

# Modelos avanzados de elección de localización residencial

Alfonso Orro Arcay  
José Benito Pérez López



Grupo de Ferrocarriles y Transportes  
*E.T.S. de Ing. de Caminos, Canales y Puertos*  
UNIVERSIDAD DE A CORUÑA



# Elección de la Localización Residencial (ELR)

- ❑ Problema: ¿dónde vive la gente y cómo elige su localización?
  - ❑ Crítico para los ciudadanos porque el hogar es un elemento muy importante en su vida
    - Generalmente supone una de sus mayores inversiones
    - Es un elemento determinante para su vida social
  - ❑ La elección es realizada por los diferentes agentes del sistema en base a sus características
  - ❑ La localización de la residencia tiene una gran influencia en las características de la vivienda
    - Precio
    - Uso habitual del transporte
    - Tipo de actividad social
    - Planificación del tiempo y las actividades dentro y fuera del hogar (Guo y Bhat, 2007)
- Modelización: **Enfoque empírico desagregado**

# Tipología del submodelo ELR en el contexto LUTI

<i>Modelo</i>	<i>Método</i>	<i>Tratamiento del espacio</i>	<i>Categorización</i>	<i>Escala territorial</i>	<i>Área geográfica</i>
<i>Edmonton</i>	<i>MED</i>	<i>Zonas discretas - Individual sólo en demanda</i>	<i>Categorías</i>	<i>Urbana</i>	<i>Edmonton, Alberta</i>
<i>DRAM</i>	<i>Interacción espacial y MED</i>	<i>Zonas discretas</i>	<i>Categorías</i>	<i>Urbana y metropolitana</i>	<i>USA</i>
<i>DELTA</i>	<i>MED</i>	<i>Zonas discretas</i>	<i>Categorías</i>	<i>Urbana y regional</i>	<i>UK y Nueva Zelanda</i>
<i>MUSSA</i>	<i>MED</i>	<i>Zonas discretas</i>	<i>Categorías</i>	<i>Urbana</i>	<i>Santiago de Chile</i>
<i>Oxford</i>	<i>Econométrico-Regresión</i>	<i>Zonas discretas</i>	<i>Categorías</i>	<i>Urbana y regional</i>	<i>Oxfordshire</i>
<i>TILT</i>	<i>MED</i>	<i>Zonas discretas</i>	<i>Categorías</i>	<i>Regional</i>	<i>Región de Stockhoom, Suecia</i>
<i>UrbanSIM</i>	<i>MED</i>	<i>Zonas discretas, celdas o parcelas</i>	<i>Categorías</i>	<i>Metropolitana</i>	<i>USA y Europa occidental</i>
<i>Oregon2</i>	<i>MED</i>	<i>Celdas</i>	<i>Microsimulación</i>	<i>Regional (estado-USA)</i>	<i>Oregon</i>
<i>ALBATROSS</i>	<i>MED</i>	<i>Zonas discretas</i>	<i>Microsimulación</i>	<i>Nacional</i>	<i>Holanda</i>
<i>RAMBLASS</i>	<i>MED</i>	<i>Zonas discretas</i>	<i>Microsimulación</i>	<i>Urbana y regional</i>	<i>Sur y Oeste de Yorkshire, UK</i>

# Modelo de Elección de la Localización Residencial

- Planteamiento teórico: Elección de un individuo  $n$  entre  $I(26)$  zonas alternativas

$$Y_{ni} = 1 \text{ si } i \text{ es la alternativa elegida y } 0 \text{ en otro caso}$$

- En este contexto de variable dependiente nominal el planteamiento econométrico de regresión lineal genera estimadores sesgados e inconsistentes, con lo que el enfoque estadístico del problema es no lineal a partir del modelo de probabilidad de cada elección (que posteriormente se agregará a toda la población)

$$P_{ni} = P(Y_{ni} = 1), i \in \{1, \dots, I\}$$

- El método se ancla a una teoría económica (a diferencia del modelado de fenómenos físicos que se derivan directamente de la teoría de probabilidad) para permitir una interpretación económica: Maximización de la Utilidad del agente económico

$$P_{ni} = P(U_{ni} \geq U_{nj}, \forall j \neq i) = P(U_{ni} \geq \max_j U_{nj})$$

- Modelos microeconómicos basados en la teoría de la utilidad aleatoria

$$U_{ni} = V_{ni} + \varepsilon_{ni}, i = 1, \dots, I$$

- Modelos de Elección Discreta (DCM) especialmente en modelos de prognosis (Ben-akiva et al., 2010)

- U: Utilidad indirecta U

- V: utilidad sistemática u observada con las siguientes características:

- Enfoque condicional: función de las variables específicas de cada zona alternativa
- Interacciones con atributos individuales para caracterizar al individuo decisor en la parte observada de la utilidad
- Forma funcional lineal en la mayoría de las ocasiones (aunque en otras aplicaciones se han estudiado otras formas funcionales como el modelo Logit Mixto Box-Cox (Orro, A., 2003)

- $\varepsilon$ : error aleatorio

# Modelo Logit de Elección Discreta

- Distribución Valor Extremo de la componente no observada (error aleatorio)

- 0. Multinomial Logit (MNL, McFadden, D., 1974)

$$P_{ni} = \frac{e^{\lambda V_{ni}}}{\sum_j e^{\lambda V_{nji}}}$$

- Como se verá, comúnmente aplicado en ELR

- Limitaciones:

1. No tiene en cuenta correlación entre alternativas ni entre individuos
2. Homogeneidad en los gustos

1. Flexibilización del MNL para incorporar correlación:

- Estructura cerrada → modelos de valor extremo generalizado (GEV),

$$P_{ni} = \frac{x_i \frac{\delta G}{\delta x_i}(x_1, \dots, x_I)}{\lambda G(x_1, \dots, x_I)}$$

$$x_j = e^{\lambda V_j} \quad \forall j \in C$$

- Modelos con simulación

- *Logit mixto (componentes de error)* → Dificultad para este tipo de problemas por el  $n^0$  de alternativas

2. Modelos con heterogeneidad aleatoria en los gustos → modelo logit mixto (coeficientes aleatorios)

# MELR con datos de corte transversal

- Características:
  1. **Correlación espacial entre zonas alternativas**
  2. **Heterogeneidad en los gustos (observada y no observada)**
- Propuesta: Modelos GEV mixtos con correlación espacial
  1. **Núcleo GEV, de estructura cerrada, que recoja la correlación espacial**
    - i. Con una estructura de nidos de zonas predefinida por el analista
      - A. NL: Logit Anidado o Jerárquico
      - B. CNL: Logit de Nidos Cruzados o GNL (Logit Anidado Generalizado)
    - ii. Sin una estructura de nidos predefinida por el analista
      - A. PCL: Logit Combinatorio Generalizado
      - B. SCL: Logit Espacialmente Correlado (basado en la vecindad de zonas)
      - C. GSCL: Logit Espacialmente Correlado Generalizado
      - D. DSCL: basado en la distancia entre zonas
      - E. NPSCL: basado en el perímetro común entre zonas adyacentes
  2. **Sobre el núcleo GEV se define un modelo mixto de coeficientes aleatorios que recoja la heterogeneidad no observada en los gustos**

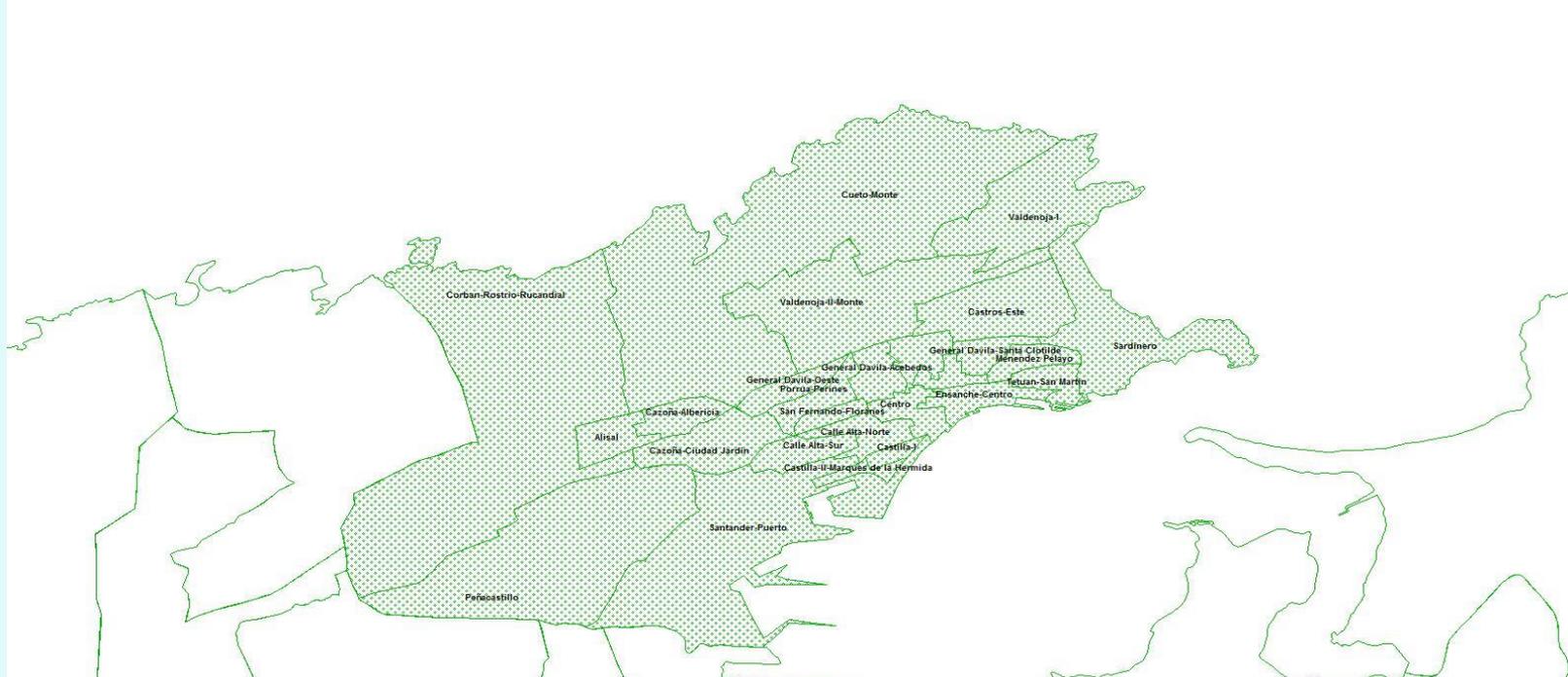
## 0. ML (M-MNL)

- i. Con nidos predefinidos por el analista
  - A. M-NL
  - B. M-CNL

- ii. Sin nidos predefinidos por el analista
  - B. M-SCL
  - C. M-GNL
  - D. M-DSCL
  - E. M-NPSCL

# Aplicación a escala metropolitana: Santander

- Zonificación del área metropolitana de Santander → 26 zonas



## Modelo estructural

- Estudiar los factores que determinan la localización de la vivienda
- Variables y atributos influyentes

## Modelo predictivo

- Estudiar los factores que determinan la localización de la vivienda
- Diseñado a partir del estructural mediante backward

# Modelo estructural TRANSPACE: Ibeas, A., Cordera, R., Dell'Olio, L. y Coppola, P. (2013)

- ❑ Variables de transporte, entorno socioeconómico de la zona y atributos individuales
- ❑ Datos del Muestra de 534 hogares realizada en el proyecto INTERLAND (GIST, 2010)

Tipo	Variable	Nombre	Descripción
TTE	Accesibilidad	ACC	Indicador de accesibilidad al empleo de la zona
TTE	Tiempo viaje	TV	Tiempo de viaje entre el centroide de la zona del uso del suelo de la vivienda y el centroide de la zona del uso del suelo del lugar de trabajo.
TTE	Tiempo espera	TE	Tiempo medio de espera en transporte público en las paradas de la zona
TTE	Interzona	IN	Variable ficticia que cuando vale uno indica que las zonas de residencia y trabajo coinciden.
ENT	Extranjeros	EXT	Nº de población extranjera no comunitaria en la zona.
ENT	Viviendas	LN_VIV	Logaritmo neperiano del nº de viviendas en la zona.
ENT	Prestigio	PG	Variable ficticia que cuando vale uno indica si la zona tiene especial prestigio (subjetivo).
ENT	Precio vivienda	PRE	Precio medio de los inmuebles de la zona.
ENT	Escuelas	ESC	Nº de escuelas en la zona
IND	Categoría de ingresos altos	H	Variable ficticia que cuando vale uno indica que ingresos altos (frente a ingresos medios o bajos) y que se ha implementado como interacción con el resto de variables (salvo la ficticia IN).

# Modelo TRANSPACE predictivo aplicado a Santander:

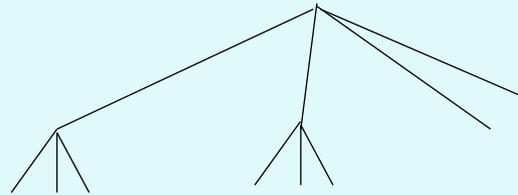
## 0. Logit Multinomial (MNL)

MNL		
Parametro variable	Valor estimado	DaP
$\beta$ TV	-0,114	0,053
$\beta$ EXT	-0,870	0,403
$\beta$ LN_VIV	1,490	-0,690
$\beta$ PG*H	1,250	-0,579
$\beta$ PRE	-2,160	
$\beta$ ESC*H	0,224	-0,104
<b>Nº parámetros estimados</b>		6
	LL	-1667,968
	$\rho^2$	0,0413
	$\rho^2$ ajustado	0,0379
	LR test – Null	143,712
	LR test – MNL	

ML-5			
Parametro variable	Valor estimado	DT estimada	DaP
$\beta$ TV	-0,118		0,056
$\beta$ EXT	-0,914		0,431
$\beta$ LN_VIV	1,490		-0,703
$\beta$ PG*H	1,120		-0,528
$\beta$ PRE	-2,120		
$\beta$ ESC*H	0,160	0,289	-0,075
<b>Nº parámetros estimados</b>			7
		LL	-1666,966
		$\rho^2$	0,0419
		$\rho^2$ ajustado	0,0379
		LR test – Null	145,716
		LR test – MNL	2,004

# i.A. NL: Con estructura de nidos predefinida por el analista

- Modelo que prevé una estructura de nidos (con un solo nivel o varios) donde cada alternativa pertenece a un solo nido

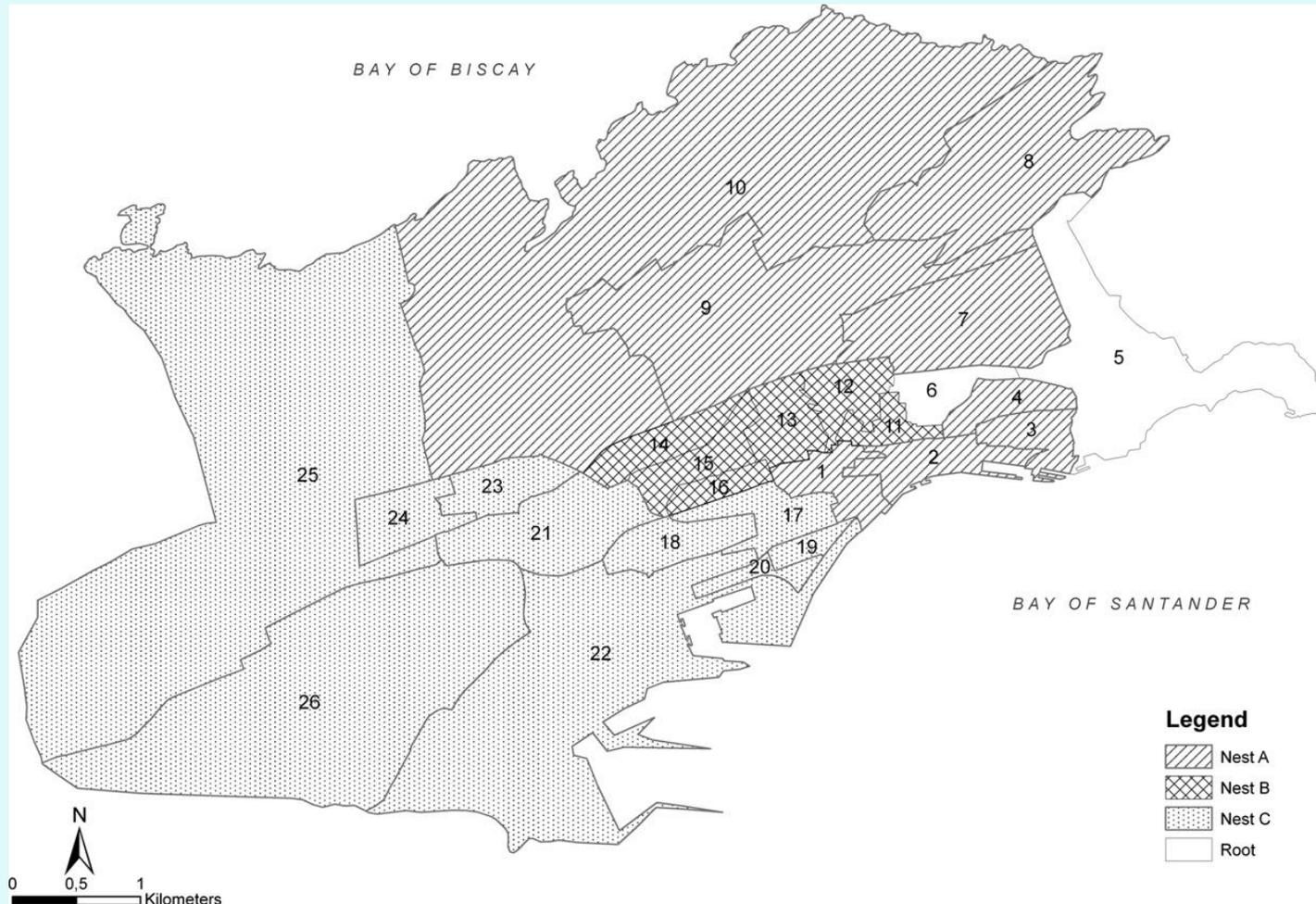


- La función GEV es

$$G(x_1, \dots, x_I) = \sum_{k=1}^K \left( \sum_{j \in C_k} x_j^{\lambda_k} \right)^{\frac{\lambda}{\lambda_k}}$$

- $\sigma_j = 1 - \lambda_j$  son parámetros a estimar que representan la correlación entre las alternativas de cada nido  $j$  (si no es significativamente distinta de cero el nido colapsa como el caso MNL)
- El analista debe definir previamente la estructura de nidos que agrupa las zonas en base a criterios no incorporados en la función de utilidad

# Ibeas, A., Cordera, R., Dell'Olio, L. y Coppola, P. (2013)



# Aplicación de NL y M-NL a Santander

NL		
Parametro variable	Valor estimado	DaP
$\beta$ TV	-0,107	0,049
$\beta$ EXT	-1,070	0,486
$\beta$ LN_VV	1,660	-0,755
$\beta$ PG*H	1,280	-0,582
$\beta$ PRE	-2,200	
$\beta$ ESC*H	0,218	-0,099
$\sigma$ A	0,160	
$\sigma$ B	0,167	
$\sigma$ C	0	
$\sigma$ D	0	
Nº parámetros estimados		8
LL		-1661,462
$\rho^2$		0,0450
$\rho^2$ ajustado		0,0404
LR test – Null		156,724
LR test – MNL		13,012

M-NL			
Parametro variable	Valor estimado	DT estimada	DaP
$\beta$ TV	-0,111		0,053
$\beta$ EXT	-0,11		0,053
$\beta$ LN_VV	1,62		-0,779
$\beta$ PG*H	1,16		-0,558
$\beta$ PRE	-2,08		
$\beta$ ESC*H	0,154	0,274	-0,074
$\sigma$ A	0,174		
$\sigma$ B	0,160		
$\sigma$ C	0		
$\sigma$ D	0		
Nº parámetros estimados			9
LL			-1660,321
$\rho^2$			0,0457
$\rho^2$ ajustado			0,0405
LR test – Null			159,006
LR test – MNL			15,294

# i.B. CNL con correlación y su aplicación a Santander

- El CNL (y GNL) permite que cada zona pertenezca a varios nidos, con una grado de pertenencia a cada uno definido por el parámetro de asignación  $\alpha_{ij}$

$$0 < \alpha_{jk} < 1, \forall j, k \text{ y } \sum_k \alpha_{jk} = 1$$

- La función GEV

$$G(x_1, \dots, x_I) = \sum_{k=1}^K \left( \sum_{j \in C_k} \alpha_{jk} x_j^{\lambda_k} \right)^{\frac{\lambda}{\lambda_k}}$$

- La especificación con parámetros  $\alpha_{ij}$  variables tiene problemas de convergencia
- La especificación con  $\alpha_{ij}$  fijos requiere una estructura espacial, por ejemplo “vecindad” como en SCL

CNL		
Parametro variable	Valor estimado	DaP
$\beta$ TV	-0,103	0,064
$\beta$ EXT	-0,696	0,430
$\beta$ LN_VIV	1,240	-0,765
$\beta$ PG*H	0,950	-0,586
$\beta$ PRE	-1,620	
$\beta$ ESC*H	0,182	-0,112
$\sigma$ A	0,091	
$\sigma$ B	0,145	
$\sigma$ C	0,200	
$\sigma$ D	0,383	
<b>Nº parámetros estimados</b>		10
LL		-1666,371
$\rho^2$		0,0422
$\rho^2$ ajustado		0,0365
LR test – Null		146,906
LR test – MNL		3,194

El modelo CNL colapsa a uno MNL

## ii. MELR con correlación espacial pero sin una estructura de nidos predefinida por el analista (I)

- ii.A. PCL → al igual que CNL
  - Tendrá problemas de convergencia si los parámetros de asignación a nidos son variables por su elevado n° en el caso ELR
  - La especificación de valores fijos de los parámetros de asignación a nido es el método seguido por los modelos con correlación
- Logit espacialmente correlacionado (SCL, Bhat y Guo, 2004) → **MSCL**

- Se basa en PCL:
  - Se definen los nidos como pares de zonas adyacentes
  - Se especifican valores fijos de los parámetros de asignación a nidos  $\alpha_{ij}$
- Variable ficticia  $\omega_{ij}$  : 1 cuando las zonas  $i$  y  $j$  son adyacentes y 0 cuando no (por convenio  $\omega_{ij}=0$ )

- Parámetro de asignación de cada zona  $i$  al nido (pareado)  $ij$

$$\alpha_{i,j} = \frac{\omega_{ij}}{\sum_k \omega_{ik}}$$

$$\sum_j \alpha_{i,j} = 1$$

- Función generatriz GEV

$$G(x_1, \dots, x_I) = \sum_{i=1}^{I-1} \sum_{j=i+1}^I ((\alpha_{i,j} x_i)^{\frac{1}{\lambda}} + (\alpha_{j,i} x_j)^{\frac{1}{\lambda}})^{\lambda}$$

# Aplicación SCL y MSCL

SCL		
Parametro variable	Valor estimado	DaP
$\beta$ TV	-0,174	0,070
$\beta$ EXT	-1,030	0,412
$\beta$ LN_VIV	1,930	-0,772
$\beta$ PG*H	1,560	-0,624
$\beta$ PRE	-2,500	
$\beta$ ESC*H	0,274	-0,110
$\sigma$	0,357	
<b>Nº parámetros estimados</b>		7
LL		-1665,94
$\rho^2$		0,0425
$\rho^2$ ajustado		0,0384
LR test – Null		147,768
LR test – MNL		4,056

M-SCL			
Parametro variable	Valor estimado	DT estimada	DaP
$\beta$ TV	-0,182		0,075
$\beta$ EXT	-1,070		0,440
$\beta$ LN_VIV	1,940		-0,798
$\beta$ PG*H	1,410		-0,580
$\beta$ PRE	-2,430		
$\beta$ ESC*H	0,206	-0,421	-0,085
$\sigma$	0,367		
<b>Nº parámetros estimados</b>			8
LL			-1664,796
$\rho^2$			0,0431
$\rho^2$ ajustado			0,0385
LR test – Null			150,056
LR test – MNL			6,344

# MELR con correlación espacial pero sin una estructura de nidos predefinida por el analista (II)

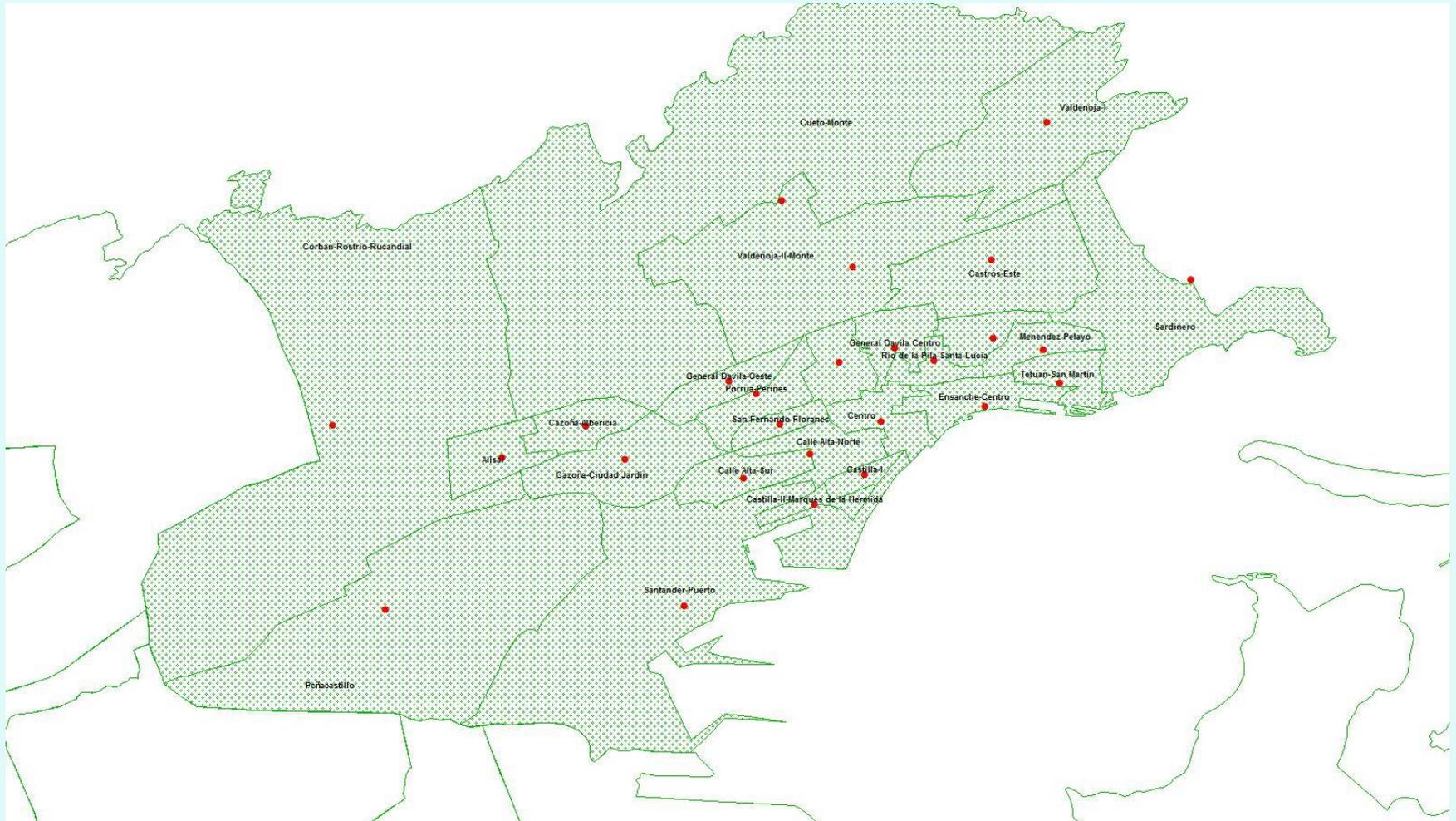
- Logit Espacialmente Correlacionado Generalizado (GSCL, Sener, Pendyala y Bhat, 2011) → **MGSCCL**
  - Se admite correlación entre zonas no adyacentes
  - Atributos  $z_{ij}$  que caracterizan la relación entre cada par de zonas

$$\alpha_{i,ij} = \frac{e^{\phi' z_{ij}}}{\sum_k e^{\phi' z_{ik}}}$$

$$\sum_j \alpha_{i,ij} = 1$$

- DSCL (Sener, Pendyala y Bhat, 2011) → **MGSCCL**
  - SCL basado en la distancia:  $z_{ij} = \ln d_{ij}$  (distancia euclídea)
  - Shen, Y., Silva, J., Martínez, L., 2014: SML (Logit Mixto Espacial) con especificación semejante a MDSCCL, aplicado a la estación de Madrid Atocha (con datos de panel):  $z_{ij} = D_{ij}$  (y 0 a partir de 3km)

# Distancias entre los centroides de las zonas



# DSCL y MDSCCL con $\phi=-2$ (fijo)

DSCL		
Parametro variable	Valor estimado	DaP
$\beta$ TV	-0,114	0,053
$\beta$ EXT	-0,870	0,403
$\beta$ LN_VIV	1,490	-0,690
$\beta$ PG*H	1,250	-0,579
$\beta$ PRE	-2,160	
$\beta$ ESC*H	0,224	-0,104
$\sigma$	0	
Nº parámetros estimados		7
LL		-1667,968
$\rho^2$		0,0413
$\rho^2$ ajustado		0,0373
LR test – Null		143,712
LR test – MNL		0

M-DSCL			
Parametro variable	Valor estimado	DT estimada	DaP
$\beta$ TV	-0,118		0,056
$\beta$ EXT	-0,915		0,432
$\beta$ LN_VIV	1,490		-0,703
$\beta$ PG*H	1,120		-0,528
$\beta$ PRE	-2,120		1,000
$\beta$ ESC*H	0,159	-0,291	-0,075
$\sigma$	0		
Nº parámetros estimados			8
LL			-1666,905
$\rho^2$			0,0419
$\rho^2$ ajustado			0,0373
LR test – Null			145,838
LR test – MNL			2,126

El modelo M-DSCL colapsa a uno ML

# MELR con correlación espacial pero sin una estructura de nidos predefinida por el analista (III)

## □ Logit Espacialmente Correlado Generalizado por Proporción de Vecindad (NPSCL) → M-NPSCL

- Sener, Pendyala y Bhat, 2011 plantea la posibilidad de utilizar el perímetro común entre pares de zonas para definir modelos GSCL
- El equipo de proyecto propone el modelo NPSCL en base a ese planteamiento con la siguiente especificación

$$\alpha_{i,j} = \frac{e^{\phi' z_{ij}}}{\sum_k e^{\phi' z_{ik}}}$$

$$\sum_j \alpha_{i,j} = 1$$

- Donde
  - $z_{ij} = 0$  si las dos zonas no son adyacentes
  - $z_{ij} = \ln p_{ij}$  de la longitud del perímetro compartido por ambas zonas adyacentes

- Luego  $\alpha_{ij} = \begin{cases} 0 & \text{si } i, j \text{ no adyacentes} \\ p_{ij}^{\phi} & \text{si } i, j \text{ adyacentes} \end{cases}$

- Se hace una primera aplicación con el valor fijo  $\phi = 1$

# Aplicación de NPSCCL y M-NPSCCL con $\phi=1$ (fijo)

NPSCCL		
Parametro variable	Valor estimado	DaP
$\beta$ TV	-0,180	0,068
$\beta$ EXT	-1,120	0,426
$\beta$ LN_VIV	2,050	-0,779
$\beta$ PG*H	1,580	-0,601
$\beta$ PRE	-2,630	
$\beta$ ESC*H	0,302	-0,115
$\sigma$	0,425	
Nº parámetros estimados		7
LL		-1663,183
$\rho^2$		0,04405101
$\rho^2$ ajustado		0,0400
LR test – Null		153,282
LR test – MNL		9,57

M-NPSCCL			
Parametro variable	Valor estimado	DT estimada	DaP
$\beta$ TV	-0,192		0,075
$\beta$ EXT	-1,170		0,455
$\beta$ LN_VIV	2,060		-0,802
$\beta$ PG*H	1,460		-0,568
$\beta$ PRE	-2,570		
$\beta$ ESC*H	0,227	-0,484	-0,088
$\sigma$	0,444		
Nº parámetros estimados			8
LL			-1661,779
$\rho^2$			0,0449
$\rho^2$ ajustado			0,0403
LR test – Null			156,09
LR test – MNL			12,378

# Líneas de trabajo en curso

- ❑ Estimación de los modelos GSCL diseñados con parámetros  $\phi$  variable
- ❑ Búsqueda de nuevas especificaciones GSCL, teóricas y especialmente con ayuda de herramientas GIS
- ❑ Analizar las especificaciones de parámetros de asignación a nido (alfas) fijos utilizadas en los modelos GSCL con la estructura de nidos predefinida en el proyecto para los modelos NL y CNL
- ❑ Utilización de microsimulación de poblaciones en la categorización

**Modelos avanzados de elección de localización residencial**

**GRACIAS POR SU ATENCIÓN**

**Alfonso Orro Arcay  
José Benito Pérez López**



**Grupo de Ferrocarriles y Transportes**  
*E.T.S. de Ing. de Caminos, Canales y Puertos*  
**UNIVERSIDAD DE A CORUÑA**

