico Tecnológico, donde se encuentra el campus de la Universidad Europea del Atlántico (UNEATLANTICO). Abunda la población joven donde se encuentran los barrios como: La Albericia, Alisal.

Entorno Cueto-Las Llamas: este barrio está situado alrededor de El Parque Atlántico de las Llamas, un gran parque de 11 hectáreas que tendrá unas 35 en un futuro. Se sitúa desde la avenida Constitución, incluyendo la zona de expansión de Valdenoja, hasta la Avenida Los Castros donde se encuentran los barrios como: La Pereda, Valdenoja, Fumoril.

Monte: este barrio limita con Cueto y tras el 'boom inmobiliario' ha quedado asimilado por la urbe en el área que limita con La Vaguada de las Llamas.

Peñacastillo: se encuentra rodeando la peña de Peñacastillo. Se caracteriza por contener los principales centros comerciales de la ciudad (C.C. Peñacastillo, El Corte Inglés...) En el centro del barrio se pueden observar antiguas casas rurales, así como un amplio contraste con las nuevas grandes edificaciones de sus alrededores, donde se encuentran los barrios como: Barrios: Nuevo Parque, Primero de Mayo, Nueva Montaña, Ojaiz, Adarzo, Camarreal, Lluja, El Empalme, La Lenteja, Campogiro-La Remonta.

San Román de la Llanilla: Se trata de una zona de chalets y pareados que se encuentra en el límite entre lo urbano y lo rural, donde se encuentran los barrios como: Latorre, Somonte, Corban, El Somo, La Llanilla, El Mazo, Canda Landáburu, Lavapiés, La Gloria, La Sierra, La Candía, Corceño, La Cueva, Pintores Montañeses, La Garita, La Canal, El Campizo, La Sota, La Regata y La Cavaduca (barrios existentes hasta 1986 cuando hubo una reordenación de calles eliminando unas y renombrando otras).

Población

Su población actual, según el último dato del Padrón municipal de Santander publicado a fecha de (01 de Enero de 2014) es de 176.064 habitantes. En su área de influencia urbana reside prácticamente la mitad de la población total de Cantabria. A nivel general la cantidad de la población en el municipio ha permanecido prácticamente constante aunque se ha observado una ligera frenada en los últimos años (ver Figura 13)



debido a los procesos de periurbanización y de dispersión urbana que se han ido presentado en varias ciudades de la península española.

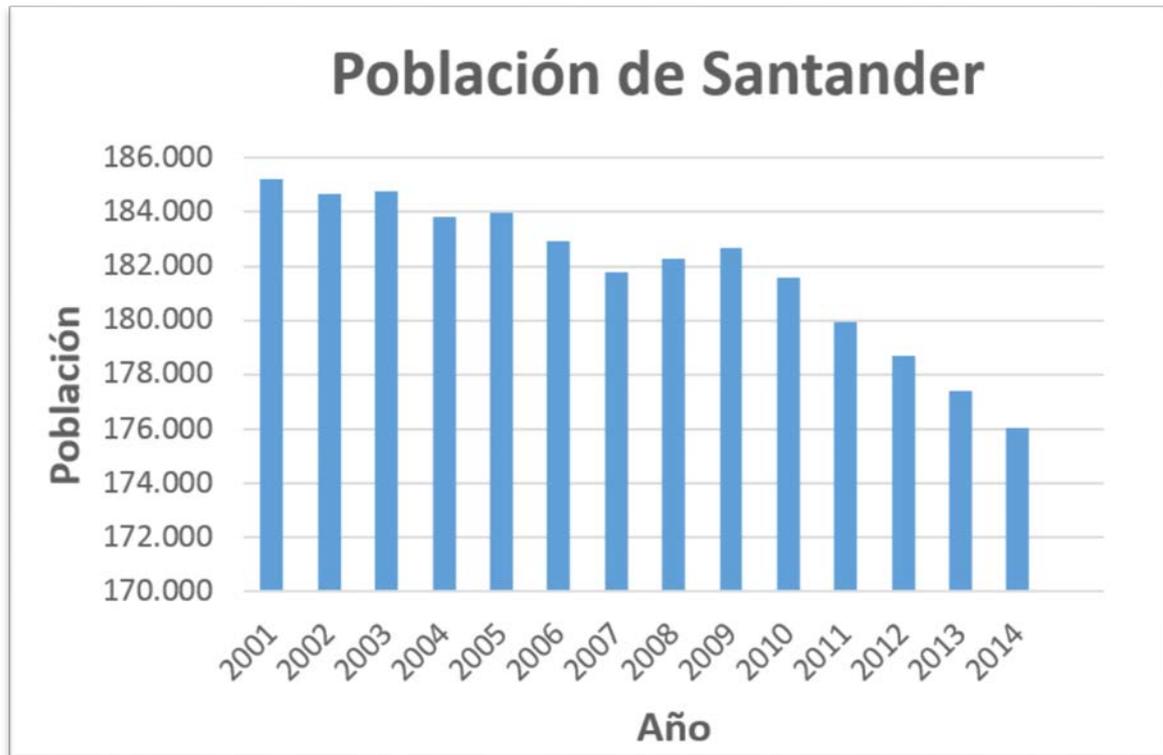


Figura 13: Evolución del municipio de Santander (2001 – 2014) – Fuente: Padrón municipal. Instituto Nacional de Estadística (INE)

Transportes

La red de transportes de Santander consta de varias líneas de autobuses urbanos, gestionados por la empresa municipal TUS (ver Figura 14 y Figura 15). Actualmente el principal medio de transporte público en esta ciudad es el autobús, con una de las medias más altas de utilización por habitante de España.

Como en la mayoría de las ciudades conocidas, el modo principal de transporte es el privado motorizado. Sin embargo cuando se habla de Santander se suele relucir un caso particular por lo cual los desplazamientos que se realizan tanto en autobús como a pie tienen un porcentaje muy elevado dentro del Plan de Movilidad Sostenible.

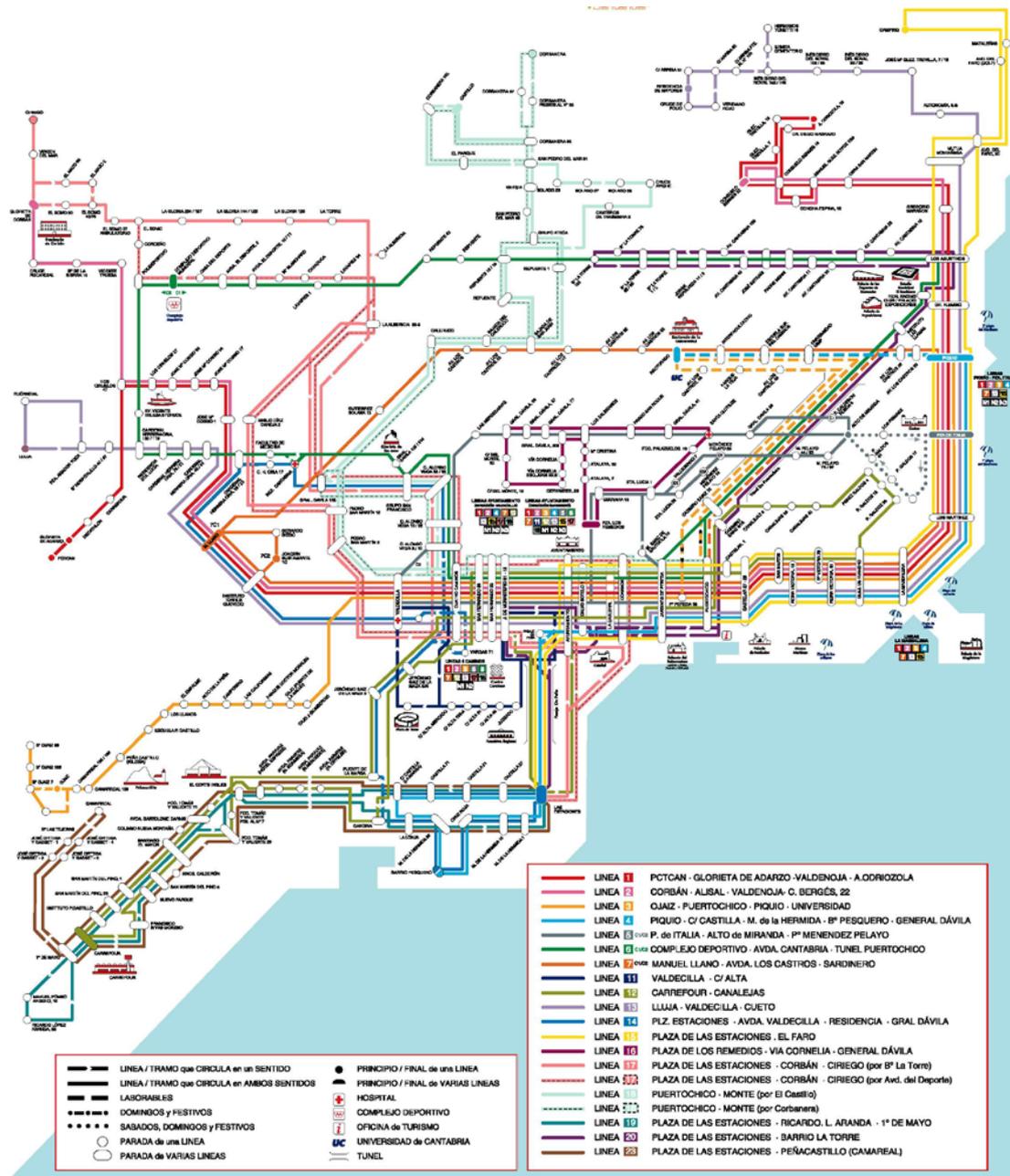


Figura 14: Plano de la red de transporte (Incluyendo líneas de ALSA) – Fuente TUS



Figura 15: Plano de la red de transporte TUS (Excepto líneas de ALSA)

4.2 Análisis y descripción de la base de datos

4.2.1 Análisis de los datos de partida

Para llevar a cabo esta Tesis Doctoral se ha partido de la recopilación de los datos de viajes mediante una encuesta domiciliaria del 2011 donde se han adquirido las informaciones sobre viviendas, los vehículos, las personas del hogar y sus correspondientes viajes realizados.

Teniendo en manos los datos recopilados correspondientes a la encuesta previamente realizada, se procede al estudio y análisis necesario de los mismos con el objetivo caracterizar tanto a los encuestados como sus viajes según las informaciones recogidas la encuesta y posteriormente se procede a la valoración de las variables que forman parte de la calidad global del transporte, tanto público como privado, comprobando la bondad de los datos recopilados, teniendo en cuenta sus caracterizaciones para proceder finalmente a su validación adecuada. Por lo tanto,



aquellos que no cumplan con la coherencia esperada se eliminan del estudio y por supuesto no forman parte de la base de datos que se modeliza basándose en una decisión rigurosa.

Ya con todos los datos validados se procede con la caracterización, tanto de los encuestados, como de sus viajes, teniendo en cuenta el resto de particularidades obtenidas de la encuesta.

Es necesario conocer: la edad, el sexo, la ocupación habitual, el nivel de ingresos, la disponibilidad de carné de conducir, la disponibilidad de vehículo propio... en porcentajes para caracterizar al usuario. Esta caracterización se realiza con el objetivo de conocer quién forma parte de nuestra muestra, de esta forma se definen las necesidades de los individuos de la población (tras la expansión de la muestra) y cuáles son sus características. Al conocer hacia quién van dirigidas las políticas de futuras actuación no es tan complicado conseguir que las mismas sean más eficientes y sus resultados, por lo tanto, mejores.

Otros datos que se ha tenido en cuenta es el comportamiento de los individuos que se mueven en el área de estudio y las razones de sus movimientos, o sea, conocer cómo y por qué se mueven. Una vez conocido sus necesidades, es mucho más sencillo mejorar las alternativas disponibles con el objetivo de facilitar, tanto sus horarios, como sus desplazamientos. No se debe olvidar la característica de la información recogida en la encuesta para cada caso de estudio.

4.2.2 Zonificación del área de estudio

Tanto el área de estudio como la propia zonificación se han determinado en base a la movilidad total, transporte público y transporte privado, a fin de poder analizar escenarios de futuro, en los cuales se contemple el cambio modal de los usuarios.

Se ha valorado la extensión del área de estudio en base a los crecimientos urbanísticos sufridos por el propio Santander y los barrios periféricos. Se ha analizado, en este sentido, la existencia de continuidad o no en el entramado urbano.

Asimismo, en la definición de los límites del área se han tenido en cuenta los nuevos crecimientos previstos en el medio plazo.



Otros factores determinantes a la hora de establecer el área de estudio han sido las características socioeconómicas de las posibles zonas implicadas, así como sus características de movilidad.

En referencia a la zonificación, esta implica la división del área de estudio en áreas agregadas que faciliten una mayor interpretación de datos y agilicen los algoritmos de resolución asociados a los distintos modelos empleados.

Se ha adoptado una zonificación del área objeto de la investigación de acuerdo con la dirección del estudio de manera consistente con los modelos de oferta y de demanda calibrados y teniendo en cuenta (Ortúzar & Willumsen, 2001):

- Desequilibrios de población entre zonas en virtud de los desarrollos producidos en los últimos años.
- Adecuación de la zonificación en virtud de la localización de las zonas respecto de ejes y nudos de transporte, tanto privado como de transporte público, existente y programado.
- Características socioeconómicas del último padrón
- Zonificación administrativa.
- Homogeneidad de las zonas.
- Calificación y uso del suelo aprobado.

Dentro del proceso de zonificación, se ha hecho especial hincapié en localizar las zonas de demanda especial, como son la universidad, centros comerciales, hospitales, áreas industriales, instalaciones deportivas, etc. El motivo de dicha desagregación radica en la diferente motivación de los viajes que presentan estas especiales zonas frente al resto de zonas del área de estudio.

En concreto para el presente estudio se ha tomado como referencia la división del municipio en secciones censales, desagregando aquellas otras zonas de especiales características (ver Tabla 4).



Candina/Parayas
Estaciones
Centro Comercial El Alisal / Mercadona
Centro Comercial Peñacastillo
Centro Comercial. Bahía de Santander
Universidad
Complejo Deportivo La Albericia
Valdecilla
Nueva Montaña
Parque de la Magdalena
Sta. Clotilde
El Sardinero / P. Deportes
Ciudad del Transportista
Residencia Cantabria

Tabla 4: Zonas de especiales características

Como resultado de la explotación de los datos provenientes de la encuesta, se agrupan las 161 zonas en iniciales en 86 zonas finales debido al problema que radica en el riesgo de tener demasiadas celdas vacías en la matriz origen-destino resultante de la encuesta, con la problemática de no contar con número suficiente de viajes para modelar el reparto modal. Estas zonas conforman una disgregación del área de estudio más adecuada y más que suficiente para calibrar con garantías los posteriores modelos.

Por otro lado, este número de zonas está dentro de los rangos recomendados por la literatura científica en materia de modelación de la demanda de transporte para áreas de estudio de la extensión de Santander (ciudad de tipo intermedio).

Finalmente se ha definido las zonas especiales correspondientes a cada zona de estudio (ver Figura 16) tales como:

- Residencia Cantabria que corresponde a la zona de estudio 16
- Santa Clotilde que corresponde a la zona de estudio 24
- Ciudad del Transporte que corresponde a la zona de estudio 28
- Parque de la Magdalena que corresponde a la zona de estudio 64
- El sardinero / Palacio de Deportes que corresponden a la zona de estudio 65



- Nueva Montaña que corresponde a la zona de estudio 70
- Centro Comercial El Alisal / Mercadona que corresponden a la zona de estudio 77
- Centro Comercial Peñacastillo que corresponde a la zona de estudio 78
- Centro Comercial Bahía de Santander que corresponde a la zona de estudio 79
- Universidad que corresponde a la zona de estudio 80
- Complejo Deportivo La Albericia que corresponde a la zona de estudio 81
- Valdecilla que corresponde a la zona de estudio 82
- Candina / Parayas que corresponden a la zona de estudio 85
- Estaciones que corresponde a la zona de estudio 86

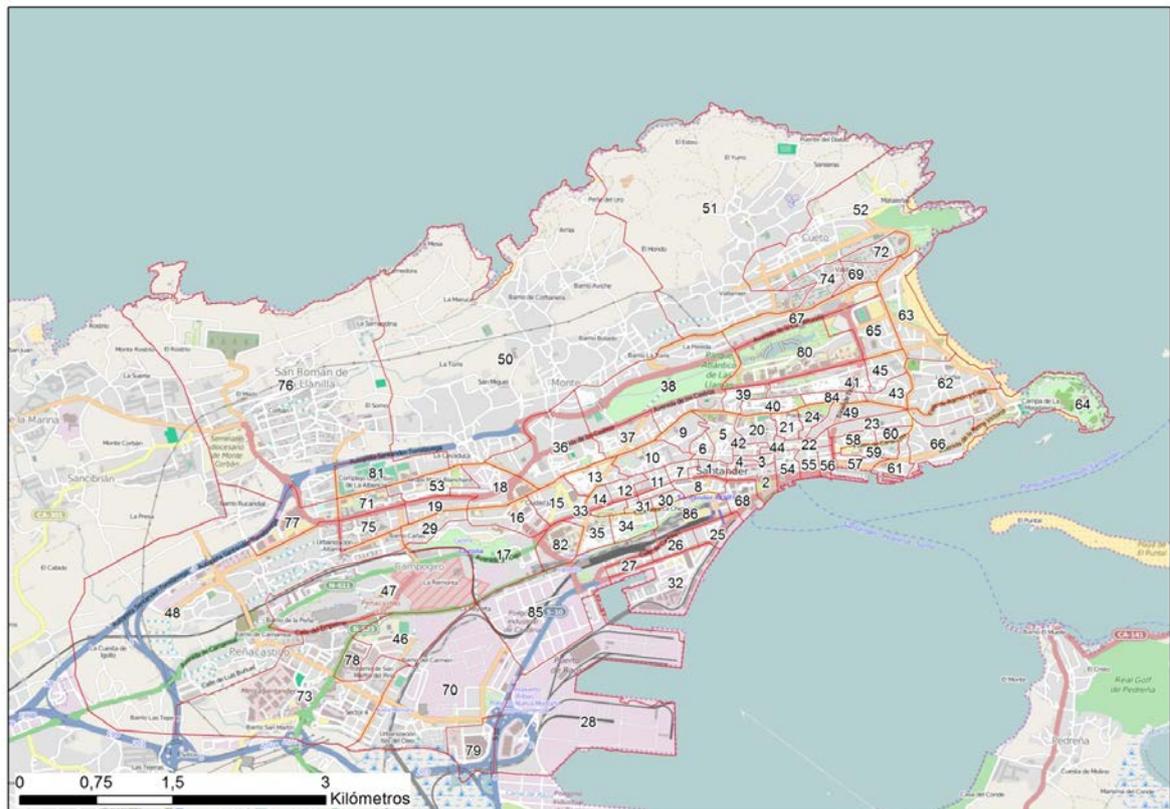


Figura 16: Plano del área de estudio con la numeración de cada una de las zonas



4.2.3 Descripción de las variables

Para la parte de la elaboración de los modelos zonales se ha contado con una base de datos con las variables socioeconómicas y viajes de los hogares encuestados y variables a nivel zonal.

Para los estudios de estimaciones por zonas, las primeras ocho (8) variables son las dependientes que corresponden a cada propósito de viaje y todas las demás son las independientes con las que se combinan hasta que se obtenga las variables independientes más significativas para cada propósito de viaje (ver Figura 17).

Campo	Descripción
PROD_TTRIPS	Viaje Producidos Totales (viajes base zonales)
PROD_HBW	Viajes Producidos basados en el hogar motivo trabajo (viajes base zonales)
PROD_HBO	Viajes Producidos basados en el hogar motivo otros (viajes base zonales)
PROD_NHB	Viajes Producidos no basado en el hogar (viajes base zonales)
ATTRAC_TTRIPS	Viajes Atraídos Totales (viajes base zonales)
ATTRAC_HBW	Viajes Atraídos Basados en el Hogar motivo trabajo (viajes base zonales)
ATTAC_HBO	Viajes Atraídos Basados en el Hogar motivo otros (viajes base zonales)
ATTRAC_NHB	Viajes Atraídos no basados en el hogar (viajes base zonales)
ACC_ACT	Accesibilidad Activa de la zona
ACC_PAS	Accesibilidad Pasiva de la zona
CBD	Tiempo de viaje al centro desde la zona
Hogares	Número de hogares en la zona
EMP	Número de empleos en la zona
POB_2011	Población en la zona
Den_POP/Hectarea	Densidad de población en la zona
DEN/EMP	Densidad de empleos en la zona
FACILITY	Equipamiento especial en la zona (Universidad, Hospital, Ayuntamiento, Gobierno Regional)
Locales_Comer	Número de locales comerciales en la zona
Vehículos	Vehículos de la zona
Pob_2001	Población de la zona en 2001
Pob_2008	Población en la zona en 2008
Ocupados	Población ocupada
Comercio_por meno	Comercios al por menor
Servicios	Empleos del sector servicios
Básica	Empleos básicos
Hectáreas	Superficien en hectáreas
sq/habitables	M2 habitables

Figura 17: Variables y categorías de viaje por zona

Sin embargo para los estudios de estimación por hogares, se ha utilizado como variables dependientes las correspondientes a cada propósito de viaje: TTRIPS, HBO, HBW, NHB como viajes producidos basados en hogar ya que a los hogares solo producen viajes y combinándolas con las variables independientes y siguiendo un sentido lógico hasta encontrarse con las más significativas correspondiente a cada modelo para los casos de estudios requeridos (ver Figura 18).



Campo	Descripción
Id_Hogar	Identificador del hogar
ZONA_TRANS	Zona de transporte donde se encuentra el hogar
X	Coord_X Localización Hogar
Y	Coord_Y Localización Hogar
TTRIPS	Viajes Totales realizados por el hogar
HBO	Viajes Home-Based Motivo Otros realizados por el hogar
HBW	Viajes Home-Based Motivo Trabajo realizados por el hogar
NHB	Viajes Non-Home-Based realizados por el hogar
INCOME	Ingresos del hogar 1)<600€, 2) 600-1500€, 3)1500-2500€, 4)>2500€
AGE	Edad Media del hogar
SIZE	Número de miembros del hogar
WORKERS	Número de trabajadores del hogar
VEH	Número de vehículos del hogar
POB_2011	Población (2011) en la zona del hogar
TIME_CBD	Tiempo de viaje (minutos) al centro de la ciudad desde la zona del hogar
ACC_ACT	Accesibilidad Activa de la zona del hogar
ACC_PAS	Accesibilidad Pasiva de la zona del hogar
HOUSEHOLDS_2001	Número de hogares (2001) de la zona del hogar
EMP	Número de empleos de la zona del hogar
DEN_POB_HA	Densidad de Población (Pob/Hectáreas) de la zona del hogar
DEN_EMP_HA	Densidad de Empleos (Emp/Hectáreas) de la zona del hogar
LOCALS	Número de locales comerciales de la zona del hogar

Figura 18: Variables y categorías de viaje por hogar





CAPÍTULO 5

Análisis y discusión de los resultados



5 ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

En este capítulo se encuentran los análisis de la precisión de las estimaciones de viajes base zonales obtenidas mediante los modelos de producción / atracción por zonas y de producción por hogares para cada una de las categorías de viajes considerando datos de los modelos utilizados.

Para ello se ha tenido en cuenta el porcentaje de error representado en los planos considerando la expresión siguiente:

$$\%error = \frac{V_{bzi} - V_{ei}}{V_{bzi}} \times 100 \quad (41)$$

Donde V_{bzi} representa los viajes de base zonales y V_{ei} los viajes estimados, para la categoría de viajes i .

Por lo, el criterio de signos elegido en las leyenda es el positivo para la sub estimación de viajes y el negativo para la sobre estimación.

También se encuentran los resultados de los modelos de viajes tanto por estimación zonales como por hogares correspondientes a las matrices de contigüidad utilizadas para el desarrollo de esta Tesis, siempre partiendo de los modelos básicos.



Para examinar la dependencia espacial después de controlar otras características relevantes podemos implementar modelos multivariados de regresión espacial. Peor aún, si las medidas exploratorias revelan la presencia de dependencia espacial debemos implementar dichos modelos, incluso si no tenemos ningún interés sustantivo en ella, esto porque ignorar la auto-correlación entre nuestras unidades geográficas podría comprometer la fiabilidad de nuestro análisis (Anselin 1992). La regresión espacial proporciona herramientas para diagnosticar y tratar los casos de correlación espacial.

Los modelos de regresión espacial se pueden dividir en dos grupos, de acuerdo a la forma de dependencia espacial que haya entre las observaciones (Anselin, 1995).

- En el **modelo espacial lag**, la variable dependiente está correlacionada espacialmente de la siguiente forma en la ecuación (35).
- En el **modelo espacial error**, la variable dependiente no está correlacionada por sí misma, sino que es el error de muestreo aleatorio el que está correlacionado en la ecuación (36).

Ambos tipos de auto-correlación espacial se pueden corregir estimando los coeficientes de correlación espacial, lo que requiere el uso de una matriz de ponderadores espaciales W , que aplica una fórmula de ponderación de datos con una ponderación directamente relacionada a la contigüidad entre observaciones.

El programa ofrece un camino para determinar que modelo necesitamos implementar. Este software permite estimar una regresión estándar de mínimos cuadrados y una serie de diagnósticos espaciales con los cuáles podemos decidir qué tipo de auto-correlación espacial está presente en nuestros datos y, por tanto, el modelo a utilizar.



5.1 Los resultados de weights (pesos) creados

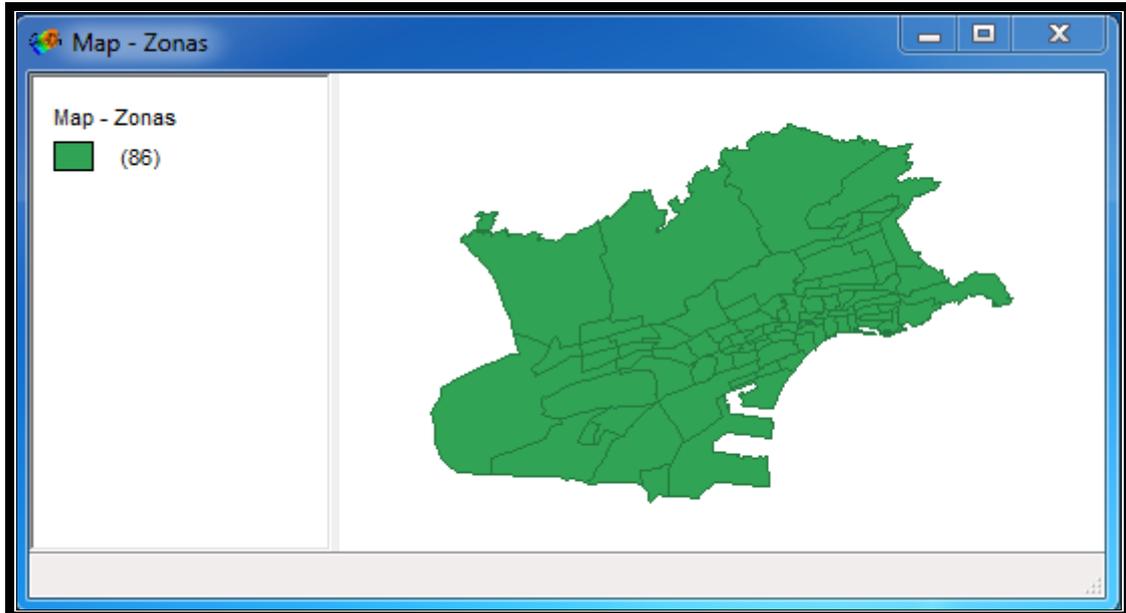


Figura 19: Construcción del Polígono de Thiessen con la ayuda de GeoDa

Estos resultados se obtienen, haciendo el uso del **Polígono Thiessen** (ver Figura 19) ya creado anteriormente. Se han obtenido los resultados de los modelos de las matrices siguientes:

- 1) Los resultados de la matriz de contigüidad Queen
- 2) Los resultados de la matriz de contrigüidad Rook
- 3) Los resultados de la matriz Threshold distance

La matriz de pesos espaciales están basados en Contigüidad Queen que incluye todos los puntos de sus fronteras comunes y sus vértices por lo tanto siempre tienen una estructura de conexión más densa (más vecinos) que en Contigüidad Rook, que sólo utiliza las fronteras comunes para definir sus vecinos.

Se define matriz Threshold Distance, con el fin de construir una matriz de pesos espacial basado en una banda de distancia, para ello se especifica una distancia crítica. El programa internamente calcula la distancia mínima necesaria para asegurar que cada observación tenga al menos un vecino, utilizando el criterio de banda distancia para que todos los puntos seleccionados estén dentro de la distancia especificada desde la observación considerada.

5.2 Modelos de viajes estimados por zonas

5.2.1 Viajes producidos totales “PROD_TTRIPS” (estimación por zonas)

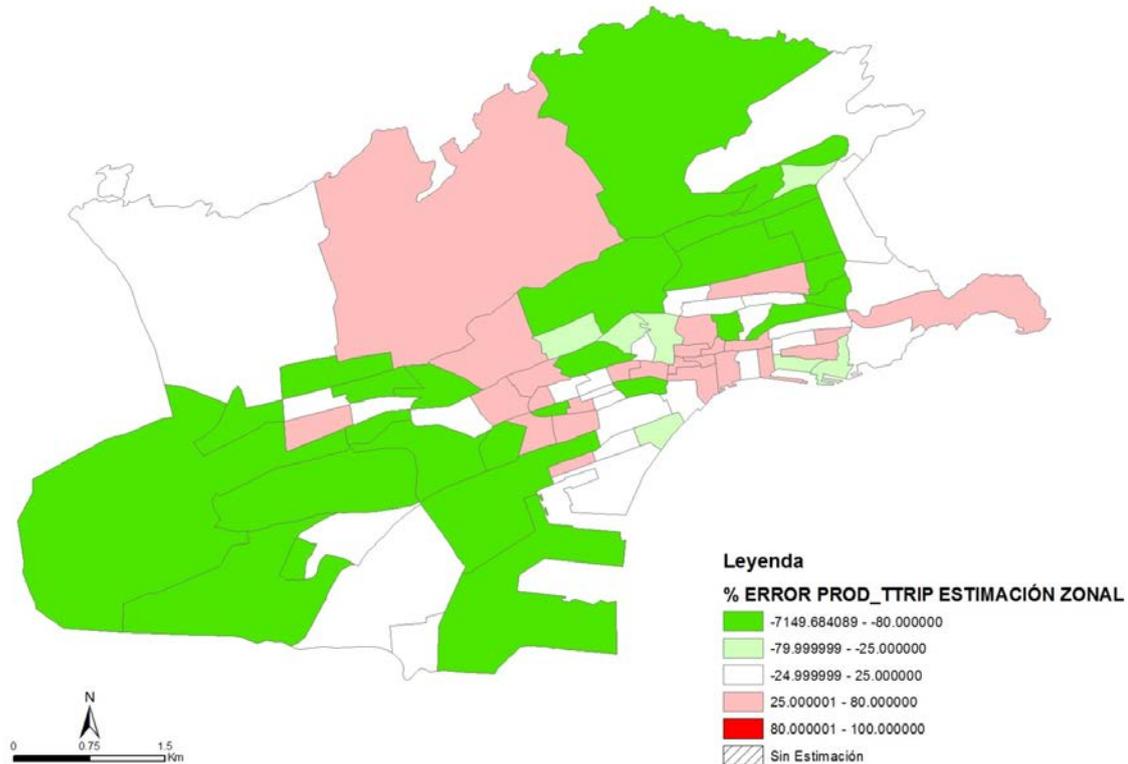


Figura 20: Plano estimación viajes producidos totales (estimación por zonas)

- **Análisis de viajes producidos totales (estimación por zonas)**

El modelo calibrado usa como variables los hogares junto con la población de la zona en el año 2011. Es un modelo bastante bueno porque a pesar de que el intercepto tiene un tamaño considerable respecto de los viajes estimados, su test “t” es muy bajo. Los coeficientes de las variables son considerables y su test “t” bastante por encima de 1,96. El R^2 es adecuado aunque no muy alto.

A la vista de los resultados representados en los planos, podemos ver como las zonas con un buen ajuste se encuentran en Cabo Mayor, el Sardinero, la Playa de los Peligros, el Barrio Pesquero, en la entrada a Santander dirección Bilbao entre la autovía S-10 y la entrada del ferrocarril, las estaciones de autobús y ferrocarril, dispersas sin patrón aparente por el centro urbano y algunas al oeste del mismo. Desde el punto de vista



de la sub estimación se observa como sólo existen zonas en el intervalo 25%-80% y están concentradas por el noroeste, el Palacio de la Magdalena y sobre todo en el entorno de la zona del Ayuntamiento, incluyéndola también. Si analizamos la sobre estimación de viajes nos damos cuenta que existe un gran error respecto de los viajes base zonales ya que sólo existen una pocas zonas en el intervalo de pequeña sobre estimación, perteneciendo la mayor parte al intervalo con mayor error. Las encontramos en la parte norte, en las zonas en torno a la universidad, incluida, al oeste, sur y suroeste.

- **Resultados de los modelos de viajes producidos totales “PROD_TTRIPS” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “PROD_TTRIPS_1”							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q3	Spacial Error Q1	Spacial Lag R1	Spacial Error R2	Spacial Lag TD1	Spacial Error TD2
CONSTANT	381,841	2012,051	466,617	-212,574	423,384	-109,519	218,230
	0,480	1,320	0,552	-0,229	0,505	-0,039	0,1926
Hogares	3,162	2,971	2,707	2,924	2,796	3,161	3,228
	2,997	2,896	2,645	2,835	2,716	3,048	3,127
POB_2011	1,075	1,099	1,203	1,068	1,189	1,085	1,090
	2,884	3,033	3,276	2,910	3,235	2,956	2,963
Parámetros de Ajustes							
W_PROD_TTRIPS		-0,302		0,158		0,090	
		-1,225		1,137		0,175	
LAMBDA			0,221		0,195		0,599
			1,394		1,243		2,123
R-squared	0,331	0,346	0,351	0,344	0,348	0,332	0,322
Log likelihood	-820,012	-819,325	-819,151	-819,410	-819,318	-820,001	-819,286
S.E of reg,	3347,900	3311,280	3298,920	3316,180	3308,210	3347,24	3372,870
Likelihood Ratio Test		0,241	0,189	0,272	0,239	0,882	0,228
		1,3745	1,723	1,205	1,387	0,022	1,453

Tabla 5: Resumen del modelo “PROD_TTRIPS_1” Viajes producidos totales (Estimación por zonas)

En éste modelo “PROD_TTRIPS_1” (estimación por zonas), las correlaciones espaciales mejoran un poco los parámetros de ajustes en los modelos pero no lo suficientemente buenos como para tenerlas en cuenta, porque en el Test razón de verosimilitud es mucho menor que 3,84 que es el valor correspondiente al 5% en la probabilidad estadística (Chi-Square) y también el índice de probabilidad en cada caso es superior a 0,05 (ver Tabla 5). Como no se cumple ninguna de las condiciones mencionadas anteriormente, nos quedamos con el modelo básico.



Percentage Points of the Chi-Square Distribution									
Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Tabla 6: The chi-square value for different probability level (0,01-0,99) for degrees of freedom ranging from 1 to 60

- A cada variable le corresponde su coeficiente y su t estadística.
- Al **Likelihood Ratio Test** (test de razón de verosimilitud) le corresponde su índice la probabilidad y valor test “t” (ver Tabla 6).
- Para obtener este resultado que es el más significativo, se ha tenido que calibrar varios puntos diferentes para cada modelo y escoger el mejor.



REGRESIÓN - MEJORES RESULTADOS DEL MODELO "PROD_TTRIPS_2"							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q3	Spacial Error Q2	Spacial Lag R1	Spacial Error R2	Spacial Lag TD2	Spacial Error TD2
CONSTANT	1002,141	2967,718	899,650	1004,852	960,691	-1841,985	882,917
	1,440	2,077	1,478	1,140	1,454	-1,051	0,431
Hogares	3,048	2,812	3,160	3,049	3,124	3,117	3,100
	3,260	3,108	3,604	3,318	3,447	3,361	3,377
Den_POP_He	8,050	8,280	8,230	8,052	8,011	8,220	8,045
	4,173	4,441	4,663	4,057	4,362	4,294	4,269
Parámetros de Ajustes							
W_PROD_TTRIPS		-0,361		-0,001		0,549	
		-1,464		-0,005		1,741	
LAMBDA			-0,348		-0,082		0,824
			-1,400		-0,471		6,607
R-squared	0,392	0,412	0,410	0,392	0,394	0,381	0,379
Log likelihood	-815,929	-814,887	-815,150	-815,929	-815,833	-814,961	-815,162
S.E of regression	3192,650	3140,670	3146,530	3192,650	3187,040	3221,060	3226,340
Likelihood Ratio Test		0,149	0,212	0,996	0,661	0,164	0,217
		2,084	1,560	0,0000	0,193	1,936	1,5332

Tabla 7: Resumen del modelo "PROD_TTRIPS_2" Viajes producidos totales (Estimación por zonas)

En éste modelo "PROD_TTRIPS_2" (estimación por zonas), nos encontramos en el mismo caso que en el modelo "PROD_TTRIPS_1" (estimación por zonas) y de nuevo nos quedamos con el **Modelo Básico** que es el que mejor ajuste nos presenta, o sea sin correlaciones espaciales.

- ✓ Por lo tanto analizando estos dos modelos, se ha quedado con el **Modelo Básico_2** correspondiente al propósito de viajes "PROD_TTRIPS_2" (estimación por zonas) como mejor modelo, porque nos presentan las variables independientes más significativas y también tiene mejor ajuste (R-squared).

5.2.2 Viajes producidos basados en el hogar motivo trabajo (estimación por zonas)

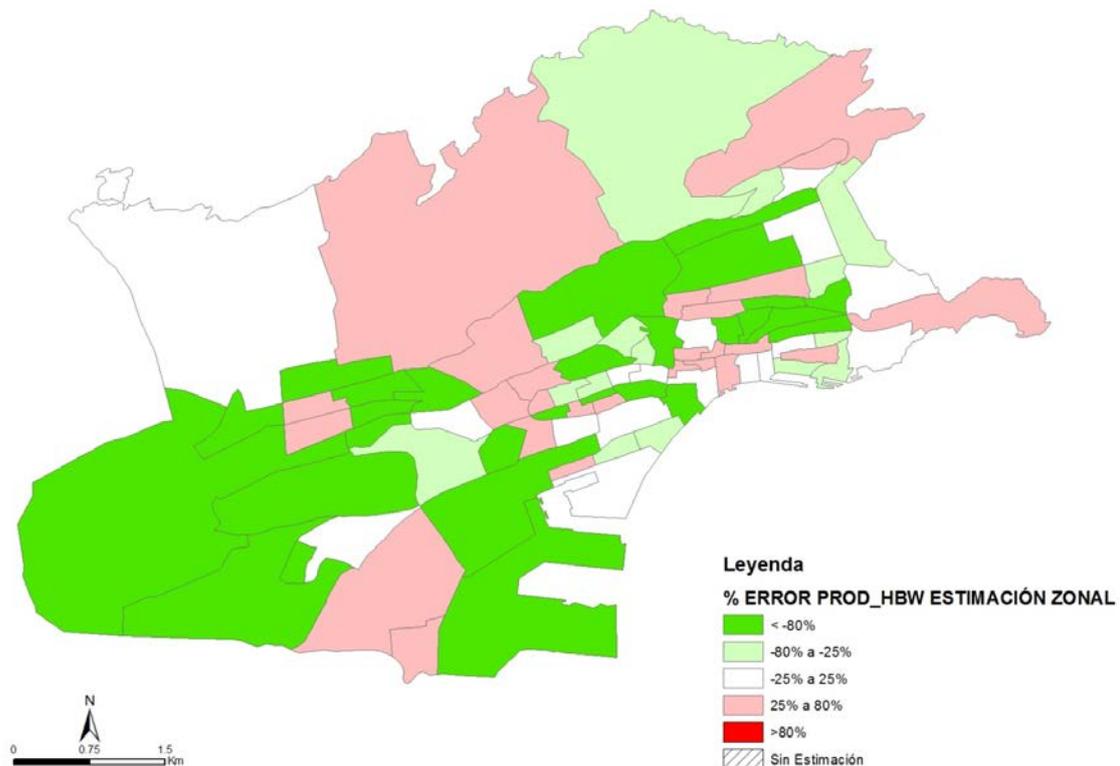


Figura21: Plano viajes producidos basados en el hogar motivo trabajo (estimación por zonas)

- **Análisis viajes producidos basados en el hogar motivo trabajo (estimación por zonas)**

Para la elaboración de este modelo solamente se ha utilizado una de las variables disponibles, esta es la población existente en cada zona en el año 2011. Sobre el modelo resultante hay varias cosas importantes que comentar, primero que el coeficiente del intercepto tiene un valor intermedio si se compara con la estimación de viajes resultante y bastante parecido al modelo utilizado para obtener los viajes producidos totales (estimación por zonas), pero su test “t” es más alto, o sea mucho mayor que 1,96 y el R^2 es aceptable. En cuanto a la variable población 2011 tiene un coeficiente relativamente pequeño, lo cual es entendible porque tiene que ser multiplicado por la población de cada zona que es un número bastante grande, y su test “t” demuestra su importancia porque es superior a cuatro.

Al analizar el porcentaje de error de la estimación en los planos, las zonas con una buena estimación son principalmente: el Barrio Pesquero, la estación de ferrocarril, una



zona al oeste, la Playa del Camello, la Playa de los Peligros, la zona del estadio del Racing y las que están alineadas a lo largo del Paseo Pereda y las calles de Calvo Sotelo, Jesús del Monasterio y Burgos. Desde el punto de vista de la sub estimación no hay zonas en el peor intervalo pero si podemos encontrar zonas pertenecientes al de 25%-80% que están diseminadas por el centro, en la Magdalena, en Cabo Mayor, al noroeste del centro urbano y justo al comienzo de las entradas por la autovía S-10 y del ferrocarril a Santander dirección Bilbao entre las mismas. En el estudio de la sobre estimación existen zonas del intervalo con el error más pequeño al norte, en la Playa del Sardinero y diseminadas por el centro. Sin embargo, destacan las de mayor error de sobre estimación en la parte oeste, sur, zona de la universidad y colindantes y en torno a la entrada sur del túnel de Tetuán.

- **Resultados de los modelos de viajes producidos basados en el hogar motivo trabajo “PROD_HBW” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “PROD_HBW_1”							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q2	Spacial Error Q2	Spacial Lag R1	Spacial Error R2	Spacial Lag TD1	Spacial Error TD1
CONSTANT	342,154	748,518	268,146	363,753	331,462	-1799,256	343,164
	1,335	1,914	1,207	1,179	1,347	-2,788	1,395
POB_2011	0,451	0,476	0,488	0,454	0,456	0,464	0,453
	4,229	4,591	4,959	4,258	4,397	4,265	4,288
Parámetros de Ajustes							
W_PROD_HBW		-0,347		-0,021		1,629	
		-1,512		-0,134		3,679	
LAMBDA			-0,499		-0,071		-0,158
			-2,026		-0,411		-0,237
R-squared	0,175	0,204	0,227	0,176	0,178	0,124	0,181
Log likelihood	-730,628	-729,566	-728,823	-730,620	-730,551	-729,393	-729,466
S.E of regression	1184,090	1163,210	1146,630	1183,940	1182,470	1220,700	1180,300
Likelihood Ratio Test		0,145	0,056	0,901	0,696	0,116	0,127
		2,123	3,600	0,015	0,1526	2,469	2,322

Tabla 8: Resumen del modelo “PROD_HBW_1” Viajes producidos basados en el hogar motivo trabajo (Estimación por zonas)

En éste modelo “PROD_HBW_1” (estimación por zonas), se puede apreciar que existe una sensible correlación espacial en el Queen - Spacial Error Q2 con una significatividad cercana al 95% siendo el que mejor ajuste nos presenta.



REGRESIÓN - MEJORES RESULTADOS DEL MODELO "PROD_HBW_2"							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q2	Spacial Error Q2	Spacial Lag R1	Spacial Error R2	Spacial Lag TD1	Spacial Error TD1
CONSTANT	434,618	831,733	281,230	401,405	422,487	117,666	469,477
	1,473	2,028	1,154	1,153	1,466	0,308	1,472
Sq/habitables	0,014	0,015	0,017	0,014	0,014	0,013	0,013
	3,212	3,659	4,391	3,211	3,314	3,134	3,052
Parámetros de Ajustes							
W_PROD_HBW		-0,369		0,032		0,276	
		-1,545		0,198		1,240	
LAMBDA			-0,603		-0,038		0,277
			-2,482		-0,221		1,165
R-squared	0,109	0,143	0,184	0,110	0,110	0,145	0,139
Log likelihood	-733,947	-732,828	-731,607	-733,930	-733,927	-732,400	-732,692
S.E of reg,	1230,690	1207,310	1178,210	1230,320	1230,230	1205,700	1209,800
Likelihood Ratio Test		0,135	0,030	0,852	0,840	0,078	0,113
		2,238	4,680	0,035	0,041	3,095	2,510

Tabla 9: Resumen del modelo "PROD_HBW_2" Viajes producidos basados en el hogar motivo trabajo (Estimación por zonas)

En éste modelo "PROD_HBW_2" (estimación por zonas), sí que hemos considerado las correlaciones espaciales porque el modelo más significativo es: **S_E_QUEEN – Spacial Error – Q2** en una matriz de contigüidad de segundo orden y presenta tanto los valores de sus parámetros como su variable son más ajustados (ver Tabla 9).

- ✓ Analizando estos dos modelos, nos quedamos con el modelo **S_E_QUEEN – Spacial Error – Q2** en una matriz de contigüidad de segundo orden que corresponde al propósito de viajes "PROD_HBW_2" (estimación por zonas) como mejor modelo, porque nos presenta una variable independiente más significativa, también tiene mejor ajuste (R-squared) y nos ha permitido tener en cuenta las correlaciones espaciales que es uno de los objetivos de éste trabajo.

5.2.3 Viajes producidos basados en el hogar por otros motivos (estimación por zonas)

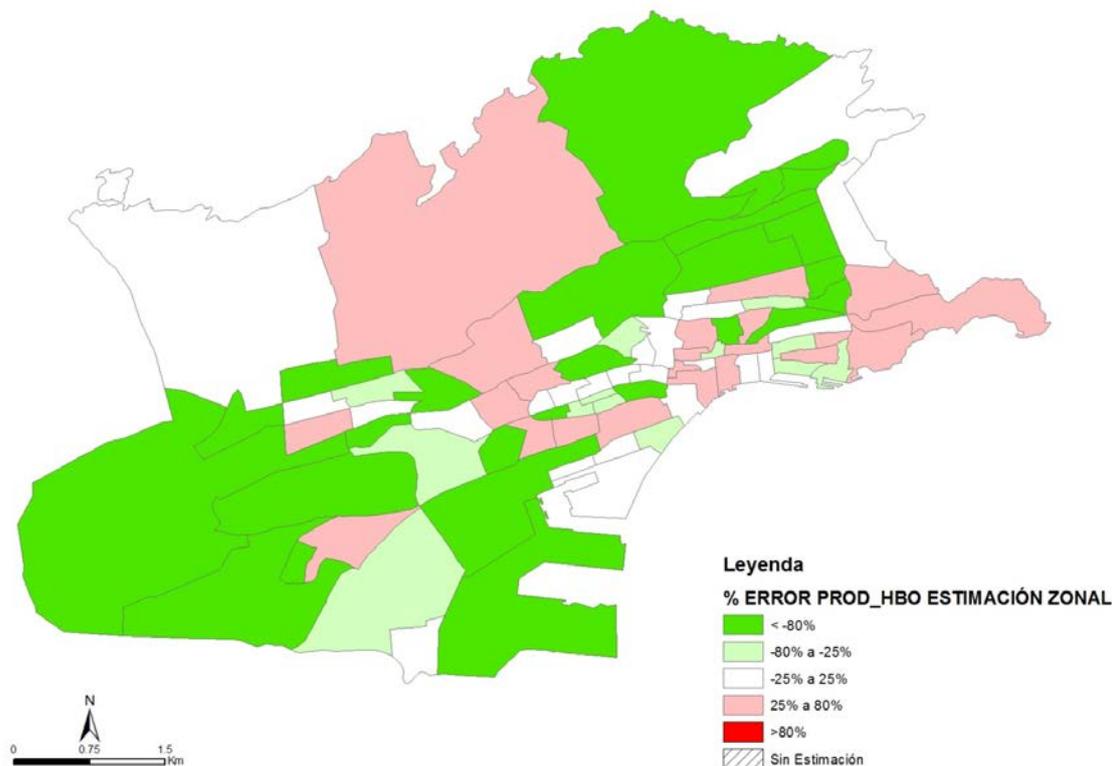


Figura 22: Plano viajes producidos basados en el hogar por otros motivos (estimación por zonas)

Para la elaboración del modelo que estima los viajes producidos basados en el hogar motivo otros con datos zonales he utilizado dos variables que no había usado antes como son la densidad de población de las zonas y los m² habitables existentes en cada una de ellas, ambas con bastante relación con las utilizadas en modelos anteriores. En cuanto al análisis específico del modelo, podemos observar que es muy bueno ya que en primer lugar el intercepto tiene un valor muy pequeño y su test “t” es ridículo. Por otro lado el R² es bastante elevado, los coeficientes de las variables son totalmente razonables y sus test “t” considerablemente por encima del 1,96.

Analizando los resultados obtenidos en los planos, las zonas con una mejor estimación están al oeste, en Cabo Mayor, la Playa del Sardinero, Barrio Pesquero y difuminadas por el centro urbano, destacando una alineación de las mismas a lo largo de las calles de Jesús del Monasterio, Burgos y San Fernando. En cuanto a la subestimación,



solo encontramos zonas que pertenecen al grupo con menos error situadas en el noroeste del centro urbano, dispersas por él sin patrón reconocible, la estación de ferrocarril y las colindantes al oeste, la Magdalena, la Playa del Camello y de los Peligros. Desde el punto de vista de la sobre estimación podemos ver como existen zonas del intervalo de error pequeño dispersas por el área de estudio pero destacan sobre todo las que tienen un mayor error al oeste, suroeste, sur, norte y la zona de las universidades junto con las que están a su alrededor excepto por su parte sur.

- **Resultados de los modelos de viajes producidos basados en el hogar por otros motivos “PROD_ HBO” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “PROD_ HBO_1”							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q3	Spacial Error Q2	Spacial Lag R1	Spacial Error R2	Spacial Lag TD2	Spacial Error TD3
CONSTANT	24,245	1908,619	259,986	88,483	-37,490	-1228,959	14,840
	0,045	2,016	0,541	0,138	-0,076	-0,731	0,002
Sq/habitables	0,031	0,028	0,027	0,031	0,032	0,031	0,031
	3,542	3,427	5,923	3,611	3,824	3,618	3,610
Den_POP_He	7,151	7,341	7,185	7,246	7,194	7,249	7,150
	5,474	5,922	5,923	5,289	6,005	5,592	5,575
Parámetros de Ajustes							
W_PROD_HBO		-0,510		-0,025		0,348	
		-2,266		-0,187		0,775	
LAMBDA			-0,474		-0,159		0,970
			-1,661		-0,903		45,974
R-squared	0,476	0,512	0,499	0,477	0,484	0,471	0,465
Log likelihood	-783,911	-781,659	-782,673	-783,894	-783,529	-782,980	-783,221
S.E of regression	2200,190	2125,060	2152,770	2199,620	2185,210	2212,330	2224,950
Likelihood Ratio Test		0,034	0,117	0,854	0,382	0,172	0,240
		4,502	2,4748	0,034	0,763	1,860	1,379

Tabla 10: Resumen del modelo “PROD_HBO_1” Viajes producidos basados en el hogar por otros motivos (Estimación por zonas)

En éste modelo “PROD_HBO_1” (estimación por zonas) hemos tenido en cuenta las correlaciones espaciales porque los valores de sus parámetros se ajustan más. El mejor modelo corresponde al S_E_QUEEN – Spacial Lag – Q3 en una matriz de contigüidad de tercer orden (ver Tabla 10).



REGRESIÓN - MEJORES RESULTADOS DEL MODELO "PROD_HBO_2"							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q2	Spacial Lag R2	Spacial Error R1	Spacial Lag TD1	Spacial Error TD1
CONSTANT	214,9477	-402,608	292,853	-360,033	356,438	-1004,816	329,977
	0,349	-0,563	0,420	-0,504	0,560	-1,332	0,434
Sq/habitables	0,052	0,049	0,052	0,049	0,050	0,048	0,048
	5,823	5,575	5,793	5,603	5,581	5,508	5,437
Parámetros de Ajustes							
W_PROD_HBO		0,239		0,219		0,423	
		1,713		1,580		2,416	
LAMBDA			0,379		0,164		0,495
			1,845		1,033		2,550
R-squared	0,287	0,319	0,321	0,314	0,300	0,344	0,339
Log likelihood	-797,167	-795,719	-795,695	-795,932	-796,654	-794,178	-794,711
S.E of regression	2566,860	2510,020	2504,940	2517,850	2544,660	2463,56	2472,150
Likelihood Ratio Test		0,089	0,086	0,116	0,311	0,014	0,027
		2,895	2,943	2,469	1,025	5,978	4,910

Tabla 11: Resumen del modelo "PROD_HBO_2" Viajes producidos basados en el hogar por otros motivos (Estimación por zonas)

En éste modelo "PROD_HBO_2" (estimación por zonas) hemos tenido en cuenta las correlaciones espaciales porque los valores de sus parámetros se ajustan más. El mejor modelo corresponde al **S_E_Threshold Distance – Spacial Lag – TD1** en una matriz de contigüidad de primer orden (ver Tabla 11).

- ✓ Analizando estos dos modelos, nos quedamos con el modelo **S_E_QUEEN – Spacial Lag – Q3** en una matriz de contigüidad de tercer orden que corresponde al "PROD_HBO_1" (estimación por zonas) como mejor modelo, porque nos presentan las variables independientes más significativas, también tiene mejor ajuste (R-squared) y nos ha permitido tener en cuenta las correlaciones espaciales por el valor de sus parámetros ya que son lo más adecuados y que es uno de los objetivos de ésta Tesis Doctoral.

5.2.4 Viajes producidos no basados en el hogar (estimación por zonas)

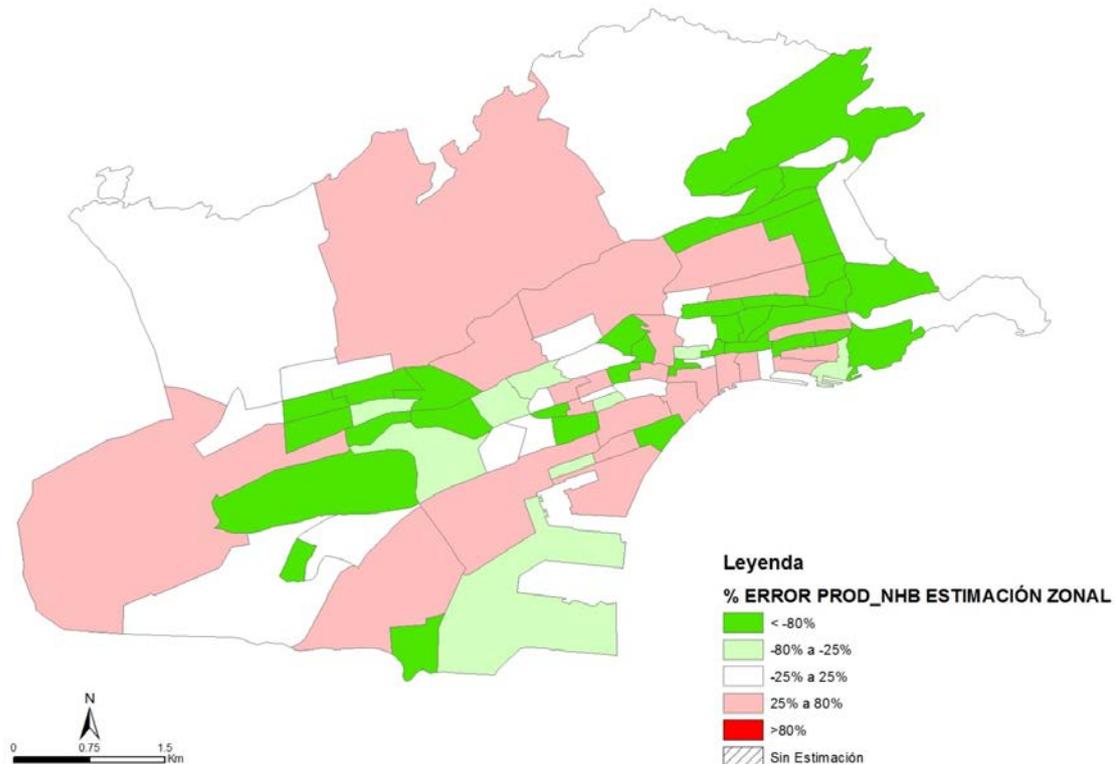


Figura 23: Plano de viajes producidos no basados en el hogar (estimación por zonas)

- **Análisis viajes producidos no basados en el hogar (estimación por zonas)**

Para la elaboración de este modelo solamente he utilizado una variable, que es la accesibilidad activa de la zona. El modelo en sí es un poco peculiar ya que a la hora de analizarlo podemos observar un intercepto con un coeficiente y test “t” razonable, el coeficiente que multiplica a la accesibilidad activa de cada zona es adecuado y con un test “t” superior al 1,96, pero el R^2 ajustado es especialmente bajo. A pesar de lo expuesto, este es el mejor según los criterios utilizados para la calibración de los modelos de viajes producidos no basados en el hogar mediante datos zonales.

A la hora de analizar los datos representados en los mapas podemos ver como las zonas con un mejor ajuste están concentradas al oeste, al norte, la Magdalena, la Playa del Sardinero, MercaSantander y dispersas por el centro de la ciudad. Hay una subestimación de viajes, perteneciente al intervalo de menor error, manifestada a lo largo de



la entrada a Santander por la S-10 hasta justo antes del Palacio de Festivales, al oeste, al noroeste y también en la parte final de la autovía S-20, más concretamente en la avenida de la Constitución. Desde el punto de vista de la sobre estimación son claramente predominantes las zonas que tiene un mayor error concentrándose en torno a la zona del Grupo María Blanchard, incluyéndola, Cabo Mayor y al sur, este y norte de la universidad.

- **Resultados de los modelos de viajes producidos no basados en el hogar “PROD_NHB” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “PROD_NHB_1”							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q3	Spacial Lag R1	Spacial Error R3	Spacial Lag TD1	Spacial Error TD1
CONSTANT	113,600	66,257	109,959	69,691	160,419	36,107	172,883
	1,201	0,702	1,256	0,736	1,420	0,334	1,565
ACC_ACT	10,734	7,439	11,124	7,751	8,540	7,809	8,244
	2,891	2,021	3,166	2,093	1,966	2,105	1,993
Parámetros de Ajustes							
W_PROD_NHB		0,329		0,301		0,390	
		2,273		2,084		1,883	
LAMBDA			-0,293		0,266		0,387
			-1,005		1,774		1,777
R-squared	0,090	0,154	0,105	0,146	0,130	0,126	0,118
Log likelihood	-623,074	-620,881	-622,603	-621,164	-621,808	-621,823	-622,187
S.E of regression	339,037	326,912	336,233	328,429	331,619	332,379	333,821
Likelihood Ratio Test		0,036	0,332	0,051	0,111	0,114	0,183
		4,386	0,942	3,821	2,532	2,503	1,773

Tabla 12: Resumen del modelo “PROD_NHB_1” Viajes producidos no basados en el hogar (Estimación por zonas)

En éste modelo “PROD_NHB_1” (estimación por zonas) hemos tenido en cuenta las correlaciones espaciales porque los valores de sus parámetros se ajustan más, a pesar de que su test “t” (test estadístico) ha bajado un poco pero sigue siendo significativo o sea, mayor que 1,96 y su R² (coeficiente de ajuste) ha sido más ajustado que él del modelo básico. También hemos tenido en cuenta su **test razón de verosimilitud** que es mucho **mayor que 3,84** y por supuesto corresponde que su índice o su **valor de probabilidad** que es **menor que 0,05**. Analizando todos los términos, mejor modelo de todos los modelos corresponde en éste caso al **S_E_QUEEN – Spacial Lag – Q1** en una matriz de contigüidad de primer orden (ver Tabla 12).



REGRESIÓN - MEJORES RESULTADOS DEL MODELO "PROD_NHB_2"							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q3	Spacial Error Q1	Spacial Lag R2	Spacial Error R2	Spacial Lag TD1	Spacial Error TD3
CONSTANT	107,749	248,337	88.197	132,204	89,868	43,894	105,772
	2,653	2,606	3,147	2,387	3,115	0,604	2,196
Locales_Comer.	1,991	2,028	2,140	2,091	2,125	1,912	1,999
	8,765	9,174	12,661	8,418	12,258	8,259	8,912
Parámetros de Ajustes							
W_PROD_NHB		-0,366		-0,098		0,198	
		-1,617		-0,717		1,095	
LAMBDA			-0,521		-0,478		0,280
			-0,521		-2,668		0,558
R-squared	0,477	0,495	0,542	0,481	0,536	0,485	0,474
Log likelihood	-599,225	-598,157	-595,521	-598,992	-595,932	-598,708	-598,556
S.E of regression	256,927	252,641	240,520	255,997	242,167	255,069	257,783
Likelihood Ratio Test		0,144	0,006	0,495	0,010	0,309	0,247
		2,135	7,407	0,465	6,586	1,033	1,338

Tabla 13: Resumen del modelo "PROD_NHB_2" Viajes producidos no basados en el hogar (Estimación por zonas)

En el caso de éste modelo "PROD_NHB_2" (estimación por zonas), tenemos dos modelos a considerar teniendo en cuenta las correlaciones espaciales que son: el modelo **S_E_QUEEN – Spacial Error – Q1** en una matriz de contigüidad de primer orden y el modelo **S_E_ROOK – Spacial Error – R2** en la matriz de contigüidad de segundo orden. Haciendo un estudio más profundo, el mejor modelo a elegir ha sido el modelo **S_E_ROOK – Spacial Error – R2** porque sus parámetros presentan unos valores más ajustados y más significativos (ver Tabla 13).

- ✓ Analizando estos dos modelos, nos quedamos con el modelo **S_E_ROOK – Spacial Error – R2** en la matriz de contigüidad de segundo orden correspondiente al propósito de viajes "PROD_NHB_2" (estimación por zonas) como mejor modelo, porque nos presentan la variables independiente más significativa, también tiene mejor ajuste (R-squared) y nos ha permitido tener en cuenta las correlaciones espaciales por el valor de sus parámetros ya que son lo más adecuados, teniendo en cuenta su **Likelihood Ratio Test** (test de razón de verosimilitud) y su correspondiente índice de probabilidad que tienen unos valores bastante adecuados para nuestro estudio.

5.2.5 Viajes producidos (estimación por zonas)

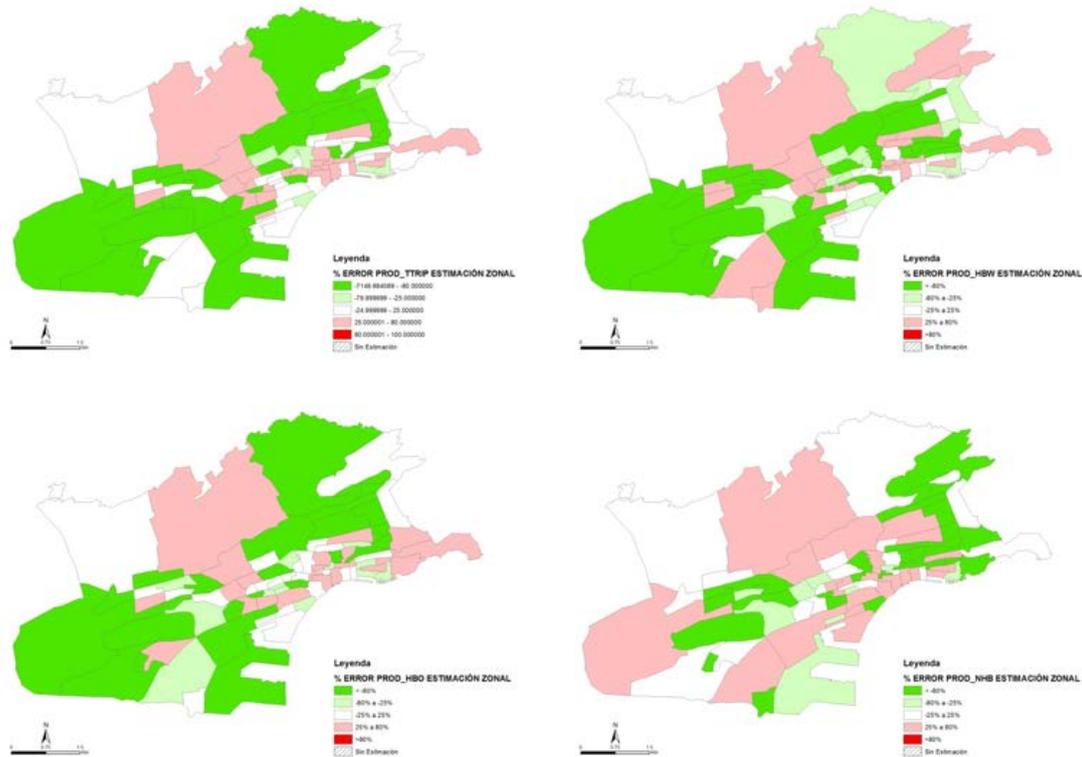


Figura 24: Figura 20 + Figura21 + Figura 22 + Figura 23

- **Análisis de viajes producidos (estimación por zonas)**

Se pueden observar ciertos patrones comunes a todos los modelos como por ejemplo que al norte de la Ciudad Jardín encontramos siempre zonas con problemas de sub-estimación o que en ninguno de los mapas existe ninguna zona con una sub-estimación de error importante. Pero sobre todo existen similitudes entre los modelos de viajes producidos totales y los basados en el hogar motivo trabajo y otros coincidiendo claramente los porcentajes de estimación, es decir, los colores de la leyenda en la mayoría de las zonas como la sub-estimación en la Magdalena y en el centro de la ciudad o fuerte sobre-estimación al norte, sur, oeste y suroeste, lo que puede hacer pensar a que está relacionado con las variables usadas en los modelos, que coinciden exactamente en algunos casos mientras que en otro existe una estrecha relación entre ellas, ya que en los totales se usan los hogares y la población, en los de motivo trabajo, la población, en los de motivo otros los m^2 habitables y la densidad de población por hectárea, estrechamente



relacionadas con las anteriores, y sin embargo los no basados en el hogar usan la accesibilidad activa, de ahí posiblemente su diferencia.

5.2.6 Viajes atraídos totales (estimación por zonas)

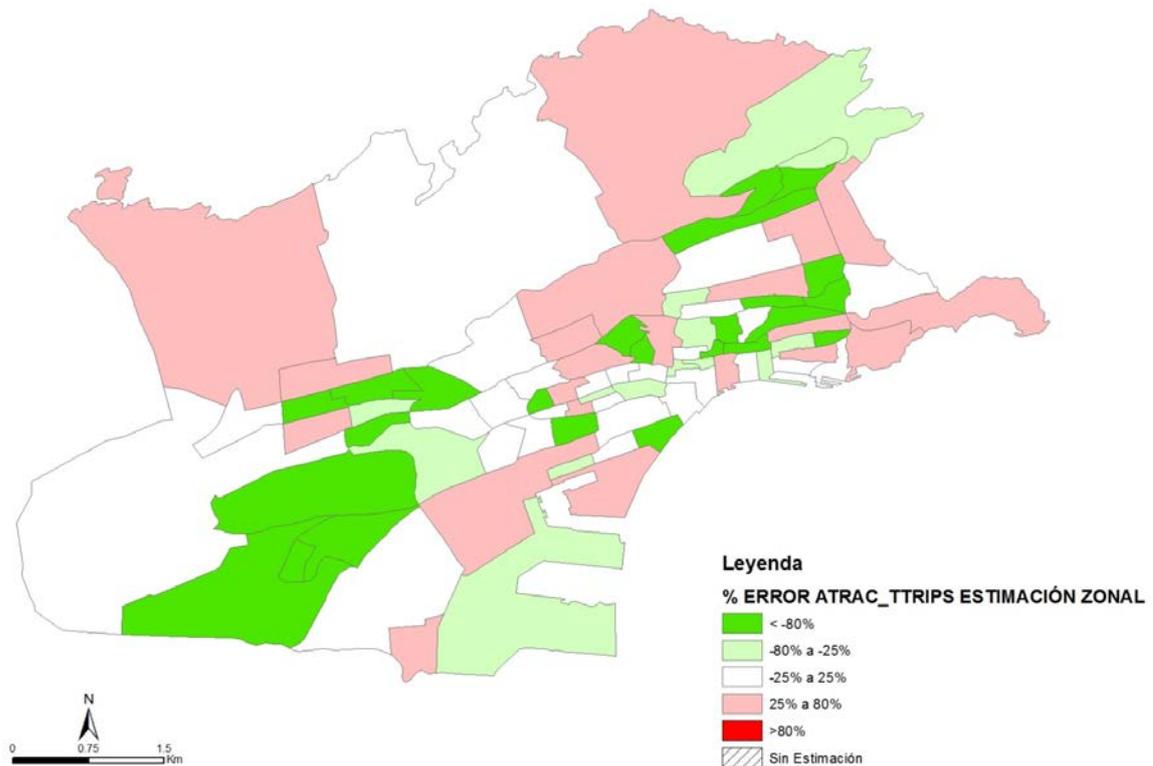


Figura 25: Planos viajes atraídos totales (estimación por zonas)

- **Análisis viajes atraídos totales (estimación por zonas)**

Este es un modelo de atracción de viajes y para la estimación de los mismos se ha utilizado dos variables que representan fielmente el concepto nombrado de atracción de viajes. En primer lugar se ha usado la variable que representa el equipamiento especial en la zona y en segundo término el número de locales comerciales que podemos encontrar en cada una de ellas. Es un modelo un poco extraño, ya que su R^2 es bastante elevado, el test “t” del intercepto y de las variables son peculiares. En primer lugar, el intercepto tiene un coeficiente con valor cercano a 2000 viajes y un test “t” mayor de 1,96 que viene a corroborar su importancia en la estimación de los viajes atraídos. Por otro lado tenemos las variables, primero la de equipamiento especial que consiste en 0 y 1, es decir, si o no con un coeficiente considerable y un test “t” bastante elevado también. En segundo lugar



tenemos los locales comerciales de la zona, que es una variable con un coeficiente adecuado al número de locales por el que se multiplica en cada zona y que además tiene un test “t” elevadísimo.

En el análisis de los datos representados en los planos podemos ver como bastantes zonas tienen un ajuste bastante bueno concentradas al oeste, al noroeste, en la universidad, en la Playa del Camello, en las estaciones, las zonas cercanas a la Ciudad Jardín y destacando alineadas a lo largo de las calles Castelar, Paseo Pereda, Calvo Sotelo, Jesús del Monasterio, Burgos y San Fernando. Desde el punto de vista de la subestimación no encontramos zonas pertenecientes al intervalo con mayor error pero sí al otro concentradas en la parte oeste, norte, Barrio Pesquero, algunas por el centro, Playa del Sardinero, la Magdalena y la Playa de los Peligros. En cuanto a las zonas con sobre estimación, existe de ambos intervalos, estando las de menor error dispersas sin ningún patrón aparente y las que tienen un error mayor se localizan en las zonas al norte de MercaSantander, incluyéndola, las que están en torno al Grupo Maria Blanchard, al norte de la universidad y en la zona sur del túnel de Tetuán.

• **Resultados de los modelos de viajes atraídos totales “ATTRAC_TTRIPS” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “ATTRAC_TTRIPS_1”							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q1	Spacial Lag R3	Spacial Error R1	Spacial Lag TD4	Spacial Error TD2
CONSTANT	2065,525	3367,715	2217,076	3178,033	2179,611	-2364,503	2028,369
	4,547	4,711	6,042	4,555	5,691	-4,069	5,323
FACILITY	6295,674	6439,533	6103,238	6415,454	6131,414	6362,267	6397,079
	4,745	5,112	4,702	5,057	4,705	4,831	4,924
Locales_Comerc.	19,182	21,627	18,308	21,369	18,534	19,384	19,390
	7,622	8,083	8,469	7,977	8,307	7,760	7,811
Parámetros de Ajustes							
W_ATTRAC_TTRIPS		-0,317		-0,273		0,896	
		-2,189		-1,963		12,085	
LAMBDA			-0,285		-0,223		-0,647
			-1,520		-1,249		-0,755
R-squared	0,501	0,536	0,521	0,529	0,514	0,490	0,511
Log likelihood	-805,455	-803,102	-804,306	-803,571	-804,681	-804,762	-804,663
S.E of regression	2826,560	2726,58	2769,550	2746,220	2788,300	2856	2798,730
Likelihood Ratio Test		0,030	0,129	0,052	0,213	0,239	0,208
		4,705	2,298	3,768	1,547	1,385	1,584

Tabla 14: Resumen del modelo “ATTRAC_TTRIPS_1” Viajes atraídos totales (Estimación por zonas)



En éste modelo “ATTRAC_TTRIPS_1” (estimación por zonas), sí que hemos tenido en cuenta las correlaciones espaciales porque los valores de los parámetros del modelo **S_E_QUEEN – Spacial Lag – Q1** en la matriz de contigüidad del primer orden han sido mejorado (ver Tabla 14).

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “ATTRAC_TTRIPS_2”							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q1	Spacial Lag R1	Spacial Error R2	Spacial Lag TD4	Spacial Error TD3
CONSTANT	2376,267	3431,396	2421,065	3275,102	2403,553	-2116,610	2389,056
	4,716	4,219	6,111	4,126	5,860	-3,552	5,509
Locales_Comerc.	19,610	21,587	19,431	21,374	19,545	19,819	19,534
	6,957	7,330	8,272	7,262	8,090	7,040	6,990
Parámetros de Ajustes							
W_ATTRAC_TTRIPS		-0,256		-0,219		0,909	
		-1,634		-1,461		14,068	
LAMBDA			-0,336		-0,285		-0,445
			-1,787		-1,592		-0,461
R-squared	0,365	0,391	0,401	0,386	0,394	0,352	0,370
Log likelihood	-815,777	-814,474	-814,102	-814,739	-814,466	-815,081	-815,363
S.E of regression	3187,02	3121,350	3095,320	3134,630	3115,180	3220,820	3174,590
Likelihood Ratio Test		0,106	0,067	0,149	0,105	0,238	0,363
		2,606	3,349	2,077	2,622	1,392	0,827

Tabla 15: Resumen del modelo “ATTRAC_TTRIPS_2” Viajes atraídos totales (Estimación por zonas)

En éste modelo “ATTRAC_TTRIPS_2” (estimación por zonas), las correlaciones espaciales mejoran un poco los parámetros de ajustes en los modelos pero no lo suficientemente buenos como para tenerlas en cuenta, porque en cada modelo, el **test razón de verosimilitud** es mucho menor que 3,84 que es el valor correspondiente 5% en la probabilidad estadística (Chi-Square) y también el **índice de probabilidad** en cada caso es superior a **0,05**. Como no se cumple ninguna de las condiciones mencionadas anteriormente, nos quedamos con el **modelo básico** (ver Tabla 15)

- ✓ Analizando estos dos modelos, nos quedamos en este caso con el modelo **S_E_QUEEN – Spacial Lag – Q1** en la matriz de contigüidad del primer orden correspondiente al propósito de viajes “ATTRAC_TTRIPS_1” (estimación por zonas) como mejor modelo, porque a pesar que tiene una variable de más que el modelo básico del propósito de viajes 2, nos presentan las variables independientes más significativa, también tiene mejor ajuste (R-squared) y nos ha permitido tener en cuenta las correlaciones espaciales por el valor de sus



parámetros ya que son lo más adecuados, teniendo en cuenta su **Likelihood Ratio Test** (test de razón de verosimilitud) y su correspondiente índice de probabilidad que tienen unos valores bastante adecuados para nuestro estudio y de esta forma se ha cumplido una vez más nuestro objetivo.

5.2.7 Viajes atraídos basados en el hogar motivo trabajo (estimación por zonas)

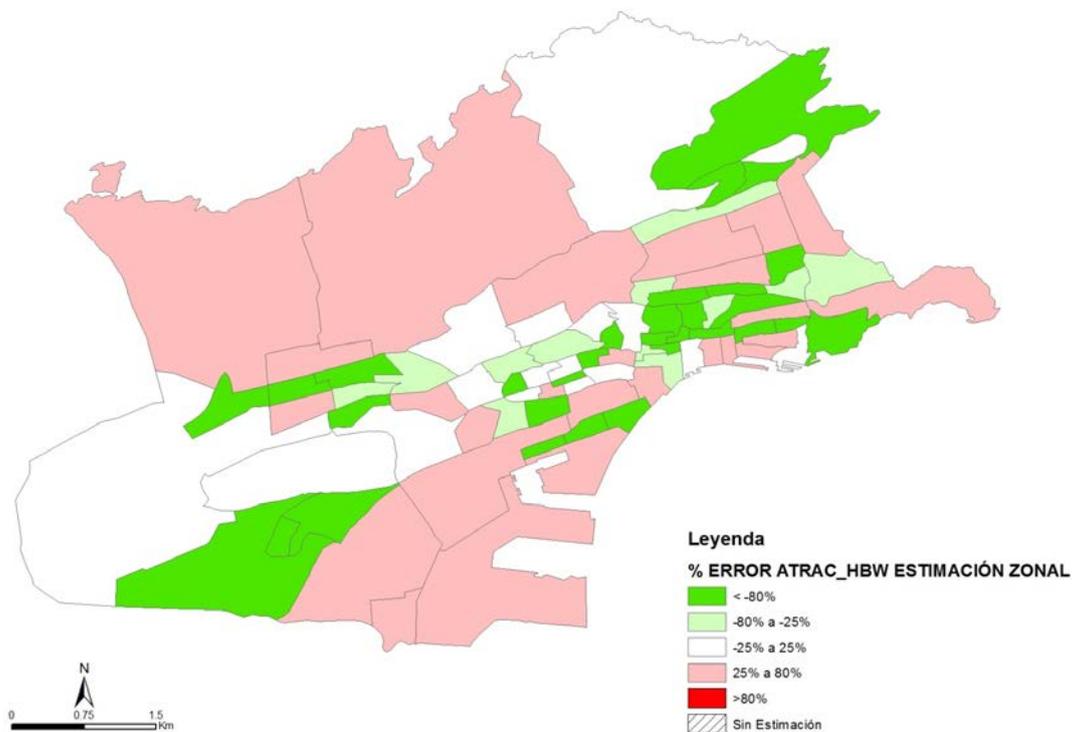


Figura 26: Viajes atraídos basados en el hogar motivo trabajo (estimación por zonas)

- **Análisis viajes atraídos basados en el hogar motivo trabajo (estimación por zonas)**

Para la elaboración de esta modelo he utilizado dos variables, siendo una de ellas la usada en el modelo de viajes atraídos totales, locales comerciales en la zona, y la otra es el número de empleos básicos que hay en cada una de ellas. Los valores de los viajes atraídos basados en el hogar motivo trabajo son bastante inferiores a los atraídos totales, por lo que se puede ver como el intercepto reduce considerablemente su valor y además tiene un test “t” cercano a 1,96 que demuestra un buen ajuste del modelo. En cuanto al R^2



tiene un valor aceptable. Finalmente, las variables tienen unos coeficientes adecuados y unos test “t” bastante elevados, sobre todo la de locales comerciales de la zona dando a entender que es la que marca la pauta más importante en la estimación de viajes.

Al visualizar el porcentaje de error de la estimación de viajes creadas por el modelo se ve como las zonas con un mejor ajuste están en el Palacio de Festivales, al norte, al oeste y al este de la Ciudad Jardín, incluida. Si analizamos la sub estimación se pueden diferenciar distintos patrones donde resulta la acumulación de zonas con error pequeño en todas las que hay a la margen derecha de la entrada del ferrocarril proveniente de Bilbao, incluyendo las estaciones, la Parte de Puerto Chico, la Magdalena, la parte final de la S-20, el estadio del Racing, la Playa del Sardinero y unas zonas al noroeste. Desde el punto de vista de la sobre estimación destacan las zonas con más error concentradas en la parte de Cabo Mayor y el campo de golf, en MercaSantander, en la parte del túnel de Tetuán, al sur de las estaciones y al norte y sur del Grupo María Blanchard.

- **Resultados de los modelos de viajes atraídos basados en el hogar motivo trabajo “ATTRAC_HBW” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “ATTRAC_HBW_1”							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q3	Spacial Lag R2	Spacial Error R3	Spacial Lag TD3	Spacial Error TD2
CONSTANT	349,638	227,080	344,957	244,030	355,165	-602,837	348,795
	1,960	1,028	2,172	1,108	1,931	-1,948	2,012
Básica	1,420	1,345	1,487	1,359	1,363	1,442	1,421
	2,579	2,501	2,726	2,524	2,541	2,631	2,628
Locales_Comerc.	5,426	5,114	5,461	5,163	5,426	5,523	5,435
	5,828	5,164	6,104	5,236	5,761	5,959	5,941
Parámetros de Ajustes							
W_ATTRAC_HBW		0,130		0,110		0,718	
		0,897		0,776		3,611	
LAMBDA			-0,290		0,082		-0,030
			-0,996		0,499		-0,050
R-squared	0,358	0,366	0,370	0,364	0,361	0,341	0,359
Log likelihood	-718,960	-718,578	-718,421	-718,661	-718,823	-717,854	-717,575
S.E of regression	1033,870	1027,650	1024,560	1029,040	1031,530	1047,350	1033,150
Likelihood Ratio Test		0,382	0,299	0,439	0,601	0,137	0,096
		0,765	1,078	0,599	0,274	2,213	2,771

Tabla 16: Resumen del modelo “ATTRAC_HBW_1” Viajes atraídos basados en el hogar motivo trabajo (Estimación por zonas)



En éste modelo “ATTRAC_HBW_1” (estimación por zonas), las correlaciones espaciales mejoran un poco los parámetros de ajustes en los modelos pero no lo suficientemente buenos como para tenerlas en cuenta, porque en cada modelo, el **test razón de verosimilitud** es mucho menor que 3,84 que es el valor correspondiente 5% en la probabilidad estadística (Chi-Square) y también el **índice de probabilidad** en cada caso es superior a **0,05**. Como no se cumple ninguna de las condiciones mencionadas anteriormente, nos quedamos con el **modelo básico** (ver Tabla 16).

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “ATTRAC_HBW_2”							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q1	Spacial Lag R3	Spacial Error R1	Spacial Lag TD3	Spacial Error TD2
CONSTANT	408,861	309,685	391,282	327,194	391,230	-130,834	395,780
	2,790	1,667	2,475	1,767	2,499	-0,235	3,079
FACILITY	2430,505	2388,643	2381,082	2397,132	2387,078	2449,960	2438,674
	5,679	5,687	5,752	5,703	5,758	5,779	5,836
Locales_Comerc.	5,632	5,380	5,736	5,429	5,736	5,688	5,733
	6,936	5,938	6,846	6,020	6,879	7,070	7,202
Parámetros de Ajustes							
W_ATTRAC_HBW		0,099		0,081		0,407	
		0,731		0,608		0,991	
LAMBDA			0,137		0,124		-0,351
			0,825		0,765		-0,470
R-squared	0,501	0,505	0,507	0,504	0,506	0,493	0,511
Log likelihood	-708,161	-707,895	-707,799	-707,969	-707,843	-707,116	-706,462
S.E of regression	911,867	908,207	906,431	909,245	907,101	918,386	902,868
Likelihood Ratio Test		0,465	0,394	0,535	0,425	0,148	0,065
		0,533	0,725	0,384	0,637	2,090	3,399

Tabla 17: Resumen del modelo “ATTRAC_HBW_2” Viajes atraídos basados en el hogar motivo trabajo (Estimación por zonas)

En éste modelo “ATTRAC_HBW_2” (estimación por zonas), las correlaciones espaciales mejoran un poco los parámetros de ajustes en los modelos pero no lo suficientemente buenos como para tenerlas en cuenta, porque en cada modelo, el **test razón de verosimilitud** es mucho menor que 3,84 que es el valor correspondiente 5% en la probabilidad estadística (Chi-Square) y también el **índice de probabilidad** en cada caso es superior a **0,05**. Como no se cumple ninguna de las condiciones mencionadas anteriormente, nos quedamos con el **modelo básico** (ver Tabla 17).

- ✓ Analizando estos dos modelos, nos quedamos en este caso con el **modelo básico_2** que corresponde al propósito de viajes “ATTRAC_HBW_2” (estimación por zonas) como mejor modelo, porque a pesar que en ninguno de



los dos modelos básicos se ha considerado las correlaciones espaciales, nos presentan las variables independientes más significativas, también teniendo en cuenta el (R-squared), sus parámetros y coeficientes se ajustan mejor.

5.2.8 Viajes atraídos basados en el hogar motivo otros (estimación por zonas)

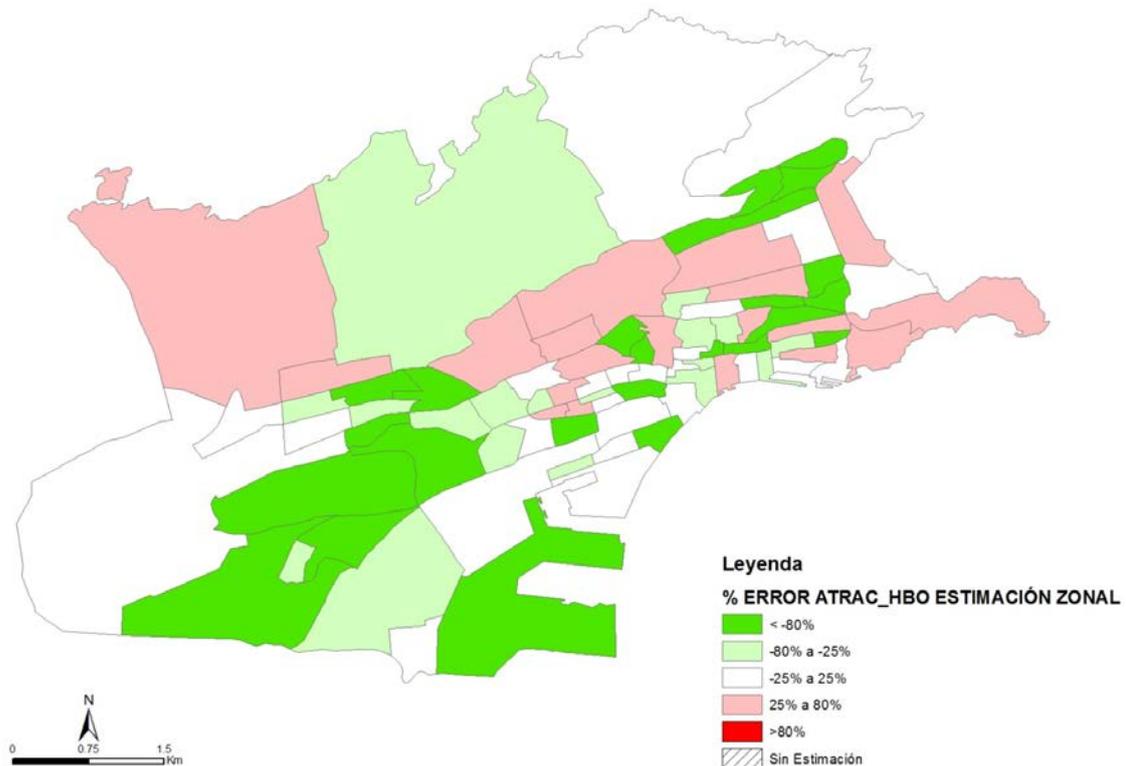


Figura 27: Viajes atraídos basados en el hogar motivo otros (estimación por zonas)

- **Análisis viajes atraídos basados en el hogar por otros motivos (estimación por zonas)**

Para la elaboración de este modelo he utilizado exactamente las mismas variables que usé para el modelo de viajes atraídos totales, estas son la de existencia o no de equipamiento especial en la zona y la del número de locales comerciales que hay en cada una de ellas. El ajuste del R^2 es aceptable. En cuanto al intercepto, tiene un peso importante en la estimación no sólo por el valor de su coeficiente sino también por la importancia de su test “t” bastante por encima del 1,96. En cuanto a las dos variables, en caso de existencia de equipamiento especial supone un incremento directo de los viajes atraídos de unos 3500, pero si analizamos el test “t” de ambas variables podemos ver como



la que lleva la parte importante del modelo es la de locales comerciales en cada una de las zonas porque casi tiene un test “t” dos veces superior en valor al de la otra variable.

Analizando los datos representados en los mapas podemos ver como las zonas con una buena estimación están al oeste, norte Cabo Mayor, estadio del Racing, Playa del Camello, Barrio Pesquero, estaciones, Palacio de Festivales y dispersas por el centro. La mayor parte de la sub estimación, perteneciente al intervalo de menor error, se encuentra a lo largo de la entrada a Santander por la autovía S-20, pero también se pueden encontrar zonas al oeste, la Playa del Sardinero, la Magdalena, la Playa de los Peligros y por el centro urbano sin concentración remarcable. La sobre estimación con menor error la podemos encontrar puntualmente por el centro, noroeste, suroeste y concentradas en torno a la Ciudad Jardín. La que tiene un error considerable está localizada en el puerto de Raos, la margen izquierda de la entrada del ferrocarril proveniente de Bilbao, al norte de las universidades, en la parte del túnel de Tetuán y en la zona del grupo María Blanchard.

- **Resultados de los Modelos de viajes atraídos basados en el hogar por otros motivos “ATTRAC_ HBO” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “ATTRAC_ HBO_1”							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q1	Spacial Lag R3	Spacial Error R1	Spacial Lag TD4	Spacial Error TD4
CONSTANT	1637,778	2456,869	1717,775	2265,160	1682,320	-1446,082	1639,264
	4,509	4,351	5,411	4,124	5,056	-3,324	0,470
FACILITY	3567,856	3575,241	3397,805	3572,04	3460,985	3605,587	3567,304
	3,363	3,510	3,251	3,477	3,309	3,424	3,424
Locales_Comerc.	12,067	13,284	11,604	13,085	11,814	12,194	12,076
	5,997	6,332	6,337	6,203	6,247	6,104	6,107
Parámetros de Ajustes							
W_ATTRAC_HBO		-0,276		-0,214		0,900	
		-1,737		-1,411		12,673	
LAMBDA			-0,171		-0,103		0,929
			-0,930		-0,588		18,468
R-squared	0,370	0,401	0,380	0,390	0,374	0,357	0,357
Log likelihood	-786,224	-784,639	-785,787	-785,192	-786,050	-785,530	-785,544
S.E of regression	2260,180	2204,330	2242,950	2223,620	2253,350	2283,870	2284,300
Likelihood Ratio Test		0,075	0,350	0,151	0,555	0,239	0,244
		3,170	0,873	2,064	0,348	1,388	1,360

Tabla 18: Resumen del modelo “ATTRAC_ HBO_1” Viajes atraídos basados en el hogar por otros motivos (Estimación por zonas)

En éste modelo “ATTRAC_ HBO_1” (estimación por zonas), las correlaciones espaciales mejoran un poco los parámetros de ajustes en los modelos pero no lo



suficientemente buenos como para tenerlas en cuenta, porque en cada modelo, el **test razón de verosimilitud** es mucho menor que **3,84** que es el valor correspondiente **5%** en la probabilidad estadística (Chi-Square) y también el **índice de probabilidad** en cada caso es superior a **0,05**. Como no se cumple ninguna de las condiciones mencionadas anteriormente, nos quedamos con el **modelo básico** (ver Tabla 18)

REGRESIÓN - MEJORES RESULTADOS DEL MODELO "ATTRAC_HBO_2"							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q1	Spacial Lag R1	Spacial Error R3	Spacial Lag TD3	Spacial Error TD3
CONSTANT	1813,880	2588,289	1864,723	2407,751	1843,153	3811,938	1817,552
	4,762	4,255	5,909	4,068	5,592	1,081	5,834
Locales_Comerc.	12,309	13,460	11,998	13,273	12,141	12,182	12,266
	5,776	6,159	6,474	6,046	6,332	5,784	5,788
Parámetros de Ajustes							
W_ATTRAC_HBO		-0,261		-0,203		-0,578	
		-1,586		-1,290		-0,570	
LAMBDA			-0,261		-0,197		-0,791
			-1,395		-1,108		-0,679
R-squared	0,284	0,314	0,310	0,304	0,300	0,291	0,293
Log likelihood	-791,718	-790,414	-790,645	-790,866	-791,050	-791,326	-791,399
S.E of regression	2409,270	2359,080	2365,430	2376,330	2381,970	2397,380	2394,730
Likelihood Ratio Test		0,106	0,143	0,192	0,248	0,375	0,425
		2,606	2,146	1,703	1,335	0,784	0,637

Tabla 19: Resumen del modelo "ATTRAC_HBO_2" Viajes atraídos basados en el hogar por otros motivos" (Estimación por zonas)

En éste modelo "ATTRAC_HBO_2" (estimación por zonas), las correlaciones espaciales mejoran un poco los parámetros de ajustes en los modelos pero no lo suficientemente buenos como para tenerlas en cuenta, porque en cada modelo, el **test razón de verosimilitud** es mucho menor que **3,84** que es el valor correspondiente **5%** en la probabilidad estadística (Chi-Square) y también el **índice de probabilidad** en cada caso es superior a **0,05**. Como no se cumple ninguna de las condiciones mencionadas anteriormente, nos quedamos con el **modelo básico** (ver Tabla 19).

- ✓ Haciendo en análisis comparativo de estos dos modelos, nos quedamos en este caso con el **modelo básico** para el propósito de viajes "ATTRAC_HBO_1" (estimación por zonas) como mejor modelo, porque a pesar que en ninguno de los dos modelos básicos se ha considerado las correlaciones espaciales, éste modelo elegido tiene una variable más y nos presentan las variables

independientes más significativas, también teniendo en cuenta el (R-squared), sus parámetros y coeficientes se ajustan mejor.

5.2.9 Viajes atraídos no basados en el hogar (estimación por zonas)

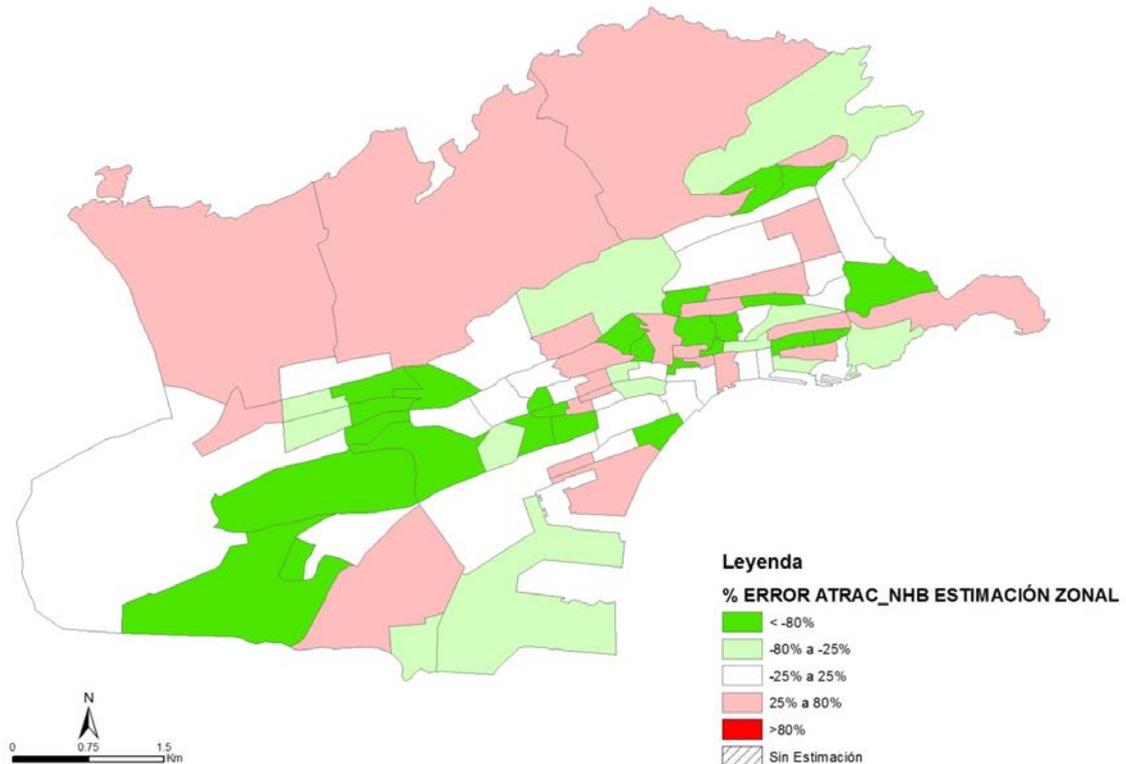


Figura 28: Viajes atraídos no basados en el hogar (estimación por zonas)

- **Análisis viajes atraídos no basados en el hogar (estimación por zonas)**

Para la estimación de los viajes atraídos no basados en el hogar vuelvo a usar un modelo que ya había utilizado anteriormente que consiste en la utilización de las variables de número de locales comerciales y de la existencia o no de equipamiento especial en la zona. Vuelve a tener la misma estructura que en casos anteriores, con un intercepto elevado en test “t” que demuestra su importancia y con un valor de coeficiente que es significativo a la hora de estimar los viajes. Con las variables pasa lo mismo, en caso de existencia de equipamientos especiales se suma casi 300 viajes. Además se puede ver



como la variable del número de locales comerciales tiene un test “t” elevado lo que le da un papel importante como en los modelos anteriores. Además el R^2 también es alto.

En los planos se puede ver como las zonas con una buena estimación de viajes se sitúan al oeste, en la universidad, en torno a la Ciudad Jardín, la Playa del Sardinero y zonas cercanas a esta y toda la entrada a Santander desde el polígono industrial de Candina hasta el Palacio de Festivales. La sub estimación con un error menor está dispersa por el centro, en la Magdalena, en el Barrio Pesquero y en las tres grandes zonas al norte que limitan con el Mar Cantábrico. Desde el punto de vista de la sobre estimación es más importante la que tiene un intervalo de mayor error, aunque la de error pequeño se puede localizar por ejemplo en Cabo Mayor y el Puerto de Raos, sin ninguna concentración fuerte de zonas, que si se manifiesta en el otro grupo de zonas en toda la margen izquierda de la entrada a Santander del ferrocarril proveniente de Bilbao y otras, eso sí dispersas, las encontramos por el resto de la ciudad.

- **Resultados de los modelos de Viajes atraídos no basados en el hogar “ATTRAC_NHB” (estimación por zonas)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “ATTRAC_NHB_1”							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q2	Spacial Error Q1	Spacial Lag R2	Spacial Error R1	Spacial Lag TD3	Spacial Error TD4
CONSTANT	158,541	82,819	154,167	186,365	155,334	17,909	157,786
	4,360	1,189	5,455	3,428	5,239	0,109	0,615
FACILITY	292,242	282,697	247,136	292,894	255,807	294,813	292,329
	2,752	2,734	2,389	2,805	2,458	2,810	2,803
Locales_Comerc.	1,468	1,398	1,531	1,542	1,513	1,479	1,469
	7,285	6,855	9,109	7,051	8,685	7,423	7,422
Parámetros de Ajustes							
W_ATTRAC_NHB		0,223		-0,099		0,376	
		1,261		-0,691		0,870	
LAMBDA			-0,342		-0,272		0,903
			-1,815		-1,520		13,175
R-squared	0,428	0,438	0,459	0,432	0,449	0,422	0,416
Log likelihood	-588,298	-587,691	-586,786	-588,061	-587,274	-587,509	-587,624
S.E of regression	226,273	224,135	220,111	225,438	222,059	227,431	228,606
Likelihood Ratio Test		0,270	0,082	0,491	0,152	0,209	0,246
		1,216	3,024	0,474	2,049	1,578	1,348

Tabla 20: Resumen del modelo “ATTRAC_NHB_1” Viajes atraídos no basados en el hogar (Estimación por zonas)

En éste modelo “ATTRAC_NHB_1” (estimación por zonas), las correlaciones espaciales mejoran un poco los parámetros de ajustes en los modelos pero no lo suficientemente buenos como para tenerlas en cuenta, porque en cada modelo, el **test razón de verosimilitud** es mucho menor que **3,84** que es el valor correspondiente **5%** en



la probabilidad estadística (Chi-Square) y también el **índice de probabilidad** en cada caso es superior a **0,05**. Como no se cumple ninguna de las condiciones mencionadas anteriormente, nos quedamos con el **modelo básico** (ver Tabla 20).

- ✓ Este modelo no se compara con ninguno por la simple razón que haciendo el estudio de la calibración con otras variables nos salen no significativas el test “t” o sea bastante menos que 1,96.

5.2.10 Viajes atraídos (estimación por zonas)

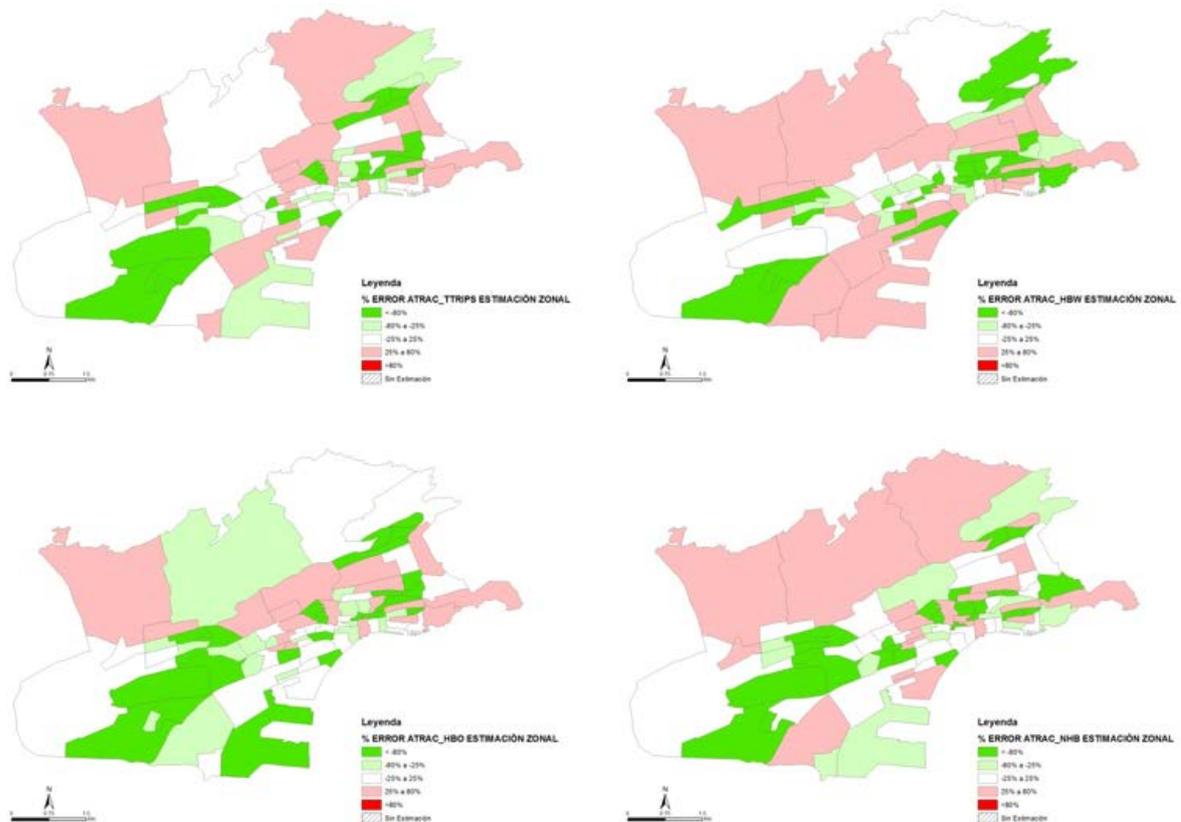


Figura 29: Figura 25 + Figura 26 + Figura 27 + Figura 28

- **Análisis viajes atraídos (estimación por zonas)**

Analizando los cuatro mapas expuesto en la tabla se puede observar claramente si trazamos dos rectas paralelas a la costa norte del área de estudio, una coincidente con la Avenida de los Castros y otra a la altura de Puerto Chico, como existe una zona norte relacionada con la sub estimación, una zona central donde destaca la sobre estimación, sobre todo la del intervalo que representa una mayor error en la estimación y finalmente, una parte sur donde destaca la sub estimación excepto en uno de los planos, el de los viajes atraídos basados en el hogar motivo otros. Esto parece estar también



biunívocamente relacionado con la utilización de determinadas variables ya que en 3 de los planos se ha utilizado el mismo modelo basado en el número de locales comerciales de la zona y la existencia o no de equipamientos especiales, siendo solo el modelos de viajes atraídos basados en el hogar motivo trabajo el que sustituye la variable de equipamiento especial en la zona por la de empleos básicos, totalmente razonable si se tiene en cuenta la categoría de viajes.

5.3 Modelos de viajes estimados por hogares

5.3.1 Viajes producidos basados en el hogar motivo trabajo (estimación por hogares)

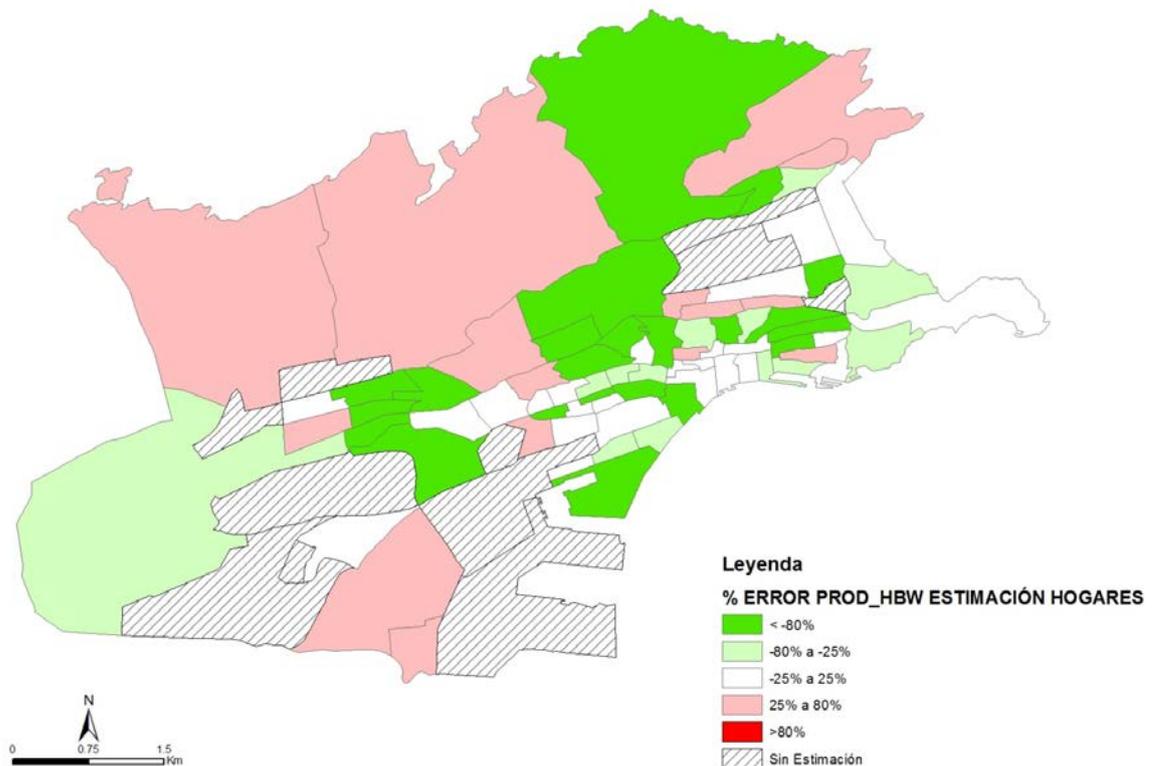


Figura 30: Viajes producidos basados en el hogar motivo trabajo (estimación por hogares)

- **Análisis viajes producidos basados en el hogar motivo trabajo (estimación por hogares)**

En la elaboración de este modelo de estimación de viajes con datos por hogares he utilizado una serie de variables que son las siguientes: número de trabajadores del hogar y el número de vehículos que existen en cada hogar. En el examen de los distintos



parámetros del modelo podemos darnos cuenta de que ajusta muy bien ya que el R^2 es muy elevado y este último tiene un coeficiente y test “t” insignificantes. Por otro lado al observar las variables podemos ver la importancia de ambas (las dos por encima claramente del 1,96) pero resalta entre ellas la del número de trabajadores que existen en el hogar, hay que tener en cuenta que estamos estimando viajes producidos basados en el hogar motivo trabajo, ya que su test “t” es elevadísimo.

En este caso, podemos ver en el plano unas zonas marcadas a rayas lo que significa que no se dispone de datos para realizar la estimación de viajes. Las zonas con una buena estimación de viajes se localizan en la Playa del Sardinero, el estadio del Racing, al sur de la zona de las universidades, la Magdalena, el palacio de Festivales, en torno a la Ciudad Jardín y la estación de ferrocarril y a lo largo de la calle de Calvo Sotelo y el Paseo Pereda. Desde el punto de vista de la sub estimación sólo hay zonas que pertenecen al intervalo de menor error, situadas al oeste, noroeste, en Cabo Mayor, justo en la entrada a Santander por la S-10 en la margen izquierda y en las zonas al sur de las universidades. En cuanto a la sobre estimación de pequeño error destacan tres zonas alineadas a lo largo de las calles San Fernando, Burgos y Jesús del Monasterio. Sin embargo es más importante la de mayor error concentrada en el Barrio Pesquero, al sur del Grupo María Blanchard, incluido, y rodeando por norte, oeste y sur la zona de las universidades que carece de estimación de viajes.

- **Resultados de los modelos de viajes producidos basados en el hogar motivo trabajo “PROD_HBW” (estimación por hogares)**

En éste modelo “**PROD_HBW_1**” (estimación por hogares), todos los modelos espaciales cumplen con todas las condiciones requeridas para tener en cuenta las correlaciones espaciales porque los valores de sus parámetros se ajustan más, sus test “t” (test estadístico) han sido muy buenos y siguen siendo significativos o sea, mayor que 1,96 y su R^2 (coeficiente de ajuste) han sido un poco más ajustado que él del modelo básico. También hemos tenido en cuenta su **test razón de verosimilitud** que es mucho **mayor que 3,84** y por supuesto corresponde que su índice o su **valor de probabilidad** que es mucho **menor que 0,05**. Analizando todos los términos, el mejor modelo de todos



los modelos corresponde en éste caso al **S_E_QUEEN – Spacial Error– Q2** en una matriz de contigüidad de segundo orden (ver Tabla 21).

REGRESIÓN - MEJORES RESULTADOS DEL MODELO "PROD_HBW_1"							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q2	Spacial Error Q2	Spacial Lag R1	Spacial Error R1	Spacial Lag TD1	Spacial Error TD1
CONSTANT	0,010	-0,416	0,017	-0,269	0,020	-0,424	0,009
	0,126	-2,871	0,173	-2,367	0,235	-2,075	0,090
WORKERS	1,934	1,931	1,937	1,913	1,921	1,928	1,927
	32,350	32,584	32,956	32,126	32,337	32,385	32,450
VEH.	0,191	0,191	0,185	0,188	0,198	0,192	0,200
	2,846	2,882	2,795	2,840	2,979	2,876	3,003
Parámetros de Ajustes							
W_PROD_HBW		0,219		0,153		0,221	
		3,436		3,342		2,271	
LAMBDA			0,387		0,231		0,366
			4,299		3,691		2,702
R-squared	0,633	0,639	0,645	0,639	0,641	0,635	0,637
Log likelihood	-1383,030	-1377,020	1372,7110	-1377,620	-1376,055	-1380,410	-1379,561
S.E of regression	1,31505	1,3037	1,293	1,304	1,299	1,310	1,308
Likelihood Ratio Test		0,00053	0,00001	0,00101	0,00019	0,02211	0,00845
		12,0215	20,6334	10,8131	13,9459	5,2369	6,9348

Tabla 21: Resumen del modelo 1 - Viajes producidos basados en el hogar motivo trabajo "PROD_HBW_1" (estimación por hogares)

REGRESIÓN - MEJORES RESULTADOS DEL MODELO "PROD_HBW_2"							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q2	Spacial Error Q2	Spacial Lag R1	Spacial Error R1	Spacial Lag TD1	Spacial Error TD1
CONSTANT	0,123	-0,301	0,126	-0,159	0,138	-0,306	0,127
	1,810	-2,143	1,408	-1,473	1,782	-1,489	1,477
WORKERS	2,008	2,005	2,010	1,987	1,998	2,002	2,004
	37,191	37,421	38,001	36,798	37,138	37,180	37,289
Parámetros de Ajustes							
W_PROD_HBW		0,219		0,154		0,219	
		3,396		3,349		2,203	
LAMBDA			0,389		0,225		0,348
			4,327		3,591		2,516
R-squared	0,629	0,635	0,642	0,635	0,637	0,632	0,633
Log likelihood	-1387,070	-1381,150	-1376,600	-1381,640	-1380,466	-1384,530	-1384,036
S.E of regression	1,322	1,310	1,299	1,311	1,307	1,317	1,315
Likelihood Ratio Test		0,00058	0,00000	0,00098	0,00028	0,02404	0,01371
		11,8479	20,9476	10,8703	13,2147	5,0914	6,0750

Tabla 22: Resumen del modelo 2 - Viajes producidos basados en el hogar motivo trabajo "PROD_HBW_2" (estimación por hogares)



En éste modelo “**PROD_HBW_2**” (estimación por hogares), todos los modelos espaciales cumplen con todas las condiciones requeridas para tener en cuenta las correlaciones espaciales porque los valores de sus parámetros se ajustan más, sus test “t” (test estadístico) han sido muy buenos y siguen siendo significativos o sea, mayor que 1,96 y su R^2 (coeficiente de ajuste) han sido un poco más ajustado que él del modelo básico. También hemos tenido en cuenta su **Test razón de verosimilitud** que es mucho **mayor que 3,84** y por supuesto corresponde que su índice o su **valor de probabilidad** que es mucho **menor que 0,05**. Analizando todos los términos, el mejor modelo de todos los modelos corresponde en éste caso al **S_E_QUEEN – Spacial Error– Q2** en una matriz de contigüidad de segundo orden (ver Tabla 22).

- ✓ Haciendo un estudio comparativo de estos dos modelos, nos quedamos en este caso con los dos modelos, o sea con el modelo **S_E_QUEEN – Spacial Error– Q2** en una matriz de contigüidad de segundo orden tanto para el propósito de viajes “**PROD_HBW_1**” (estimación por hogares) como para el propósito de viajes “**PROD_HBW_2**” (estimación por hogares) como mejores modelos, porque a pesar que el modelo 1 tiene una variable más que el modelo 2, los dos modelos nos presentan las variables independientes más significativas, también tiene mejor ajuste (R-squared) muy parecido y nos ha permitido tener en cuenta las correlaciones espaciales por el valor de sus parámetros ya que son lo más adecuados y bastante parecidos y también teniendo en cuenta sus **Likelihood Ratio Test** (test de razón de verosimilitud) y su correspondiente índice de probabilidad que tienen unos valores bastante parecidos y adecuados para nuestro estudio; de esta forma se ha cumplido una vez más nuestro objetivo.

5.3.2 Viajes producidos basados en el hogar motivo otros (estimación por hogares)

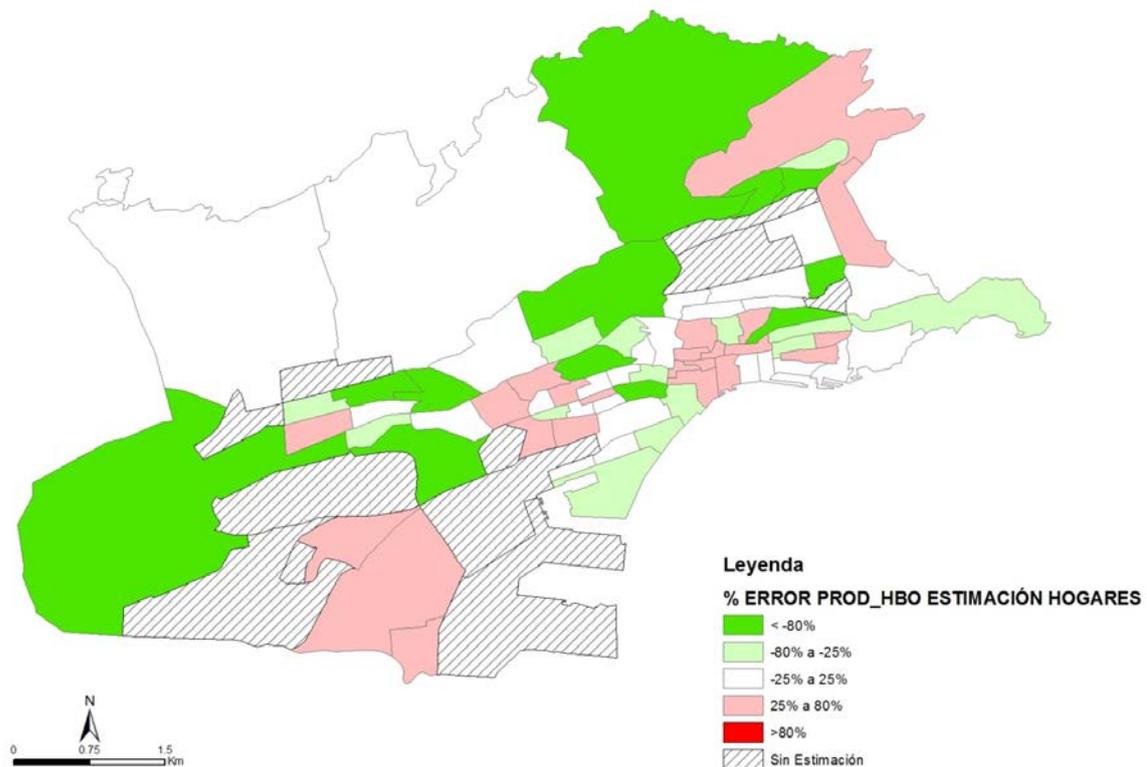


Figura 31: Viajes producidos basados en el hogar motivo otros (estimación por hogares)

- **Análisis viajes producidos basados en el hogar motivos otros (estimación por hogares)**

Desde el punto de vista del modelo utilizado para la estimación de viajes, he utilizado como variables el número de miembros que existen en el hogar y la densidad de empleos por hectárea que existe en la zona del hogar, lo que es una variable de carácter zonal. En cuanto al modelo, tenemos un R^2 no excesivamente alto junto con un intercepto de coeficiente y test “t” bastante pequeños. Al analizar las variables, las dos tienen test “t” elevados pero resalta por el resultado de los mismos la importancia de la variable que consiste en el número de miembros del hogar. Con todo esto, este es un modelo que ajusta bastante bien.

Desde el punto de vista del análisis de los datos plasmados en los planos mediante los porcentajes de error de las estimaciones, podemos ver como algunas zonas están rayadas, estas carecen de estimación, y las de mejor ajuste se localizan al oeste, al noroeste, en el estadio del Racing, la Playa del Camello, al sur de la universidades, en la estación de ferrocarriles y desde el Paseo Pereda hasta la Playa de los Peligros. La sub



estimación de menor error, está localizada en la entrada de Santander por la S-10, en Cabo Mayor y la playa del Sardinero y en torno a la Ciudad Jardín y las zonas al este del Ayuntamiento. Hay algunas zonas con sobre estimación pequeña como la Magdalena o el Barrio Pesquero, destacando las que tienen un mayor error concentradas al oeste y norte de las universidades y también al oeste de la zona de estudio.

- **Resultados de los modelos de viajes producidos basados en el hogar motivos otros “HBO” (estimación por hogares)**

REGRESIÓN - MEJORES RESULTADOS DEL MODELO 1							
Variables	S_E_Modelo Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q1	Spacial Lag R2	Spacial Error R2	Spacial Lag TD1	Spacial Error TD1
CONSTANT	-0,122	-1,154	-0,010	-1,558	0,165	-1,489	-0,006
	-0,480	-3,675	-0,036	-4,073	0,580	-3,033	-0,022
SIZE	1,802	1,771	1,776	1,753	1,747	1,776	1,779
	17,931	18,006	18,086	17,781	17,584	17,786	17,752
DEN_EMP_HA	0,004	0,003	0,004	0,003	0,003	0,004	0,004
	5,264	3,703	3,364	3,544	2,524	4,101	4,022
Parámetros de Ajustes							
W_HBO		0,278		0,386		0,347	
		5,280		5,153		3,283	
LAMBDA			0,319		0,406		0,366
			5,443		4,567		2,711
R-squared	0,305	0,336	0,337	0,334	0,326	0,315	0,311
Log likelihood	-1976,450	-1961,820	-1962,296	-1962,690	-1967,624	-1971,640	-1973,745
S.E of regression	2,719	2,657	2,654	2,662	2,677	2,700	2,706
Likelihood Ratio Test		0,00000	0,00000	0,00000	0,00003	0,00193	0,02012
		29,2476	28,3002	27,5073	17,6456	9,6195	5,4019

Tabla 23: Resumen del modelo 1 - Viajes producidos basados en el hogar motivos otros “HBO” (estimación por hogares)

En éste modelo “**PROD_HBO_1**” (estimación por hogares), todos los modelos espaciales cumplen con todas las condiciones requeridas para tener en cuenta las correlaciones espaciales porque los valores de sus parámetros se ajustan más, sus test “t” (test estadístico) han sido muy buenos y siguen siendo significativos o sea, mayor que 1,96 y su R² (coeficiente de ajuste) han sido un poco más ajustado que el del modelo básico. También hemos tenido en cuenta su **Test razón de verosimilitud** que es **mucho mayor que 3,84** y por supuesto corresponde que su índice o su **valor de probabilidad** que es **mucho menor que 0,05**. Analizando todos los términos, el mejor modelo de todos



los modelos corresponde en éste caso al **S_E_QUEEN – Spacial Error– Q1** en una matriz de contigüidad de primer orden (ver Tabla 23); aunque comparando bien, todos los demás modelos tienen los valores muy similares al modelo elegido.

REGRESIÓN - MEJORES RESULTADOS DEL MODELO “PROD_HBO_2”							
Variables	S_E_Model o Básico	S_E_Queen		S_E_Rook		S_E_Threshold Distance	
		Spacial Lag Q1	Spacial Error Q1	Spacial Lag R2	Spacial Error R2	Spacial Lag TD1	Spacial Error TD1
CONSTANT	-0,125	-1,291	0,013	-1,759	0,165	-1,852	0,001
	-0,472	-4,025	0,044	-4,532	0,553	-3,867	0,004
SIZE	1,819	1,779	1,780	1,758	1,750	1,782	1,782
	17,982	18,068	18,140	17,816	17,670	17,811	17,767
EMP.	0,001	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005
	3,947	3,091	2,602	3,024	2,274	3,347	3,242
Parámetros de Ajustes							
W_HBO		0,305		0,427		0,429	
		5,920		5,819		4,311	
LAMBDA			0,349		0,460		0,478
			6,113		5,390		4,113
R-squared	0,295	0,334	0,336	0,333	0,328	0,312	0,310
Log likelihood	-1982,380	-1963,760	-1964,086	-1964,090	-1967,587	-1974,150	-1975,702
S.E of regression	2,739	2,661	2,656	2,664	2,673	2,706	2,709
Likelihood Ratio Test		0,00000	0,00000	0,00000	0,00000	0,00005	0,00026
		37,2435	36,5881	36,5714	29,5852	16,4620	13,3567

Tabla 24: Resumen del modelo 2 - Viajes producidos basados en el hogar motivos otros “HBO” (estimación por hogares)

En éste modelo “**PROD_HBO_2**” (estimación por hogares), todos los modelos espaciales cumplen con todas las condiciones requeridas para tener en cuenta las correlaciones espaciales porque los valores de sus parámetros se ajustan más, sus test “t” (test estadístico) han sido muy buenos y siguen siendo significativos o sea, mayor que 1,96 y su R^2 (coeficiente de ajuste) han sido un poco más ajustado que él del modelo básico. También hemos tenido en cuenta su **Test razón de verosimilitud** que es **mucho mayor que 3,84** y por supuesto corresponde que su índice o su **valor de probabilidad** que es **mucho menor que 0,05**. Analizando todos los términos, el mejor modelo de todos los modelos corresponde en éste caso al **S_E_QUEEN – Spacial Error– Q1** en una matriz de contigüidad de primer orden (ver Tabla 24), aunque comparándolo a los demás modelos, el modelo elegido tiene los valores muy similares.

- ✓ Haciendo un estudio comparativo de estos dos modelos, nos quedamos en este caso con los dos modelos, o sea con el modelo **S_E_QUEEN – Spacial Error– Q1** en una matriz de contigüidad de primer orden tanto para el propósito de viaje



“**PROD_HBO_1**” (estimación por hogares) como para el propósito de viajes
 “**PROD_HBW_2**” (estimación por hogares) como mejores modelos, porque a pesar que los dos modelos tienen dos variables cada uno, los dos modelos nos presentan las variables independientes más significativas, también tienen mejores ajustes (R-squared) muy parecidos y nos ha permitido tener en cuenta las correlaciones espaciales por el valor de sus parámetros ya que son lo más adecuados y bastante parecidos y también teniendo en cuenta sus **Likelihood Ratio Test** (test de razón de verosimilitud) y su correspondiente índice de probabilidad que tienen unos valores bastante parecidos y adecuados para nuestro estudio; de esta forma se ha cumplido una vez más nuestro objetivo.

5.3.3 Viajes producidos basados en el hogar (estimación por hogares)

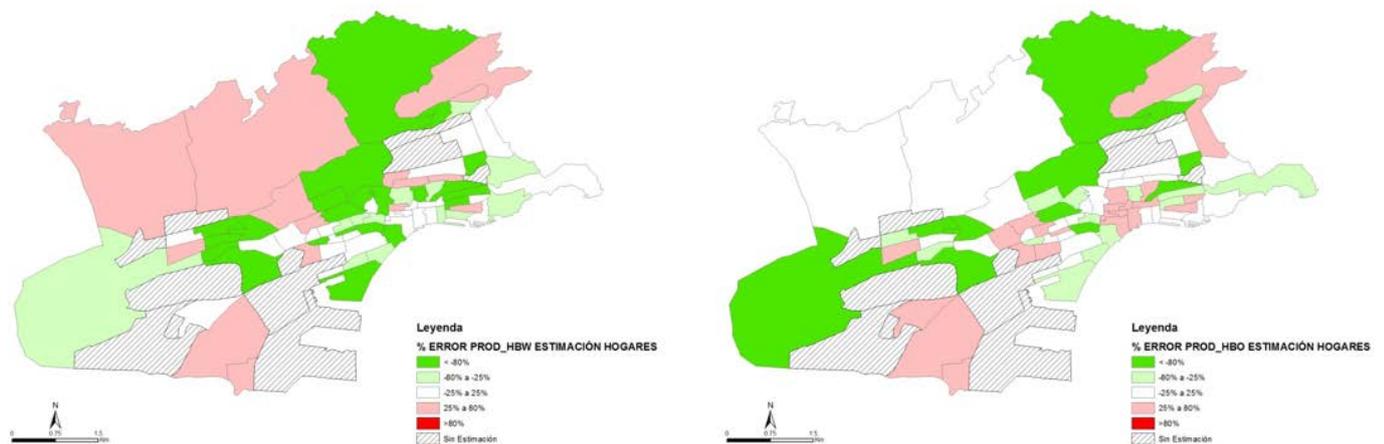


Figura 32: Figura 30 + Figura 31

- **Análisis viajes producidos (estimación por hogares)**

En el estudio de estos dos mapas, parece como que algunas zonas coinciden en aspecto como al oeste y al norte, pero el resto del área de estudio no parece tener ninguna similitud, por lo que puede que esta sea una mera coincidencia. Además esta idea está apoyada por los modelos usados, ya que mientras uno usa trabajadores y vehículos, el otro tiene en cuenta una variable zonal como es la densidad de población en la zona y el tamaño del hogar.



CAPÍTULO 6

Conclusiones



6 CONCLUSIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

6.1 Conclusiones

En esta tesis doctoral se han propuesto un conjunto de modelos avanzados aplicados a la generación de viajes, mediante el uso de modelos de regresión lineal múltiple (MLR), a partir de datos zonales y los datos de los hogares, teniendo en cuenta relaciones espaciales.

Se han calibrado modelos tanto a nivel zonal como a nivel de hogares para el período Punta Mañana. Para el primer caso, se han estimado modelos de producción y atracción de viajes agregados (viajes totales) y desagregados para distintos propósitos de viaje: basados en el hogar motivo trabajo (HBW), basados en el hogar motivo otros (HBO) y no basados en el hogar (NHB). Para el caso de viajes generados a nivel de hogar, se han estimado modelos de producción de viajes para los motivos más comunes: basados en el hogar motivo trabajo (HBW) y basados en el hogar motivo otros (HBO); todos ellos para el caso de estudio de la ciudad de Santander.

Del análisis de los residuos de los modelos RLM clásicos, se detectó la presencia de correlación espacial en muchos de los casos que justificó la utilización de los modelos econométricos espaciales. Por lo tanto, la utilización de los modelos **Spatial Lag** y



Spatial Error mejoró de manera significativa el ajuste estadístico en los modelos de generación de viajes para esos casos.

Por otra parte, los resultados de la calibración demuestran que en términos generales, los modelos **Spatial Lag** y **Spatial Error** calibrados son más adecuados para representar la generación de viajes a nivel de hogar.

En lo que respecta a los viajes a nivel zonal, la atracción de viajes mostró igualmente una sensible presencia de correlación espacial tanto para los viajes totales como para cada motivo analizado. Esto es debido a que los viajes están muy concentrados en la zona centro y en los polos atractores como el Hospital. Por su parte, los viajes producidos mostraron esta correlación de una manera más significativa, para los viajes desagregados por cada motivo.

Las técnicas de econometría espacial, han permitido mostrar y cuantificar las interrelaciones que tienen los datos en el espacio geográfico, encontrándose que existe auto-correlación espacial en los datos en la ciudad de Santander para la producción viajes a nivel de hogar y zonal, siendo en los viajes del primer tipo más significativa.

6.2 Líneas futuras de investigación

A partir de los resultados y conclusiones obtenidas en esta tesis, se proponen las siguientes líneas futuras de investigación:

- Diferenciar por modo de transporte para analizar si existen diferencias significativas entre la generación de viajes en vehículo privado y otros modos.
- Incluir la accesibilidad como variable zonal independiente. De esta manera y combinando su análisis con el uso de los modos de transporte se podría inferir si la generación de viajes es sensible a las características de los usos del suelo y sistemas de transporte.
- Siguiendo la línea de investigación propuesta en el punto anterior, se propone incluir modelos que consideren estas interacciones y la existencia de la dependencia espacial entre observaciones en los conocidos como modelos LUTI (Land Use Transport Interaction).



-
- Emplear otra tipología de modelos, como los modelos Spatial Durbin que consideran a la vez la existencia de correlación espacial en las variables dependientes y las variables independientes de los modelos.
 - Aplicar estos modelos en otras áreas de estudio para realizar comparativas con tipologías de ciudades diversas.



CAPÍTULO 7

Referencias



7 REFERENCIAS

Aldana, C.M. (2006). Modelación de la generación y atracción de viajes en el valle de Aburrá. Medellín: Facultad Nacional de Minas, Universidad Nacional de Colombia Sede Medellín.

Allen, R. L. (1976). Method for comparing fish growth curves. N. Z. J. Mar. Freshwater Res., 10: 687 - 692

Anselin, L. (1980), "Estimation methods for spatial autoregressive structures". Regional Science Dissertation and Monograph Series (Ithaca, NY).

Anselin, L. (1988), "Spatial econometrics: methods and models". Kluwer Academic Publishers.

Anselin, L. (1988a), "*Spatial econometrics: methods and models*". Kluwer Academic Publishers.

Anselin, L. (1988b), "Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity". Geographical Analysis, 20; pp. 1-17.

Anselin, L. (1992). Spatial dependence and spatial heterogeneity: model specifications issues in the spatial expansion paradigm. In E. Casetti, & J.P. Jones III (Eds.) Applications of the expansion method (pp. 334-354). London: Rudledge.

Anselin, L. (1992a), "Discrete space autoregressive models". En "GIS and Environmental Modeling", M.F. Goodchild, B. Parks, L.T. Steyaert (Eds.), Oxford: Oxford University Press.



Anselin, L. (1995), "Local Indicators of Spatial Association-LISA". *Geographical Analysis*, vol. 27(2); pp. 93-115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M. Fischer, H. Scholten, & D. Unwin (Eds.), *Spatial analytical perspectives on GIS* (pp. 111-125). London: Taylor & Francis.

Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment, *Proceedings of Conference on Geocomputation 1998*. Bristol, United Kingdom.

Anselin, L. (1999a), "The future of spatial analysis in the social sciences". *Geographic Information Sciences*, 5 (2); pp. 67-76.

Anselin, L. (2001a), "Spatial econometrics". En "A companion to theoretical econometrics", ed. Baltagi, Oxford: Basil Blackwell; pp. 310-330.

Anselin, L. (2001b), "Spatial effects in econometric practice in environmental and resource economics". *American Journal of Agricultural Economics*, 83 (3), pp. 705-710.

Anselin, L. (2001c) "Rao's score test in spatial econometrics". *Journal of Statistical Planning and Inference* (próxima publicación).

Anselin, L. (2001d), "Spatial econometrics". En "A companion to theoretical econometrics", ed. Baltagi, Oxford: Basil Blackwell; pp. 310-330.

Anselin, L. (2002) "Under the hood. Issues in the specification and interpretation of spatial regression models". *Agricultural Economics* 27; pp. 247-267.

Anselin, L. (2003). *Geoda 0.9. User's guide*. Urbana-Champaign, IL: Department of Agricultural and Consumer Economics - University Of Illinois.

Anselin, L. y S. Bao (1997), "Exploratory Spatial Data Analysis". En "Recent developments in spatial analysis" (Eds. Fischer y Getis), Springer-Verlag, Berlín; pp. 35-59.

Anselin, L., & Bera, A. (1988). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah, & D.E.A. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237-289). New York: Marcel Dekker.

Anselin, L. y A. Bera (1998), "Spatial dependence in linear regression models with an introduction to spatial econometrics". *Handbook of applied economic statistics*, eds. A. Ullah y D. Giles, New York: Marcel Dekker, Inc.; pp. 237-289.

Anselin, L. y S. Rey (1991), "Properties of tests for spatial dependence in linear regression models". *Geographical Analysis*, 23, 2; pp. 113-131

Anselin, L. Syabri, I., & Kho, Y. (2004). *GeoDa: An introduction to spatial data analysis - Spatial analysis laboratory*. Urbana-Champaign, IL: Department of Agricultural and Consumer Economics - University Of Illinois.

Aznar, A., J. Mur y F. J. Trávez (1996), "Métodos econométricos en el análisis regional". *Actas de la XXII Reunión de Estudios Regionales, de la Asociación Española de Ciencia Regional*. Pamplona/Iruña, 20-22 de noviembre de 1996; pp. 237-264.



Badoe, D. (2007). Forecasting Travel Demand With Alternatively Structured Models Of Trip Frequency. *Transportation Planning And Technology*, 30(5), 455 – 475.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G., & Hawkins, D.F. (2001). Structural covariates of U.S. county homicide rates: incorporation spatial effects. *Criminology* 39(3), 561-590.

Bao, S., L. Anselin, D. Martin & D. Stralberg (2000), “Seamless integration of spatial statistics and GIS: the S-Plus for ArcView and the S+Grassland links”. *Journal of Geographical Systems* 2 (3), pp. 287–306.

Baronio, A., Vianco, A., & Rabanal, C. (2012). Una Introducción a la Econometría Espacial: Dependencia y Heterogeneidad. *Catedra de econometría*.

Bodson, P. y D. Peeters (1975), “Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation. An application to a Belgian labour-demand function”. *Environment and Planning A*, 7; pp. 455-472.

Buzai, G.D., & Baxendale, C.A. (2009) (Buzai and Baxendale 2009). Análisis exploratorio de datos espaciales. *Geografía y sistemas de información geográfica. Geografía y Sistemas de Información Geográfica (GEOSIG)*, Año 1, n° 1, Sección III, 1-11.

Cameron, C. A. & Trivedi, P. K. (1998) *Regression Analysis of Count Data* (New York: Cambridge University Press).

Case, A., H. Rosen y J. Hines (1993), “Budget spillovers and fiscal policy interdependence: evidence from the states”. *Journal of Public Economics*, 52; pp. 285-307.

Cerrato, R. M. (1990). Interpretable statistical test for growth comparisons using parameters in the von Bertalanffy equation. *Can. J. Fish. Aquat.Sci.*, 47: 1416 - 1426

Chasco, C. Y. (2003), “Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales (Yrigoyen 2003)”. *Consejería de Economía e Innovación Tecnológica de la Comunidad de Madrid*.

Chasco, C.Y. (2006). Análisis estadístico de datos geográficos en geo marketing: El programa GeoDa (Yrigoyen 2006). Madrid: Departamento de Economía Aplicada, Universidad Autónoma de Madrid.

Chasco, C. & J. Vicéns (2000), “EU-membership impacts in the Spanish province income convergence: a spatial autocorrelation perspective (Yrigoyen and Otero)”. *Anales de Economía Aplicada XIV Reunión de ASEPELT-España, Oviedo*.

Cliff, A. & J. Ord (1973), “Spatial autocorrelation”. London: Pion.

Cliff, A. & J. Ord (1981), “Spatial processes, models and applications”. London: Pion

Cotrus, A. V., Prashker, J. N. & Shiftan, Y. (2005) Spatial And Temporal Transferability Of Trip Generation Demand Models In Israel, *Journal Of Transportation And Statistics* , 8(1), Pp. 37-56.



- Cressie, N. (1993), "Statistics for spatial data..Revised edition". New York: Wiley.
- Daly, A. J. (1997) Improved Methods For Trip Generation, In: Proceedings Of The 25th European Transport Conference, Vol. P415, Pp. 207-222 (London: Ptrc).
- Florax, R. y S. Rey (1995), "The Impacts Of Misspecified Spatial Interaction In Linear Regression Models". En "New directions in spatial econometrics". Ed. Springer, pp. 111-135.
- Galucci, V. F. & Quinn, T. J. (1979). Reparameterizing, fitting and testing a simple growth model. *Trans. Am. Stat. Assoc.*, 108: 14 - 25
- Getis, A. & J. Ord (1992), "The analysis of spatial association by use of distance statistics". *Geographical Analysis*, 24; pp. 189-206.
- Haining, R., S. Wise y P. Signoretta (2000), "Providing scientific visualization for spatial data analysis: Criteria and an assessment of SAGE". *Journal of Geographical Systems*, 2; pp.121-140.
- Hansen, J. E., Cordo, H. D. & Hernandez, D. R. (1993). COMPALLEN. Un programa BASIC para la comparación de curvas de crecimiento del tipo von Bertalanffy. *Frente Marítimo*, 14 (A): 41 - 48
- Isard, W. (1969), "*General Theory*". Cambridge: MIT Press.
- Kappenman, R. F. (1981). A method for growth curve comparisons. *U. S. Fish. Bull.*, 79 (1): 95 - 101
- Kimura, D. K. (1980). Likelihood methods for the von Bertalanffy growth curve. *Fishery Bulletin*, 77, 4: 765 - 776
- Kumar, A. & Levinson, D. M. (1994). Specifying, Estimating And Validating A New Trip Generation Model: Case Study In Montgomery County, Maryland. *Transportation Research Record*, 107-113.
- "LA CAIXA" (2001), "Anuario Económico de España 2001". Obra elaborada por el Instituto; L. R. Klein (Universidad Autónoma de Madrid). Barcelona.
- LeSage, J. (1995), "A multiprocess mixture model to estimate space-time dimensions of weekly pricing of certificates of deposit". En "New directions in spatial econometrics", eds. L. Anselin y R.J. Florax, Springer-Verlag, Berlín; pp. 359-416.
- LeSage, J. (1999). "Spatial econometrics". Department of Economics, University of Toledo.
- LeSage, J. (2000). "Bayesian estimation of limited dependent variable spatial autoregressive models". *Geographical Analysis*, 32.
- Long, J. S. (1997). *Regression Models For Categorical And Limited Dependent Variables* (Thousand Oaks, Ca: Sage Publications).
- Maddala, G. S. (1983) *Limited Dependent And Qualitative Variables In Econometrics* (New York: Cambridge University Press).



Mapinfo (1995), "MapInfo Professional. Guía del usuario". MapInfo Corporation, Nueva York.

MIDEPLAN – FDC (2010). Análisis y formulación de nuevos modelos de generación y atracción de viajes. Santiago, Chile (Videla, Bravo et al. 2010): Ministerio de Planificación y Cooperación.

Moreno, R., & Vayá, E. (2000). Técnicas econométricas para el tratamiento de datos espaciales (Serrano and Valcarce 2000): La econometría espacial. Barcelona: Edicions Universitat de Barcelona.

Mur, J. (1990), "Correlación espacial: un estudio de Monte Carlo (Mur 1990)". Tesis doctoral. Departamento de Análisis Económico, Universidad de Zaragoza.

Mur, J. (1999a), "Heterocedasticidad y autocorrelación espacial: una propuesta de series espaciales (Mur Lacambra 1999)". Estadística Española, vol. 41, nº 144; pp. 79-104.

Mur, J. (1999b), "Testing for spatial autocorrelation: moving average versus autorregressive processes (Mur 1999)". Environment and Planning A, 31(8), pp. 1371-1382.

Mur, J. (2000), "Scale elements in spatial autocorrelation tests". Actas del 40º Congreso de la ERSA (ASOCIACIÓN EUROPEA DE CIENCIA REGIONAL), Barcelona.

Ord, J.K. & A. Getis (1995), "Local spatial autocorrelation statistics: distributional issues and an application". Geographical Analysis, 27.4; pp. 286-306.

Ortuzar, J. D. (2000). Modelos de demanda de transporte. Santiago, Chile: Ediciones Alfaomega.

Ortuzar, J. D., & Willumsen, L.G. (2001). Modelling transport. Chichester, United Kingdom: John Wiley & Sons.

Ortuzar, J. D., & Willumsen, L.G. (2011). Modelos de Transporte traducción de la tercera edición inglesa, Ángel Ibeas Portilla y Luigi dell'Olio. Santander, Ediciones de la Universidad de Cantabria.

Paelinck, J.H.P. & L.H. Klaasen (1979), "Spatial econometrics". Saxon House.

Páez A, Potoglou D, Scott Dm, Kanaroglou Ps, Newbold Kb (2005) The Spatial Analysis Of Trip Generation: A Case Study Of The Hamilton Cma, 10 Th Conference On Computers In Urban Planning And Urban Management, London, Uk, June 29–July 1.

Pauly, D. & Munro, J. L. (1984). Once more on growth comparison in fish and invertebrates. Fish-byte, 2 (1): 21.

Rao, C. (1958). Some statistical methods for comparisons of growth curves. Bionetrics, 14: 1 - 17

Serrano, R. M. and E. V. Valcarce (2000). Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial, Edicions Universitat Barcelona.



Stetzer, F. (1982), "Specifying weights in spatial forecasting models: the results of some experiments". *Environmental and Planning A*, 14; pp. 571-584.

Tobler, W. (1979), "Cellular Geography". En "Philosophy in Geography", ed. S. Gale y G. Olsson, Dordrecht: Reidel; pp. 379-86.

Tukey, J.W. (1977), "Exploratory Data Analysis". Reading: Addison-Wesley.

Vayá, E. y J. Suriñach (1996), "Contrastes de autocorrelación espacial: una aplicación al ámbito de las provincias españolas". *Actas de la X Reunión Asepelt-España*, com. B-3-9. Ed. Junta de Comunidades de Castilla-La Mancha. Albacete.







8 ANEXOS

En este anexo, se detallan los resultados de todos los casos analizados en el capítulo 5. En concreto se detalla, para cada caso, las estimaciones del mejor modelo, y del modelo sin considerar la existencia de correlación espacial, que en dicho capítulo se ha denominado como “modelo básico”.



8.1 MODELOS DE PRODUCCIÓN DE VIAJES TOTALES (ESTIMACIÓN POR ZONAS)

• LOS RESULTADOS DEL MEJOR MODELO 2

RESULTADO DEL MODELO BASICO – 2

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```

Data set           : Zonas_Alex_3
Dependent variable : PROD_TTRIP   Number of observations: 86
Mean dependent var : 4912.35    Number of variables   : 3
S.D. dependent var : 4095.9      Degrees of Freedom    : 83

R-squared          : 0.392417    F-statistic           : 26.8034
Adjusted R-squared : 0.377776    Prob(F-statistic)    : 1.04626e-009
Sum squared residual: 8.76602e+008  Log likelihood        : -815.929
Sigma-square       : 1.05615e+007 Akaike info criterion : 1637.86
S.E. of regression : 3249.84      Schwarz criterion     : 1645.22
Sigma-square ML    : 1.0193e+007
S.E of regression ML: 3192.65
  
```

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	1002.141	695.653	1.440576	0.15347
Hogares	3.048123	0.9350348	3.259903	0.00162
Den_POP_He	8.049349	1.928784	4.173277	0.00007

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 4.515716

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	25.9166	0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	7.1810	0.02758
Koenker-Bassett test	2	4.1437	0.12595

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : Zonas_Alex_3.gal
(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	-0.0380	-0.3167	0.75147
Lagrange Multiplier (lag)	1	0.0062	0.93741
Robust LM (lag)	1	1.0294	0.31030
Lagrange Multiplier (error)	1	0.3304	0.56542
Robust LM (error)	1	1.3536	0.24465
Lagrange Multiplier (SARMA)	2	1.3598	0.50667

Modelo Básico_2



8.2 MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR MOTIVO TRABAJO (ESTIMACIÓN POR ZONAS)

- LOS RESULTADOS DEL MEJOR MODELO 2

RESULTADO DEL MODELO BASICO – 2

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : Zonas_Alex_3
Dependent variable : PROD_HBW   Number of observations: 86
Mean dependent var : 1278.49   Number of variables   : 2
S.D. dependent var : 1304.08   Degrees of Freedom    : 84

R-squared      : 0.109383   F-statistic           : 10.3167
Adjusted R-squared : 0.098781   Prob(F-statistic)    : 0.00186948
Sum squared residual:1.30255e+008   Log likelihood       : -733.947
Sigma-square   :1.55066e+006   Akaike info criterion : 1471.89
S.E. of regression : 1245.25   Schwarz criterion    : 1476.8
Sigma-square ML : 1.5146e+006
S.E of regression ML: 1230.69
```

variable	Coefficient	std.Error	t-statistic	Probability
CONSTANT	434.6179	295.0541	1.473011	0.14448
sq_habitab	0.01385593	0.004313859	3.211958	0.00187

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 4.153899
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 198.8748 0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 1 2.5379 0.11114
Koenker-Bassett test 1 0.6260 0.42882

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Resultado Queen_1.gal
(row-standardized weights)
TEST MI/DF VALUE PROB
Moran's I (error) 0.0001 0.2130 0.83132
Lagrange Multiplier (lag) 1 0.1770 0.67398
Robust LM (lag) 1 2.2846 0.13066
Lagrange Multiplier (error) 1 0.0000 0.99834
```

Modelo Básico – 2



QUEEN – 2

```
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set      : Zonas_Alex_3
Spatial weight : Resultado Queen_2.gal
Dependent Variable : PROD_HBW      Number of Observations: 86
Mean dependent var : 1278.489173   Number of Variables : 2
S.D. dependent var : 1304.075915   Degrees of Freedom : 84
Lag coeff. (Lambda) : -0.603091

R-squared      : 0.183716   R-squared (BUSE) : -
Sq. Correlation : -         Log likelihood : -731.607189
Sigma-square   : 1.38818e+006 Akaike info criterion : 1467.21
S.E of regression : 1178.21   Schwarz criterion : 1472.12
```

variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	281.23	243.6212	1.154374	0.24835
sq_habitab	0.01662529	0.003786292	4.390917	0.00001
LAMBDA	-0.6030915	0.2430124	-2.481731	0.01307

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	1	3.4854	0.06191

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Resultado Queen_2.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	4.6797	0.03052

Modelo_2_Queen_2_Spacial Error



8.3 MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR OTROS MOTIVOS (ESTIMACIÓN POR ZONAS)

- LOS RESULTADOS DEL MEJOR MODELO 1

RESULTADO DEL MODELO BASICO – 1

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : Zonas_Alex_3
Dependent variable : PROD_HBO   Number of observations: 86
Mean dependent var : 3405.79   Number of variables : 3
S.D. dependent var : 3041.12   Degrees of Freedom  : 83

R-squared      : 0.476574   F-statistic      : 37.7853
Adjusted R-squared : 0.463961   Prob(F-statistic) : 2.15029e-012
Sum squared residual: 4.16313e+008   Log likelihood    : -783.911
Sigma-square    : 5.01582e+006   Akaike info criterion : 1573.82
S.E. of regression : 2239.6   Schwarz criterion : 1581.18
Sigma-square ML : 4.84085e+006
S.E of regression ML: 2200.19
```

Variable	Coefficient	Std. Error	t-statistic	Probability
CONSTANT	24.2449	531.8	0.04559026	0.96375
sq_habitab	0.03082226	0.008701654	3.542115	0.00065
Den_POP_He	7.150961	1.306238	5.47447	0.00000

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 5.050193
(Extreme Multicollinearity)
```

```
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 26.7718 0.00000
```

```
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 2 0.8292 0.66059
Koenker-Bassett test 2 0.4595 0.79474
```

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Resultado Queen_1.gal
(row-standardized weights)
TEST MI/DF VALUE PROB
Moran's I (error) -0.0683 -0.8014 0.42289
Lagrange Multiplier (lag) 1 0.0311 0.86004
Robust LM (lag) 1 1.0641 0.30229
Lagrange Multiplier (error) 1 1.0664 0.30176
Robust LM (error) 1 2.0994 0.14736
Lagrange Multiplier (SARMA) 2 2.1305 0.34464
```

Modelo Básico – 1



QUEEN – 3

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
 Data set : Zonas_Alex_3
 Spatial weight : Resultado_Queen_3.gal
 Dependent Variable : PROD_HBO Number of observations: 86
 Mean dependent var : 3405.79 Number of Variables : 4
 S.D. dependent var : 3041.12 Degrees of Freedom : 82
 Lag coeff. (Rho) : -0.510086
 R-squared : 0.511713 Log likelihood : -781.659
 Sq. Correlation : - Akaike info criterion : 1571.32
 Sigma-square : 4.51587e+006 Schwarz criterion : 1581.14
 S.E of regression : 2125.06

variable	Coefficient	Std.Error	z-value	Probability
W_PROD_HBO	-0.5100863	0.2251071	-2.265971	0.02345
CONSTANT	1908.619	946.6814	2.016116	0.04379
sq_habitab	0.02834997	0.008273448	3.426621	0.00061
Den_POP_He	7.340874	1.239522	5.922342	0.00000

REGRESSION DIAGNOSTICS
 DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	0.4146	0.81278

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	4.5023	0.03385

Modelo_1_Queen_3_Spacial Lag



8.4 MODELOS DE VIAJES PRODUCIDOS NO BASADOS EN EL HOGAR (ESTIMACIÓN POR ZONAS)

- LOS RESULTADOS DEL MEJOR MODELO 2

RESULTADO DEL MODELO BASICO – 2

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : Zonas_Alex_3
Dependent variable : PROD_NHB      Number of observations: 86
Mean dependent var : 365.335      Number of variables : 2
S.D. dependent var : 355.51       Degrees of Freedom : 84

R-squared      : 0.477703      F-statistic      : 76.8282
Adjusted R-squared : 0.471486      Prob(F-statistic) : 1.76235e-013
Sum squared residual: 5.677e+006      Log likelihood   : -599.225
Sigma-square   : 67583.3      Akaike info criterion : 1202.45
S.E. of regression : 259.968      Schwarz criterion : 1207.36
Sigma-square ML : 66011.6
S.E of regression ML: 256.927
```

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	107.7494	40.61367	2.653032	0.00954
Locales_Co	1.991761	0.2272359	8.765167	0.00000

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 2.497089
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 3.7421 0.15396

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 1 21.3904 0.00000
Koenker-Bassett test 1 15.5335 0.00008

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Resultado Queen_1.gal
(row-standardized weights)
TEST MI/DF VALUE PROB
Moran's I (error) -0.1584 -2.2366 0.02531
Lagrange Multiplier (lag) 1 0.3321 0.56443
Robust LM (lag) 1 4.6724 0.03065
Lagrange Multiplier (error) 1 5.7336 0.01664
Robust LM (error) 1 10.0738 0.00150
Lagrange Multiplier (SARMA) 2 10.4059 0.00550
```

Modelo Básico – 2



ROOK - 2

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
 Data set : Zonas_Alex_3
 Spatial weight : Resultado Rook 2.gal
 Dependent Variable : PROD_NHB Number of Observations: 86
 Mean dependent var : 365.334657 Number of Variables : 2
 S.D. dependent var : 355.509751 Degrees of Freedom : 84
 Lag coeff. (Lambda) : -0.477983
 R-squared : 0.535989 R-squared (BUSE) : -
 Sq. Correlation : - Log likelihood : -595.931684
 Sigma-square : 58645 Akaike info criterion : 1195.86
 S.E of regression : 242.167 Schwarz criterion : 1200.77

Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	89.86788	28.84859	3.115157	0.00184
Locales_Co	2.124706	0.173336	12.25774	0.00000
LAMBDA	-0.477983	0.1791196	-2.668512	0.00762

REGRESSION DIAGNOSTICS
 DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	1	18.3964	0.00002

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Resultado Rook 2.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	6.5864	0.01028

Modelo_2_Rook_2_Spatial Error



8.5 MODELOS DE VIAJES ATRAÍDOS TOTALES (ESTIMACIÓN POR ZONAS)

- LOS RESULTADOS DEL MEJOR MODELO 1

RESULTADO DEL MODELO BASICO – 1

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : Zonas_Alex_3
Dependent Variable : ATTRAC_TTR      Number of Observations: 86
Mean dependent var : 4912.35        Number of Variables   : 3
S.D. dependent var : 4001.2         Degrees of Freedom    : 83

R-squared      : 0.500960      F-statistic          : 41.6597
Adjusted R-squared : 0.488935      Prob(F-statistic)   : 2.96898e-013
Sum squared residual: 6.87091e+008    Log likelihood       : -805.455
Sigma-square    : 8.27821e+006    Akaike info criterion : 1616.91
S.E. of regression : 2877.19      Schwarz criterion    : 1624.27
Sigma-square ML : 7.98943e+006
S.E of regression ML: 2826.56
```

variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	2065.525	454.2354	4.547257	0.00002
FACILITY	6295.674	1326.686	4.745412	0.00001
Locales_Co	19.18257	2.51654	7.622596	0.00000

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 2.585704
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 25.4495 0.00000
```

```
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 2 3.7041 0.15692
Koenker-Bassett test 2 2.0283 0.36272
```

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Resultado Queen_1.gal
(row-standardized weights)
TEST MI/DF VALUE PROB
Moran's I (error) -0.0934 -1.2150 0.22436
Lagrange Multiplier (lag) 1 4.1130 0.04255
Robust LM (lag) 1 2.2217 0.13608
Lagrange Multiplier (error) 1 1.9950 0.15781
Robust LM (error) 1 0.1037 0.74739
Lagrange Multiplier (SARMA) 2 4.2168 0.12143
```

Modelo Básico – 1



QUEEN - 1

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
 Data set : Zonas_Alex_3
 Spatial weight : Resultado Queen_1.gal
 Dependent variable : ATTRAC_TTR Number of observations: 86
 Mean dependent var : 4912.35 Number of variables : 4
 S.D. dependent var : 4001.2 Degrees of Freedom : 82
 Lag coeff. (Rho) : -0.317426
 R-squared : 0.535638 Log likelihood : -803.102
 Sq. Correlation : - Akaike info criterion : 1614.2
 Sigma-square : 7.43426e+006 Schwarz criterion : 1624.02
 S.E of regression : 2726.58

Variable	Coefficient	Std. Error	z-value	Probability
W_ATTRAC_TTR	-0.3174256	0.1449709	-2.189581	0.02855
CONSTANT	3367.715	714.7904	4.711472	0.00000
FACILITY	6439.533	1259.753	5.111743	0.00000
Locales_Co	21.62667	2.67552	8.083165	0.00000

REGRESSION DIAGNOSTICS
 DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	2.6999	0.25925

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	4.7054	0.03007

Modelo_1_Queen_1_Spacial Lag



8.6 MODELOS DE VIAJES ATRAÍDOS BASADOS EN EL HOGAR POR EL MOTIVO TRABAJO (ESTIMACIÓN POR ZONAS)

- LOS RESULTADOS DEL MEJOR MODELO 2

RESULTADO DEL MODELO BASICO – 2

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set          : Zonas_Alex_3
Dependent variable : ATTRAC_HBW  Number of Observations: 86
Mean dependent var : 1278.49    Number of Variables : 3
S.D. dependent var : 1290.72    Degrees of Freedom : 83

R-squared         : 0.500887    F-statistic          : 41.6475
Adjusted R-squared : 0.488860    Prob(F-statistic)   : 2.98717e-013
Sum squared residual: 7.15092e+007  Log likelihood       : -708.161
Sigma-square      : 861556     Akaike info criterion : 1422.32
S.E. of regression : 928.201    Schwarz criterion    : 1429.69
Sigma-square ML   : 831502
S.E. of regression ML: 911.867
```

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	408.8615	146.5395	2.790111	0.00653
FACILITY	2430.505	427.9982	5.678773	0.00000
Locales_Co	5.63168	0.8118532	6.93682	0.00000

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 2.585704
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 41.8235 0.00000
```

```
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 2 16.6176 0.00025
Koenker-Bassett test 2 7.6117 0.02224
```

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Resultado Queen_1.gal
(row-standardized weights)
TEST MI/DF VALUE PROB
Moran's I (error) 0.0608 1.2125 0.22534
Lagrange Multiplier (lag) 1 0.5621 0.45343
Robust LM (lag) 1 0.0005 0.98216
Lagrange Multiplier (error) 1 0.8445 0.35811
Robust LM (error) 1 0.2830 0.59477
Lagrange Multiplier (SARMA) 2 0.8450 0.65540
```

Modelo Básico – 2



8.7 MODELOS DE VIAJES ATRAÍDOS BASADOS EN EL HOGAR POR OTROS MOTIVOS (ESTIMACIÓN POR ZONAS)

• LOS RESULTADOS DEL MEJOR MODELO 1

RESULTADO DEL MODELO BASICO – 1

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : Zonas_Alex_3
Dependent variable : ATTAC_HBO      Number of Observations: 86
Mean dependent var : 3405.79      Number of Variables : 3
S.D. dependent var : 2847.91      Degrees of Freedom : 83

R-squared      : 0.370154      F-statistic      : 24.3891
Adjusted R-squared : 0.354977      Prob(F-statistic) : 4.65839e-009
Sum squared residual : 4.39324e+008      Log likelihood : -786.224
Sigma-square    : 5.29306e+006      Akaike info criterion : 1578.45
S.E. of regression : 2300.67      Schwarz criterion : 1585.81
Sigma-square ML : 5.10842e+006
S.E of regression ML: 2260.18
```

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	1637.778	363.2172	4.509087	0.00002
FACILITY	3567.856	1060.849	3.363208	0.00117
Locales_Co	12.06709	2.012284	5.996714	0.00000

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 2.585704
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 69.7330 0.00000
```

```
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 2 2.8438 0.24125
Koenker-Bassett test 2 1.0904 0.57973
```

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Resultado Queen_1.gal
(row-standardized weights)
TEST MI/DF VALUE PROB
Moran's I (error) -0.0593 -0.6779 0.49786
Lagrange Multiplier (lag) 1 2.9511 0.08582
Robust LM (lag) 1 3.4507 0.06323
Lagrange Multiplier (error) 1 0.8038 0.36996
Robust LM (error) 1 1.3034 0.25360
Lagrange Multiplier (SARMA) 2 4.2545 0.11917
```

Modelo Básico – 1



8.8 MODELOS DE VIAJES ATRAÍDOS NO BASADOS EN EL HOGAR (ESTIMACIÓN POR ZONAS)

• LOS RESULTADOS DEL MEJOR MODELO 1

RESULTADO DEL MODELO BASICO – 1

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
 Data set : Zonas_Alex_3
 Dependent variable : ATTRAC_NHB Number of Observations: 86
 Mean dependent var : 365.335 Number of Variables : 3
 S.D. dependent var : 299.239 Degrees of Freedom : 83

R-squared : 0.428221 F-statistic : 31.0805
 Adjusted R-squared : 0.414443 Prob(F-statistic) : 8.4133e-011
 Sum squared residual: 4.40314e+006 Log likelihood : -588.298
 Sigma-square : 53049.8 Akaike info criterion : 1182.6
 S.E. of regression : 230.326 Schwarz criterion : 1189.96
 Sigma-square ML : 51199.3
 S.E of regression ML: 226.273

variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	158.5406	36.3626	4.359989	0.00004
FACILITY	292.2419	106.2043	2.751695	0.00728
Locales_Co	1.467641	0.2014549	7.285211	0.00000

REGRESSION DIAGNOSTICS
 MULTICOLLINEARITY CONDITION NUMBER 2.585704
 TEST ON NORMALITY OF ERRORS
 TEST DF VALUE PROB
 Jarque-Bera 2 12.2080 0.00223

DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS
 TEST DF VALUE PROB
 Breusch-Pagan test 2 9.6659 0.00796
 Koenker-Basset test 2 6.3137 0.04256

DIAGNOSTICS FOR SPATIAL DEPENDENCE
 FOR WEIGHT MATRIX : Resultado Queen_1.gal
 (row-standardized weights)
 TEST MI/DF VALUE PROB
 Moran's I (error) -0.1021 -1.3504 0.17688
 Lagrange Multiplier (lag) 1 0.5809 0.44594
 Robust LM (lag) 1 0.8307 0.36207
 Lagrange Multiplier (error) 1 2.3793 0.12295
 Robust LM (error) 1 2.6290 0.10493
 Lagrange Multiplier (SARMA) 2 3.2100 0.20089

Modelo Básico – 1



8.9 MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR EL MOTIVO TRABAJO (ESTIMACIÓN POR HOGARES)

- **LOS RESULTADOS DEL MEJOR MODELO 1**

RESULTADO DEL MODELO BASICO – 1

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : Domiciliaria_2010
Dependent variable : HBW      Number of observations: 817
Mean dependent var : 1.98164  Number of variables   : 3
S.D. dependent var : 2.17044  Degrees of Freedom   : 814

R-squared      : 0.632896  F-statistic          : 701.677
Adjusted R-squared : 0.631994  Prob(F-statistic)   : 0
Sum squared residual: 1412.88  Log likelihood       : -1383.03
Sigma-square    : 1.73573  Akaike info criterion : 2772.06
S.E. of regression : 1.31747  Schwarz criterion    : 2786.17
Sigma-square ML  : 1.72935
S.E. of regression ML: 1.31505
```

variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	0.009886357	0.07865608	0.1256909	0.89995
WORKERS	1.933921	0.05978094	32.35012	0.00000
VEH	0.1906257	0.06697269	2.84632	0.00453

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 3.419016
TEST ON NORMALITY OF ERRORS
TEST      DF      VALUE      PROB
Jarque-Bera      2      200.8975      0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST      DF      VALUE      PROB
Breusch-Pagan test      2      363.2763      0.00000
Koenker-Basset test      2      164.2405      0.00000
```

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Resultado Threshold Distance (Maximo).gwt
(row-standardized weights)
TEST      MI/DF      VALUE      PROB
Moran's I (error)      -0.0012      -0.0000      1.00000
Lagrange Multiplier (lag)      1      0.5006      0.47923
Robust LM (lag)      1      0.0000      0.99999
Lagrange Multiplier (error)      1      0.5006      0.47923
Robust LM (error)      1      0.0000      0.99999
Lagrange Multiplier (SARMA)      2      0.5006      0.77856
```

Modelo Básico – 1



QUEEN – 2

```

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set          : Poligonos
Spatial weight    : Resultado Queen_2.gal
Dependent Variable : HBW      Number of observations: 817
Mean dependent var : 1.981640 Number of Variables : 3
S.D. dependent var : 2.170438 Degrees of Freedom : 814
Lag coeff. (Lambda) : 0.387033

R-squared          : 0.645080 R-squared (BUSE) : -
Sq. Correlation    : -      Log likelihood : -1372.711341
Sigma-square       : 1.67196 Akaike info criterion : 2751.42
S.E of regression  : 1.29304 Schwarz criterion : 2765.54
    
```

Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	0.01675554	0.09683272	0.1730359	0.86262
WORKERS	1.936923	0.0587735	32.95572	0.00000
VEH	0.1848724	0.06613258	2.795481	0.00518
LAMBDA	0.3870335	0.09001913	4.299458	0.00002

```

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST
Breusch-Pagan test          DF      VALUE      PROB
                             2      360.5060   0.00000
    
```

```

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Resultado Queen_2.gal
TEST
Likelihood Ratio Test      DF      VALUE      PROB
                             1      20.6334   0.00001
    
```

Modelo_1_Queen_2_Spacial Error



• **LOS RESULTADOS DEL MEJOR MODELO 2**

RESULTADO DEL MODELO BASICO – 2

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```

Data set           : Domiciliaria_2010
Dependent variable : HBW      Number of observations: 817
Mean dependent var : 1.98164  Number of variables   : 2
S.D. dependent var : 2.17044  Degrees of Freedom    : 815

R-squared          : 0.629242  F-statistic           : 1383.2
Adjusted R-squared: 0.628787  Prob(F-statistic)    : 0
Sum squared residual: 1426.95  Log likelihood        : -1387.07
Sigma-square       : 1.75085  Akaike info criterion: 2778.15
S.E. of regression : 1.3232   Schwarz criterion     : 2787.56
Sigma-square ML    : 1.74657
S.E of regression ML: 1.32158
  
```

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.1232785	0.06811593	1.809834	0.07069
WORKERS	2.008309	0.05399929	37.1914	0.00000

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 2.550793

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	195.1332	0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	1	332.5634	0.00000
Koenker-Bassett test	1	151.3739	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : Resultado Threshold Distance (Maximo).gwt
(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	-0.0012	-0.0000	1.00000
Lagrange Multiplier (lag)	1	0.5006	0.47923
Robust LM (lag)	1	-0.0000	-1.00000
Lagrange Multiplier (error)	1	0.5006	0.47923
Robust LM (error)	1	-0.0000	-1.00000
Lagrange Multiplier (SARMA)	2	0.5006	0.77856

Modelo Básico – 2



QUEEN – 2

```
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set          : Poligonos
Spatial weight    : Resultado Queen_2.gal
Dependent Variable : HBW      Number of observations: 817
Mean dependent var : 1.981640 Number of Variables : 2
S.D. dependent var : 2.170438 Degrees of Freedom : 815
Lag coeff. (Lambda) : 0.389028

R-squared          : 0.641718 R-squared (BUSE) : -
Sq. Correlation    : -      Log likelihood : -1376.599835
Sigma-square       : 1.68779 Akaike info criterion : 2757.2
S.E of regression  : 1.29915 Schwarz criterion : 2766.61
```

Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	0.1257203	0.08926615	1.408376	0.15902
WORKERS	2.00997	0.05289303	38.00066	0.00000
LAMBDA	0.3890283	0.08990309	4.327196	0.00002

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	1	336.7437	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Resultado Queen_2.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	20.9476	0.00000

Modelo_2_Queen_2_Spatial Error



8.10 MODELOS DE VIAJES PRODUCIDOS BASADOS EN EL HOGAR POR OTROS MOTIVOS (ESTIMACIÓN POR HOGARES)

- **LOS RESULTADOS DEL MEJOR MODELO 1**

RESULTADO DEL MODELO BASICO – 1

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
 Data set : Domiciliaria_2010
 Dependent Variable : HBO Number of Observations: 817
 Mean dependent var : 4.34149 Number of Variables : 3
 S.D. dependent var : 3.26131 Degrees of Freedom : 814

R-squared : 0.304992 F-statistic : 178.605
 Adjusted R-squared : 0.303284 Prob(F-statistic) : 0
 Sum squared residual: 6039.43 Log likelihood : -1976.45
 Sigma-square : 7.41944 Akaike info criterion : 3958.89
 S.E. of regression : 2.72387 Schwarz criterion : 3973.01
 Sigma-square ML : 7.3922
 S.E of regression ML: 2.71886

variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	-0.1222395	0.2547501	-0.479841	0.63148
SIZE	1.801677	0.1004799	17.93073	0.00000
DEN_EMP_HA	0.004564291	0.0008669979	5.264478	0.00000

REGRESSION DIAGNOSTICS
 MULTICOLLINEARITY CONDITION NUMBER 5.629384
 TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	61.6770	0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	113.5965	0.00000
Koenker-Bassett test	2	78.9246	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
 FOR WEIGHT MATRIX : Resultado Threshold Distance (Maximo).gwt
 (row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	-0.0012	-0.0000	1.00000
Lagrange Multiplier (lag)	1	0.5006	0.47923
Robust LM (lag)	1	0.0000	0.99999
Lagrange Multiplier (error)	1	0.5006	0.47923
Robust LM (error)	1	0.0000	0.99999
Lagrange Multiplier (SARMA)	2	0.5006	0.77856

Modelo Básico – 1



QUEEN – 1

```

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set      : Poligonos
Spatial weight : Resultado Queen_1.gal
Dependent variable : HBO      Number of Observations: 817
Mean dependent var : 4.341493 Number of Variables : 3
S.D. dependent var : 3.261309 Degrees of Freedom : 814
Lag coeff. (Lambda) : 0.319116

R-squared      : 0.337562 R-squared (BUSE) : -
Sq. Correlation : -      Log likelihood : -1962.296436
Sigma-square   : 7.04578 Akaike info criterion : 3930.59
S.E of regression : 2.65439 Schwarz criterion : 3944.71
    
```

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-0.009791151	0.2748845	-0.03561915	0.97159
SIZE	1.775891	0.09819152	18.086	0.00000
DEN_EMP_HA	0.00367831	0.001093444	3.363969	0.00077
LAMBDA	0.3191155	0.05862612	5.443232	0.00000

REGRESSION DIAGNOSTICS
 DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	122.2232	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Resultado Queen_1.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	28.3002	0.00000

Modelo_1_Queen_1_Spacial Error



• LOS RESULTADOS DEL MEJOR MODELO 2

RESULTADO DEL MODELO BASICO – 2

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
 Data set : Domiciliaria_2010
 Dependent variable : HBO Number of Observations: 817
 Mean dependent var : 4.34149 Number of Variables : 3
 S.D. dependent var : 3.26131 Degrees of Freedom : 814

 R-squared : 0.294823 F-statistic : 170.16
 Adjusted R-squared : 0.293091 Prob(F-statistic) : 0
 Sum squared residual: 6127.79 Log likelihood : -1982.38
 Sigma-square : 7.528 Akaike info criterion : 3970.76
 S.E. of regression : 2.74372 Schwarz criterion : 3984.88
 Sigma-square ML : 7.50036
 S.E of regression ML: 2.73868

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	-0.1248781	0.2643813	-0.4723408	0.63682
SIZE	1.818717	0.1011384	17.98246	0.00000
EMP	0.0005646558	0.0001430679	3.946768	0.00009

REGRESSION DIAGNOSTICS
 MULTICOLLINEARITY CONDITION NUMBER 5.796010
 TEST ON NORMALITY OF ERRORS
 TEST DF VALUE PROB
 Jarque-Bera 2 56.1334 0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS
 TEST DF VALUE PROB
 Breusch-Pagan test 2 113.1785 0.00000
 Koenker-Bassett test 2 81.5070 0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
 FOR WEIGHT MATRIX : Resultado Threshold Distance (Maximo).gwt
 (row-standardized weights)
 TEST MI/DF VALUE PROB
 Moran's I (error) -0.0012 -0.0000 1.00000
 Lagrange Multiplier (lag) 1 0.5006 0.47923
 Robust LM (lag) 1 0.0000 0.99999
 Lagrange Multiplier (error) 1 0.5006 0.47923
 Robust LM (error) 1 0.0000 0.99998
 Lagrange Multiplier (SARMA) 2 0.5006 0.77856

Modelo Básico – 2



QUEEN – 1

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
 Data set : Poligonos
 Spatial weight : Resultado Queen_1.gal
 Dependent variable : HBO Number of Observations: 817
 Mean dependent var : 4.341493 Number of Variables : 3
 S.D. dependent var : 3.261309 Degrees of Freedom : 814
 Lag coeff. (Lambda) : 0.349581
 R-squared : 0.336573 R-squared (BUSE) : -
 Sq. Correlation : - Log likelihood :-1964.086024
 Sigma-square : 7.0563 Akaike info criterion : 3934.17
 S.E of regression : 2.65637 Schwarz criterion : 3948.29

Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	0.0126471	0.2851088	0.04435885	0.96462
SIZE	1.780551	0.09815314	18.14054	0.00000
EMP	0.0004445494	0.0001708557	2.601899	0.00927
LAMBDA	0.3495805	0.05718682	6.112956	0.00000

REGRESSION DIAGNOSTICS
 DIAGNOSTICS FOR HETEROSKEDASTICITY
 RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	124.4132	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Resultado Queen_1.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	36.5881	0.00000

Modelo_2_Queen_1_Spacial Error